Six Genes Involved in Prognosis of Hepatocellular Carcinoma Identified by Cox Hazard Regression

Qinghong Dai
Department of Clinical Pharmacology, Xiangya Hospital, Central South University, 87 Xiangya Road, Changsha 410008

Tao Liu
Shenzhen Center for Chronic Disease Control, Shenzhen

Yongchao Gao
Department of Clinical Pharmacology, Xiangya Hospital, Central South University, 87 Xiangya Road, Changsha 410008

Honghao Zhou
Department of Clinical Pharmacology, Xiangya Hospital, Central South University, 87 Xiangya Road, Changsha 410008

Xiong Li
the First Affiliated Hospital of Guangdong Pharmaceutical University, Guangzhou510060

Wei Zhang (csuzhangwei@csu.edu.cn)
Department of Clinical Pharmacology, Xiangya Hospital, Central South University, 87 Xiangya Road, Changsha 410008

Research Article

Keywords: Cox hazard regression, DEGs, HCC, hub gene, risk score, prognostic model

Posted Date: January 19th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-140844/v1

Version of Record: A version of this preprint was published at BMC Bioinformatics on March 30th, 2021.
See the published version at https://doi.org/10.1186/s12859-021-04095-7.

Six genes involved in prognosis of hepatocellular carcinoma identified by Cox hazard regression

Qinghong Dai[3-6], Tao Liu[1], Yongchao Gao[3-6], Honghao Zhou[3-6], Xiong Li[2*], Wei Zhang[3-6*]

[1]Shenzhen Center for Chronic Disease Control, Shenzhen, China

[2]the First Affiliated Hospital of Guangdong Pharmaceutical University, Guangzhou510060, China

[3]Department of Clinical Pharmacology, Xiangya Hospital, Central South University, 87 Xiangya Road, Changsha 410008, P. R. China

[4]Institute of Clinical Pharmacology, Central South university, Hunan Key Laboratory of Pharmacogenetics, 110 Xiangya Road, Changsha 410078, P. R. China

[5]Engineering Research Center of Applied Technology of Pharmacogenomics, Ministry of Education, 110 Xiangya Road, Changsha 410078, P. R. China

[6]National Clinical Research Center for Geriatric Disorders, 87 Xiangya Road, Changsha 410008, P. R. China

Qinghong Dai and Tao Liu are both regarded as the first authors.

***Correspondence**

Wei Zhang,

Department of Clinical Pharmacology, Xiangya Hospital, Central South University, Changsha 410008; P. R. China

Email: csuzhangwei@csu.edu.cn

Xiong Li,

the First Affiliated Hospital of Guangdong Pharmaceutical University, Guangzhou510060, China

Email: 1403873213@qq.com

Abstract

Background: Hepatocellular carcinoma (HCC), derived from hepatocytes, is the main histological subtype of primary liver cancer and poses a serious threat to human health due to *the high incidence* and *poor prognosis*. This study aimed to establish a multigene prognostic model to predict the prognosis of patients with HCC.

Results: Gene expression datasets (GSE121248, GSE40873, GSE62232) were used to identify differentially expressed genes (DEGs) between tumor and adjacent or normal tissues, and then hub genes were screened by protein–protein interaction (PPI) network and Cytoscap software. Seventeen genes among hub genes were significantly associated with prognosis and used to construct a prognostic model through COX hazard regression analysis. The predictive performance of this model was evaluated with TCGA data and was further validated with independent dataset GSE14520. Six genes (*CDKN3, ZWINT, KIF20A, NUSAP1, HMMR, DLGAP5*) were involved in the prognostic model, which separated HCC patients from TCGA dataset into high- and low-risk groups. Kaplan-Meier (KM) survival analysis and risk score analysis demonstrated that low-risk group represented a survival advantage. Univariate and multivariate regression analysis showed risk score could be an independent prognostic factor. The receiver operating characteristic (ROC) curve showed there was a better predictive power of the risk score than that of other clinical indicators. At last, the results from GSE14520 demonstrated the reliability of this prognostic model in some extent.

Conclusion: This prognostic model represented significance for prognosis of HCC, and the risk score according to this model may be a better prognostic factor than other traditional clinical indicators.

Keywords: Cox hazard regression, DEGs, HCC, hub gene, risk score, prognostic model

Background

Liver cancer represent currently the sixth most frequent malignancy and the second mortality of cancer-related deaths, with more than 85000 new cases annually in the world. HCC accounts for approximately 85%~90% of liver cancer(1). The majority of HCCs occur in patients with underlying chronic liver disease and the main risk factors are the presence of hepatitis virus, alcohol abuse, obesity, *nonalcoholic* steatohepatitis and metabolic syndrome(2). Currently available treatments for HCC include *surgical resection*, *liver transplantation, chemotherapy,* radiofrequency ablation and the multikinase inhibitor sorafenib(3). However, only a small part of patients are eligible for these therapies, and the clinical efficacy is also variable and very limited for advanced HCC due to the inherent biological and genetic heterogeneity(4). Given the high incidence and mortality of HCC, which lead to serious *health problems and heavy social burden*, identifying new biomarkers to further reveal pathogenesis, predict clinical prognosis and provide individualized treatment for HCC patient are critical and urgently demanded.

The rapid development of high-throughput technology make the researches of disease-related biomarker more and more feasible and reliable(5). Generally, the occurrence and further development of tumors are caused by multiple gene abnormalities, so it is difficult for a single gene to accurately reflect the tumor characteristics. Recently, there was a view that using multiple genes to predict tumor biological features seems more convincing(6, 7). The purpose of this study was to use the gene expression data in Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) database to develop a multigene model to predict the prognosis of patients with HCC.

In this study, three GEO datasets were used to screen out hub genes. Then, a prognostic model was constructed using TCGA data on the basis of these hub genes and the predictive performance of this

model was evaluated. Finally, an independent GEO dataset was further used to validate the significance of this model. All processes of this study were based on R, Perl software and several online tools.

Methods

Dataset preparation

In this study, three raw gene expression profiles (GSE121248, GSE40873, GSE62232) were downloaded from GEO database (https://www.ncbi.nlm.nih.gov/geo)(8). GPL570 (HG-U1331_Plus_2) Affymetrix Human Genome U133 Plus 2.0 Array was performed for these datasets. The fragments per million (FPKM) expression profile of 424 HCC samples were retrieved from TCGA database (https://cancergenome.nih.gov/). In addition, GSE14520 was used as validation cohort. Table 1 listed the sample size of each dataset.

Data preprocessing and identification of DEGs

The raw data of gene expression profiles from GEO were preprocessed for background correction, log2 transformation, quantile normalization and then probeset summarization to gain gene expression matrix by using the Robust Multi-array Average (RMA) algorithm of the “affy” R package(9). GSE62232 and GSE40873 were merged into an merged dataset by Perl due to the scant nontumor samples in GSE62232, and no tumor samples in GSE40873. Given the batch effects in two datasets, the ComBat algorithm of “sva” R package was employed to remove batch effects(10). The DEGs of the merged dataset and GSE121248 were analyzed through the Empirical Bayes function in “limma” R package(11), with the thresholds of adjust $p < 0.05$ and log fold changes (log FC) > 2.0. Visualization of the overlapping genes among the merged dataset and GSE121248 was achieved by online software VENNY (https://bioinfogp.cnb.csic.es/tools/venny/).

Construction of PPI network

The Search Tool for the Retrieval of Interacting Genes database (STRING,http://string-db.org) was utilized to construct PPI network with interaction score >= 0.7(12). The subnetworks were generated by Molecular Complex Detetion (MCODE) at *default* parameters, a plugin for Cytoscape software used for clustering a significant subnetwork in the PPI network to screen hub genes(13).

Differential expression and Functional enrichment of hub genes in TCGA cohort

HCC samples in TCGA cohort were uesd to perform differential expression analysis, Gene Ontology (GO) enrichment analysis (achieved by “enrichiplot” and “org.Hs.eg.db” R packages) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis (achieved by “digest” and “Goplot” R packages) based on hub genes, Functional categories with FDR<0.05 and log FC > 2.0 were considered as significant pathways.

Construction of the prognostic model and predictive performance evaluation

Hub genes that could predict prognosis independently ($P < 0.05$) in univariate hazard regression analysis, were used to construct the prognostic model through COX hazard regression. The risk scores of TCGA samples were calculated and these samples were divided into high- and low- risk groups according to the median of risk score. Subsequently, KM survival curve, risk score analysis, independent prognostic analysis and ROC curve were implemented to evaluate the performance of this model, and the correlation between risk score and survival state was also analyzed. At last, the predictive value of the model was validated by GSE14520. The overall workflow of this study was shown in Fig. 1.

Results

Identification of DEGs

DEGs were identified by “limma” R package with the threshold of adjust $p < 0.05$ and log FC > 2.0. A total of 47 upregulated as well as 121 downegulated genes were identified from the merged dataset (Fig. 2a), and 164 upregulated and 38 downregulated genes were obtained from GSE121248 (Fig. 2b). 36 upregulated and 34 downregulated genes were further filtered through overlapping DEGs of two datasets (Fig. 2c, d), which were used to construct PPI network.

Construction of PPI network and Decision of Hub genes

PPI network was constructed by STRING online tool with interaction score >= 0.7, and 35 genes were involved in the PPI network (Fig. 3a). Further, MCODE, a plugin of Cytoscape software, was employed to filter hub genes with default parameters. Two subnetworks were found. There were 17 nodes and 135 edges in subnetwork 1 (Fig. 3b) and 3 nodes and 3 edges in subnetwork 2 (Fig. 3c).

Expression of hub genes and Functional enrichment in TCGA cohort

Expression of hub genes in TCGA cohort were analyzed, and the results exhibited that the level of all hub genes were significantly different between tumor and non-tumor samples ($p < 0.001$) (Fig. 3d, e). To functionally characterize the hub genes, GO and KEGG pathway enrichment analysis were carried out. Nuclear division and organelle fission were the most enriched GO terms (Fig. 4a, b), and KEGG pathway analysis showed that hub genes were significantly enriched in p53 signaling pathway, Rheumatoid arthritis, Cell cycle and Viral protein interaction with cytokine and cytokine receptor pathways (Fig. 4c).

Construction and predictive performance evaluation of the prognostic model

Seventeen hub genes in subnetwork 1 were applied to construct the prognostic model, while three genes in subnetwork 2 were discarded because they were not independent prognostic factors ($p > 0.05$) (Fig. 4d). Finally, the prognostic model was consisted of six genes, and risk score = 0.65619*$KIF20A$ - 0.40871*$CDKN3$ + 0.391238*$ZWINT$ - 1.07861*$NUSAP1$ + 0.757771*$DLGAP5$ + 0.479682*$HMMR$. The detailed information was shown in Table 2. The risk scores of HCC patients were calculated according to the prognostic model, and the median of risk score was defined as the cutoff to divide patients into high- and low-risk groups (n=370, which have complete survival state and risk score information). KM survival analysis showed low-risk group represented survival advantage compared with high-risk group ($p = 1.553e-06$) (Fig. 5a). ROC curve revealed that the AUC of risk score (AUC=0.792) was higher than that of other clinical parameters (AUC = 0.511, 0.504, 0.478, 0.703, 0.708, 0.508, 0.508) (n=235) (Fig. 5b). Univariate hazard regression analysis dispalyed that potential prognostic factors contained riskscore and several clinical indicators. However, only the satisfactory predictive performance of risk score persisted regardless of other clinical parameters in the multivariate hazard regression analysis ($p < 0.001$, n=235, which have complete clinical and risk score information) (Fig. 5c, d). Risk score analysis illustrated that death cases were increased and survival time was incrementally reduced along with increased risk score (n=370) (Fig. 6a-c). In adition, the risk score of death cases were significantly higher than that of alive individuals ($p = 2.0e-05$) (Fig. 6d), and the distribution of risk score relative to tumor size was displayed in Fig. 6e. These results suggested the potential significance of the prognostic model.

Validation cohort

The predictive stability of the prognostic model was validated with GSE14520 dataset. The risk scores of tumor patients were significantly higher than that of the normal controls (Fig. 7a). KM survival

analysis showed that the high-risk group displayed poorer survival compared with the low-risk group, while it did not reach statistical significance (Fig. 7b). Similarly, there was not significant correlation between risk score and TNM stage although the risk score gradually increased as the development of TNM stage (Fig. 7c). Tumor samples were divided into large and small groups with a diameter of 5cm, and lower expression scores were significantly associated with smaller tumor size (Fig. 7d). These results suggested that the prognostic model may function as an independent biomarker to predict the outcome of patients with HCC.

Discussion

Patients with HCC generally have a poor prognosis, and there have been numerous studies to explore clinical biological signatures. In this study, three GEO datasets were used to analysis of DEGs. Subsequently, 35 genes were selected by PPI, and then twenty hub genes were generated by Cytoscape software. In order to explore the function of these hub genes, GO and KEGG enrichment analysis were carried out hosted on the TCGA cohort, and the results displayed that nuclear- and chromosome-related GO term, p53 signaling pathway and cell cycle were the main enrichment pathways. Seventeen genes ($P < 0.05$) were selected from twenty hub genes by univariate regression analysis to construct the prognostic model by COX hazard regression analysis using TCGA data, and finally, six genes (*CDKN3, ZWINT, NUSAP1, DLGAP5, HMMR, KIF20A*) were involved in the prognostic model.

KIF20A is associated with drug resistance and the clinical prognosis in diverse cancers. Previous studies suggest high expression of *KIF20A* is linked with poor clinical outcomes(14), and maybe involved in process of transformation of cirrhosis to HCC(15). In terms of drug resistance, *KIF20A* promotes paclitaxel resistance of breast cancer(16), and also insensitizes colorectal tumor to

chemotherapy(17). In this study, the expression of *KIF20A* was positively correlated with the risk score that indicated poor outcomes.

DLGAP5, also known as *HURP*, is an important mediator for chromosome congression and alignment. Compelling evidence elucidates that *DLGAP5* promotes the development of non-small cell lung cancer(18), and is overexpressed in HCC and plays a critical role in the cancer cell cycle(19). Vice versa, a study confirms *DLGAP5* silence could inhibit HCC cell cycle and proliferation(20). We also suggested *DLGAP5* was a risk factor for HCC.

Many reports show that *ZWINT* is a predictor of tumor development. *ZWINT* is relative to risk index in pulmonary adenocarcinoma, that implies high level of *ZWINT* is correlated with poor outcomes(21). Similarly, elevated *ZWINT* could promote HCC clinicopathological features, and also possibly result in reduced overall survival and rising tumor recurrence(22). A study of prostate suggest *ZWINT* upregulation is correlated with higher Gleson scores and tumor grade(23).

The correlation between increased HMMR and poor prognosis has been reported in a variety of malignant tumors, including breast cancer(24), lung cancer(25), stomach cancer(26) and glioblastoma(27). Our results were in accordance with previous studies. In addition, *HMMR* may be contributed to proliferation, metastasis and invasion of breast cancer(28).

CDKN3, as tumor repressor, encodes protein that belongs to the dual-specificity protein phosphatase family. The role of *CDKN3* has been controversial in tumor progression. Increasing evidences suggest that *CDKN3* could promote tumor progression. Overexpression of *CDKN3* is associated with poor prognosis in lung adenocarcinoma(29), and the silence blocks proliferation and metastasis of pancreatic ductal adenocarcinoma(30). In contrast, *CDKN3* is relatively downregulated in brian tumor compared with normal brain tissue(31). In adition, the level of *CDKN3* is negatively

correlated with HCC clinical pathological stage, and downregulation of *CDKN3* promotes tumor clonogenic ability(32). The present study was consistent with later that *CDKN3* was a protective factor in tumor development. The role of *CDKN3* in tumors needs to be further investigated.

The function of *NUSAP1*, the last signature of the prognostic model, has also been controversial in tumor progression. It's reported that HCC patients with upregulated *NUSAP1* possess reduced survival times(33). Similar results are observed in a study of melanoma(34). Moreover, *NUSAP1* is involved in the resistance to antitumor therapy(35). However, current understanding of cervical cancer debates that low expression of *NUSAP1* is associated with higher tumor stage, and results in worse clinical outcomes(36). Our results illustrated the coefficient of *NUSAP1* was negative which implied the high level of *NUSAP1* predicted the survival advantage of HCC patients. The function of *NUSAP1* in tumor development need to be further explored by biomolecular and cellular research.

Finally, the predictive performance of the prognostic model was evaluated. K-M curve and risk score analysis indicated low-risk group had better prognosis than high-risk group, and univariate and multivariate regression analysis showed risk score might be an independent prognostic factor. Meanwhile, ROC analysis displayed the AUC of risk score was higher than that of other clinical indicators which illustrated risk score hold more prognostic value. In addition, the risk score of death cases were higher significantly than that of alive patients. The results of validation cohort also showed this prognostic model represented a prognostic significance for patients with HCC. Based on the analysis above, it was reasonable to regard risk score as a prognostic biomarker for HCC.

Conclusion

We utilized bioinformatics methods to analyze HCC-related gene expression profiles from GEO

and TCGA data. A prognostic model involving six genes was constructed through Cox hazard regression analysis, and the results of predictive performance evaluation represented the clinical value of this modle. At last, the consistent findings in validation cohort demonstrated that the prognostic model may be used as a tool to achive risk stratification of patients with HCC. For patients with higher risk score, more intensive systemic surveillance and therapy could be considered. Considering our attempt was definitely exploratory and the clinical value of the prognostic model to accurately predict prognosis was the ultimal goal, this work should not be regarded as the definitive result and more external verification work are needed to validate the predictive performance of this prognostic model.

Abbreviations

DEG: Differentially Expressed Gene; FC: Fold Change; GEO: Gene Expression Omnibus; GO: Gene Ontology; HCC: Hepatocellular Carcinoma; KM: Kaplan-Meier; PPI: Protein–Protein Interaction; ROC: Receiver Operating Characteristic; TCGA: The Cancer Genome Atlas

Declarations

Ethics approval and consent to participate Not applicable.

Consent to publication Not applicable.

Availability of data and materials The datasets analyzed during this study are publicly available in GEO database at https://www.ncbi.nlm.nih.gov/geo/ and TCGA database at https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga.

Competing interests The authors declare that they have no conflict of interests.

Funding This study was supported by the National Natural Science Foundation of China (No. 81874329 and 82073945).

Authors' contributions Qinghong Dai, Tao Liu and Wei Zhang designed the study; Qinghong Dai analyzed the data, made the figures and drafted the manuscript; Yongchao Gao polished the whole manuscript; Wei Zhang, Xiong Li and Honghao Zhou revised the whole paper and approved the final paper.

Acknowledgements Not applicable

References

1. Llovet JM, Zucman-Rossi J, Pikarsky E, Sangro B, Schwartz M, Sherman M, et al. Hepatocellular carcinoma. Nat Rev Dis Primers. 2016;2:16018.
2. Fujiwara N, Friedman SL, Goossens N, Hoshida Y. Risk factors and prevention of hepatocellular carcinoma in the era of precision medicine. J Hepatol. 2018;68(3):526-49.
3. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. Lancet. 2018;391(10127):1301-14.
4. Reig M, da Fonseca LG, Faivre S. New trials and results in systemic treatment of HCC. J Hepatol. 2018;69(2):525-33.
5. Reuter JA, Spacek DV, Snyder MP. High-throughput sequencing technologies. Mol Cell. 2015;58(4):586-97.
6. Jeong DH, Kim WR, Min BS, Kim YW, Song MK, Kim NK. Validation of a quantitative 12-multigene expression assay (Oncotype DX((R)) Colon Cancer Assay) in Korean patients with stage II colon cancer: implication of ethnic differences contributing to differences in gene expression. Onco Targets Ther. 2015;8:3817-25.
7. You YN, Rustin RB, Sullivan JD. Oncotype DX((R)) colon cancer assay for prediction of recurrence risk in patients with stage II and III colon cancer: A review of the evidence. Surg Oncol. 2015;24(2):61-6.
8. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30(1):207-10.
9. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics. 2003;19(2):185-93.
10. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch

effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28(6):882-3.
11. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
12. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43(Database issue):D447-52.
13. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics. 2003;4:2.
14. Bayo J, Fiore EJ, Dominguez LM, Real A, Malvicini M, Rizzo M, et al. A comprehensive study of epigenetic alterations in hepatocellular carcinoma identifies potential therapeutic targets. J Hepatol. 2019;71(1):78-90.
15. Lin Y, Liang R, Ye J, Li Q, Liu Z, Gao X, et al. A twenty gene-based gene set variation score reflects the pathological progression from cirrhosis to hepatocellular carcinoma. Aging (Albany NY). 2019;11(23):11157-69.
16. Khongkow P, Gomes AR, Gong C, Man EP, Tsang JW, Zhao F, et al. Paclitaxel targets FOXM1 to regulate KIF20A in mitotic catastrophe and breast cancer paclitaxel resistance. Oncogene. 2016;35(8):990-1002.
17. Xiong M, Zhuang K, Luo Y, Lai Q, Luo X, Fang Y, et al. KIF20A promotes cellular malignant behavior and enhances resistance to chemotherapy in colorectal cancer through regulation of the JAK/STAT3 signaling pathway. Aging (Albany NY). 2019;11(24):11905-21.
18. Tagal V, Wei S, Zhang W, Brekken RA, Posner BA, Peyton M, et al. SMARCA4-inactivating mutations increase sensitivity to Aurora kinase A inhibitor VX-680 in non-small cell lung cancers. Nat Commun. 2017;8:14098.
19. Tsou AP, Yang CW, Huang CY, Yu RC, Lee YC, Chang CW, et al. Identification of a novel cell cycle regulated gene, HURP, overexpressed in human hepatocellular carcinoma. Oncogene. 2003;22(2):298-307.
20. Breuer M, Kolano A, Kwon M, Li CC, Tsai TF, Pellman D, et al. HURP permits MTOC sorting for robust meiotic spindle bipolarity, similar to extra centrosome clustering in cancer cells. J Cell Biol. 2010;191(7):1251-60.
21. Endoh H, Tomida S, Yatabe Y, Konishi H, Osada H, Tajima K, et al. Prognostic model of pulmonary adenocarcinoma by expression profiling of eight genes as determined by quantitative real-time reverse transcriptase polymerase chain reaction. J Clin Oncol. 2004;22(5):811-9.
22. Ying H, Xu Z, Chen M, Zhou S, Liang X, Cai X. Overexpression of Zwint predicts poor prognosis and promotes the proliferation of hepatocellular carcinoma by regulating cell-cycle-related proteins. Onco Targets Ther. 2018;11:689-702.
23. Song ZY, Chao F, Zhuo Z, Ma Z, Li W, Chen G. Identification of hub genes in prostate cancer using robust rank aggregation and weighted gene co-expression network analysis. Aging (Albany NY). 2019;11(13):4736-56.
24. Pujana MA, Han JD, Starita LM, Stevens KN, Tewari M, Ahn JS, et al. Network modeling links breast cancer susceptibility and centrosome dysfunction. Nat Genet. 2007;39(11):1338-49.
25. Stevens LE, Cheung WKC, Adua SJ, Arnal-Estape A, Zhao M, Liu Z, et al. Extracellular Matrix Receptor Expression in Subtypes of Lung Adenocarcinoma Potentiates Outgrowth of Micrometastases. Cancer Res. 2017;77(8):1905-17.

26. Zhang H, Ren L, Ding Y, Li F, Chen X, Ouyang Y, et al. Hyaluronan-mediated motility receptor confers resistance to chemotherapy via TGFbeta/Smad2-induced epithelial-mesenchymal transition in gastric cancer. FASEB J. 2019;33(5):6365-77.
27. Tilghman J, Wu H, Sang Y, Shi X, Guerrero-Cazares H, Quinones-Hinojosa A, et al. HMMR maintains the stemness and tumorigenicity of glioblastoma stem-like cells. Cancer Res. 2014;74(11):3168-79.
28. Schwertfeger KL, Cowman MK, Telmer PG, Turley EA, McCarthy JB. Hyaluronan, Inflammation, and Breast Cancer Progression. Front Immunol. 2015;6:236.
29. Fan C, Chen L, Huang Q, Shen T, Welsh EA, Teer JK, et al. Overexpression of major CDKN3 transcripts is associated with poor survival in lung adenocarcinoma. Br J Cancer. 2015;113(12):1735-43.
30. Liu D, Zhang J, Wu Y, Shi G, Yuan H, Lu Z, et al. YY1 suppresses proliferation and migration of pancreatic ductal adenocarcinoma by regulating the CDKN3/MdM2/P53/P21 signaling pathway. Int J Cancer. 2018;142(7):1392-404.
31. Nalepa G, Barnholtz-Sloan J, Enzor R, Dey D, He Y, Gehlhausen JR, et al. The tumor suppressor CDKN3 controls mitosis. J Cell Biol. 2013;201(7):997-1012.
32. Dai W, Miao H, Fang S, Fang T, Chen N, Li M. CDKN3 expression is negatively associated with pathological tumor stage and CDKN3 inhibition promotes cell survival in hepatocellular carcinoma. Mol Med Rep. 2016;14(2):1509-14.
33. Roy S, Hooiveld GJ, Seehawer M, Caruso S, Heinzmann F, Schneider AT, et al. microRNA 193a-5p Regulates Levels of Nucleolar- and Spindle-Associated Protein 1 to Suppress Hepatocarcinogenesis. Gastroenterology. 2018;155(6):1951-66 e26.
34. Bogunovic D, O'Neill DW, Belitskaya-Levy I, Vacic V, Yu YL, Adams S, et al. Immune profile and mitotic index of metastatic melanoma lesions enhance clinical staging in predicting patient survival. Proc Natl Acad Sci U S A. 2009;106(48):20429-34.
35. Emanuele MJ, Elia AE, Xu Q, Thoma CR, Izhar L, Leng Y, et al. Global identification of modular cullin-RING ligase substrates. Cell. 2011;147(2):459-74.
36. Xie Q, Ou-Yang W, Zhang M, Wang H, Yue Q. Decreased Expression of NUSAP1 Predicts Poor Overall Survival in Cervical Cancer. J Cancer. 2020;11(10):2852-63.

Figure legends

Fig. 1 Overall process of this study. MCODE, a plugin for Cytoscape.

Fig. 2 Differantially expressed genes. (a) Merged dataset generated from GSE62232 and GSE40873. (b) GSE121248. (c, d) Venn plot of shared gene between merged dataset and GSE121248.

Fig. 3 Hub genes. (a) PPI network was consisted of 35 genes,interaction score >= 0.7 was the cutoff value. (b, c) Subnetwork 1 and subnetwork 2 identified by MCODE. Red represented upregulated gene and green represent downregulated gene. (d, e) Twenty hub genes were differencially expressed between tumor and nontumor samples in TCGA cohort ($p < 0.001$). “N” meant nontumor group and “T” meant tumor group.

Fig. 4 Functional enrichment and univariate regression analysis of hub genes in TCGA cohort. (a, b) GO enrichment analysis. CC, cellular component. BP, biological process. MF, molecular function. (c) Circle plot of KEGG pathway. (d) Univariate hazard regression analysis of hub genes.

Fig. 5 Predictive performance of prognostic model. (a) KM survival curve. (b) ROC curve of multiple indicators. (c) univariate hazard regression analysis. (d) multivariate hazard regression analysis.

Fig. 6 Risk score analysis. (a) Samples were sorted according to risk score from low to high. (b) Correlation between survical time and risk score. (c) Heatmap of six genes expression involved in prognosis model. (d) Correlation between risk score and fustate. (e) Box plot of risk score relative to tumor size.

Fig. 7 Results of Validation cohort. (a) Comparison of risk score between tumor and normal samples. (b) KM survival curve , the cutoff divided tumor samples into two groups was the median of risk scores. (c) Distribution of risk score relative to tumor stage in HCC. (d) Comparison of risk score between small and large tumors. The diameter of 5 cm was the cutoff value.

Figures

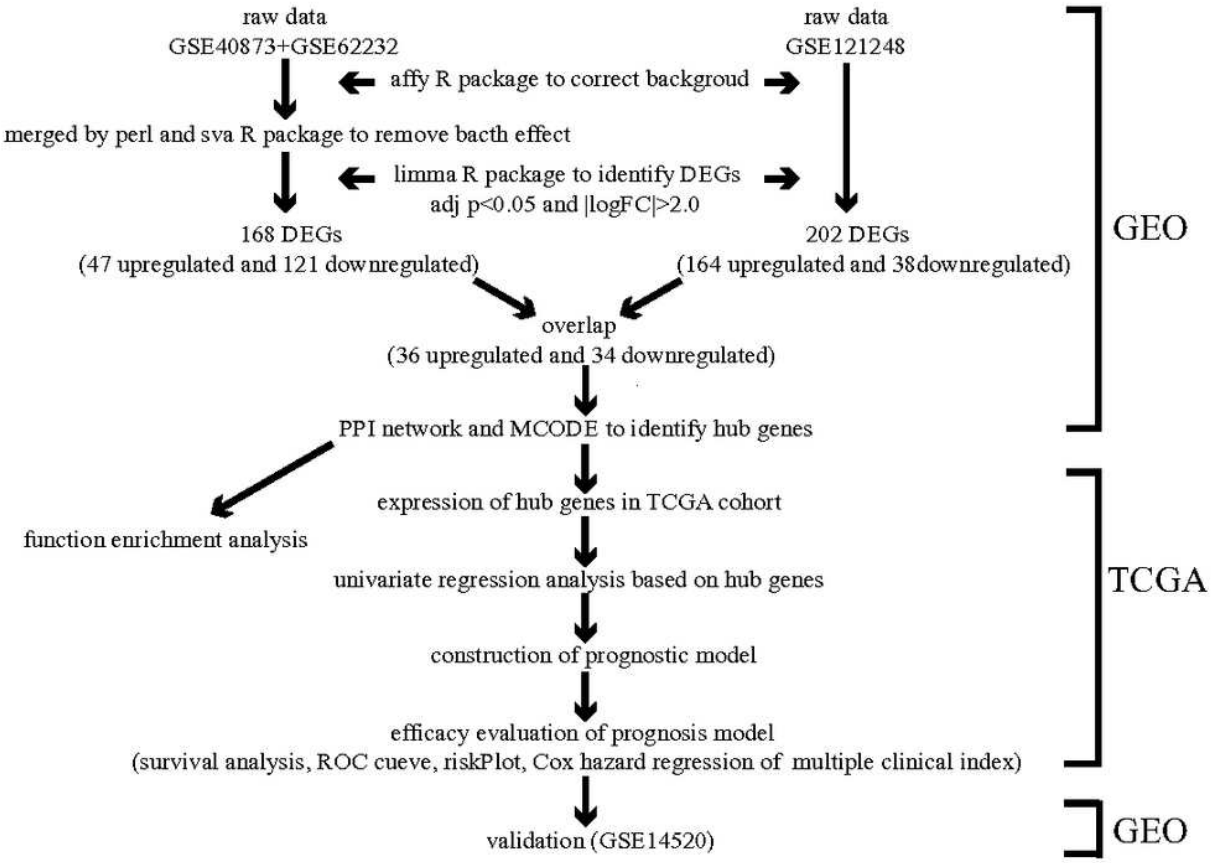

Figure 1

Overall process of this study. MCODE, a plugin for Cytoscape.

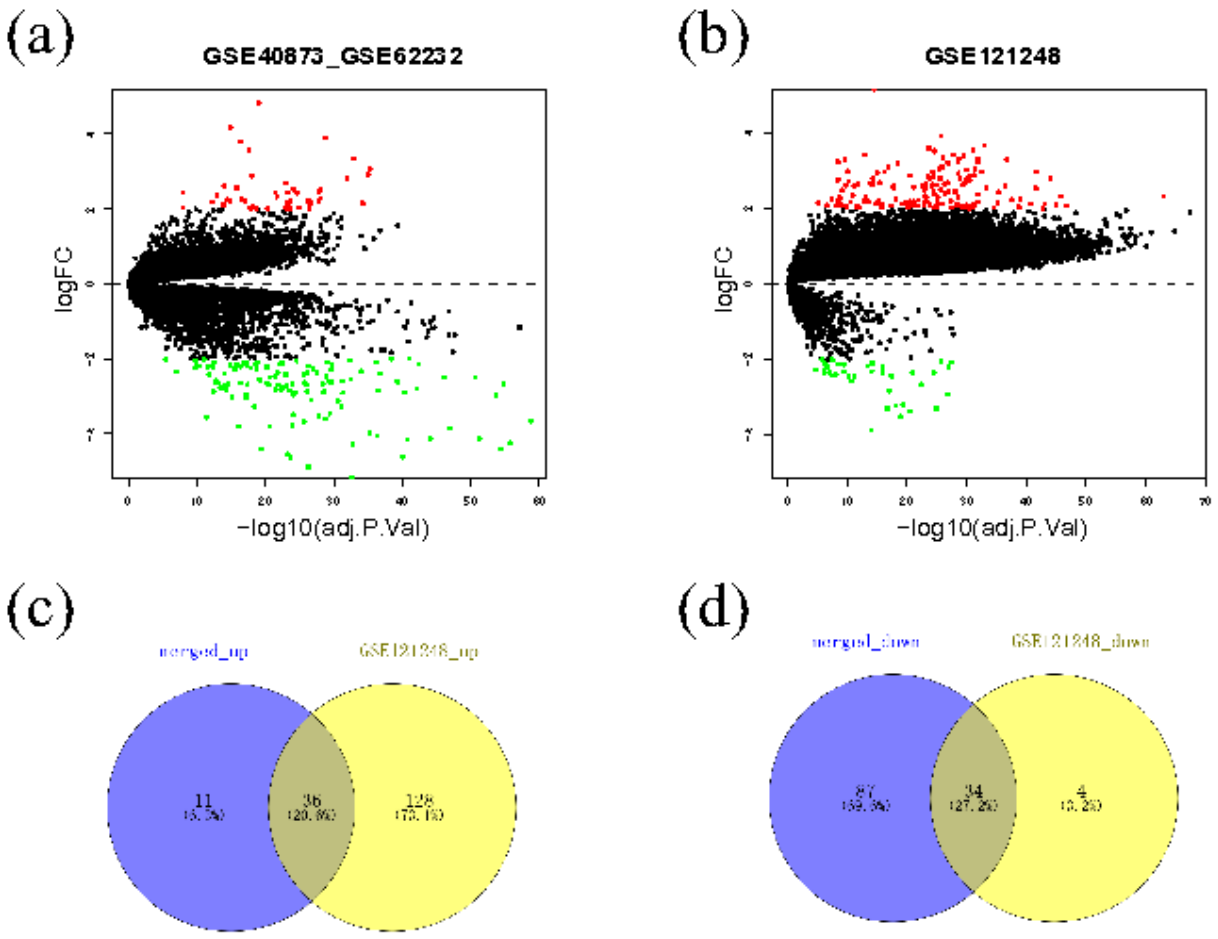

Figure 2

Differantially expressed genes. (a) Merged dataset generated from GSE62232 and GSE40873. (b) GSE121248. (c, d) Venn plot of shared gene between merged dataset and GSE121248.

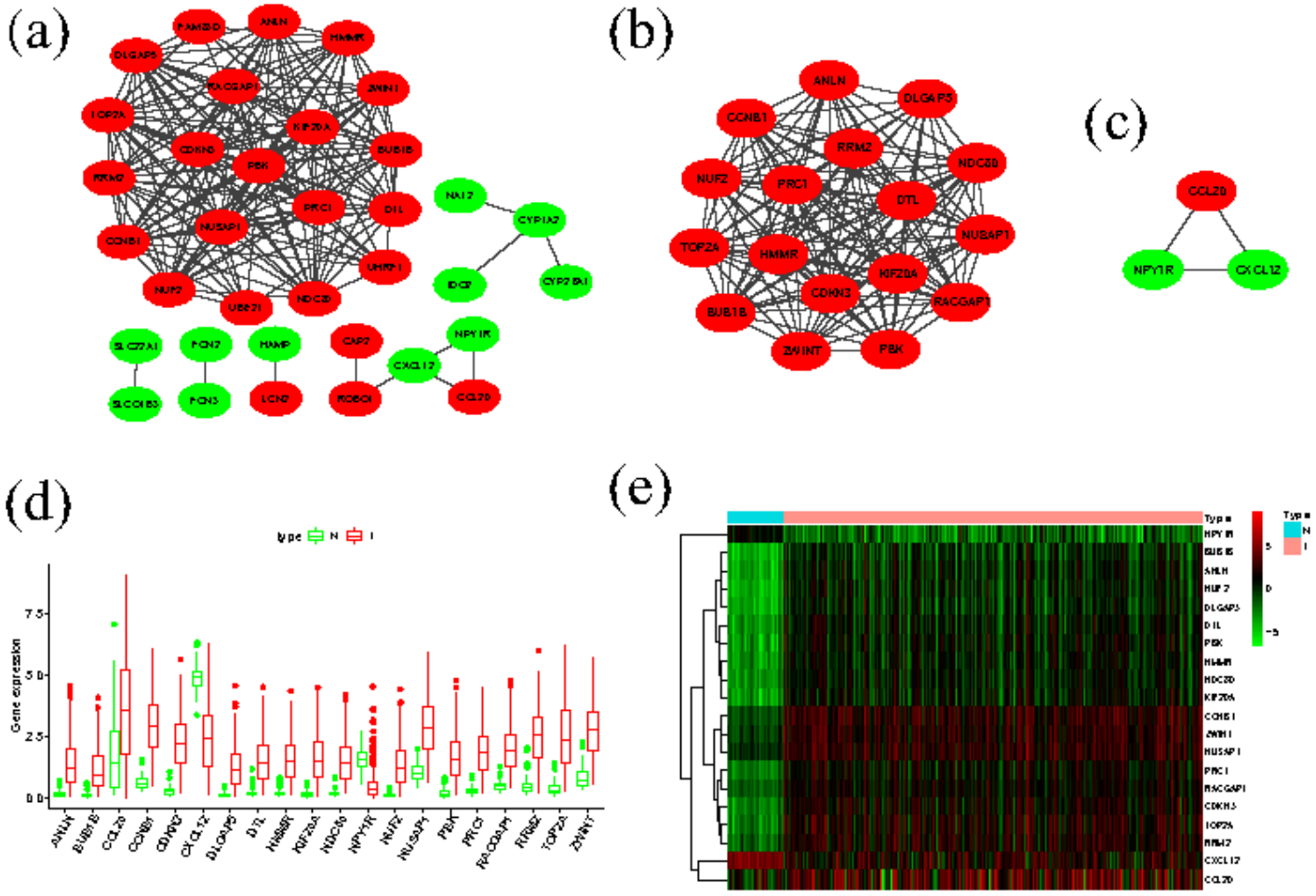

Figure 3

Hub genes. (a) PPI network was consisted of 35 genes, interaction score >= 0.7 was the cutoff value. (b, c) Subnetwork 1 and subnetwork 2 identified by MCODE. Red represented upregulated gene and green represent downregulated gene. (d, e) Twenty hub genes were differencially expressed between tumor and nontumor samples in TCGA cohort ($p < 0.001$). “N” meant nontumor group and “T” meant tumor group.

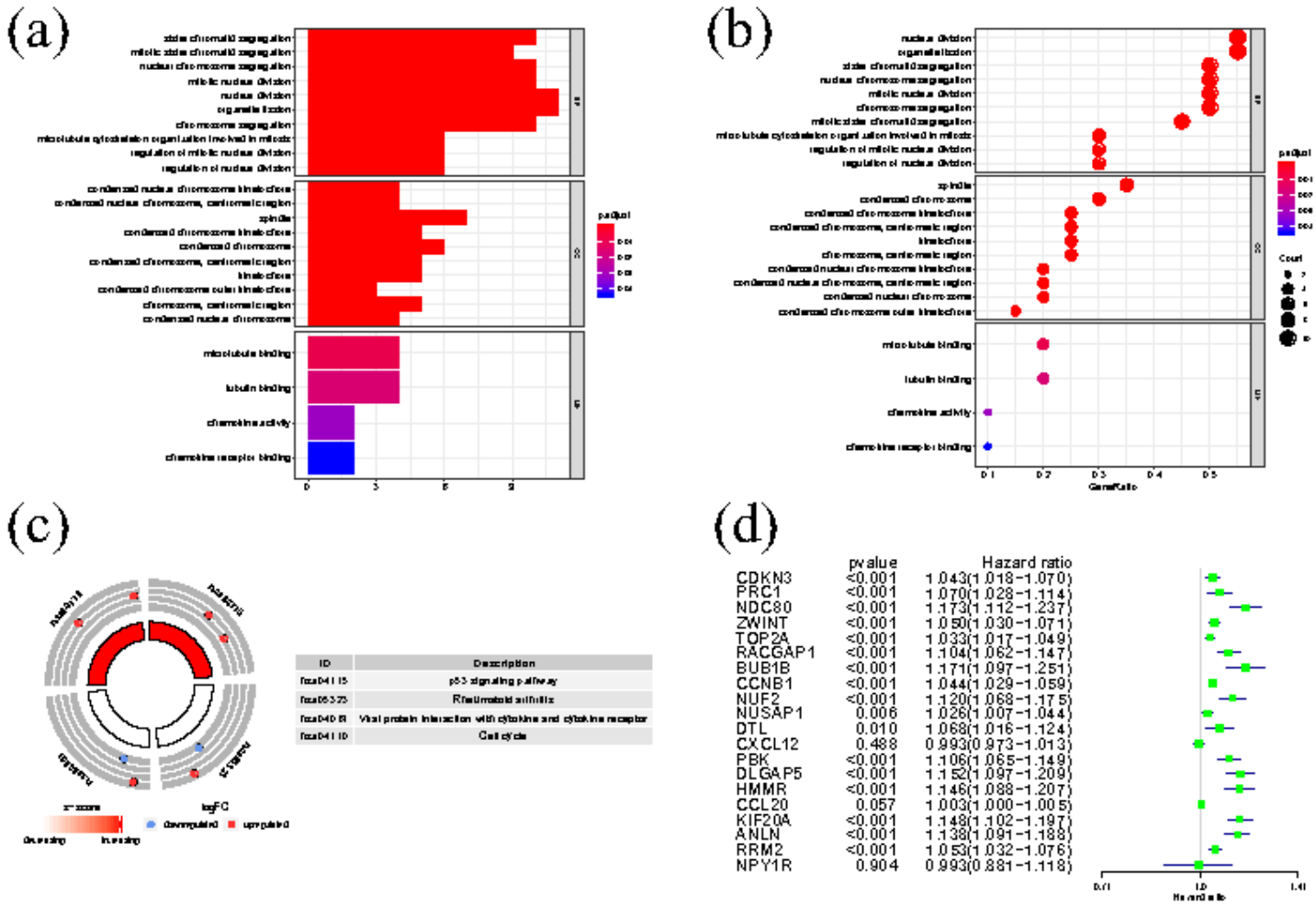

Figure 4

Functional enrichment and univariate regression analysis of hub genes in TCGA cohort. (a, b) GO enrichment analysis. CC, cellular component. BP, biological process. MF, molecular function. (c) Circle plot of KEGG pathway. (d) Univariate hazard regression analysis of hub genes.

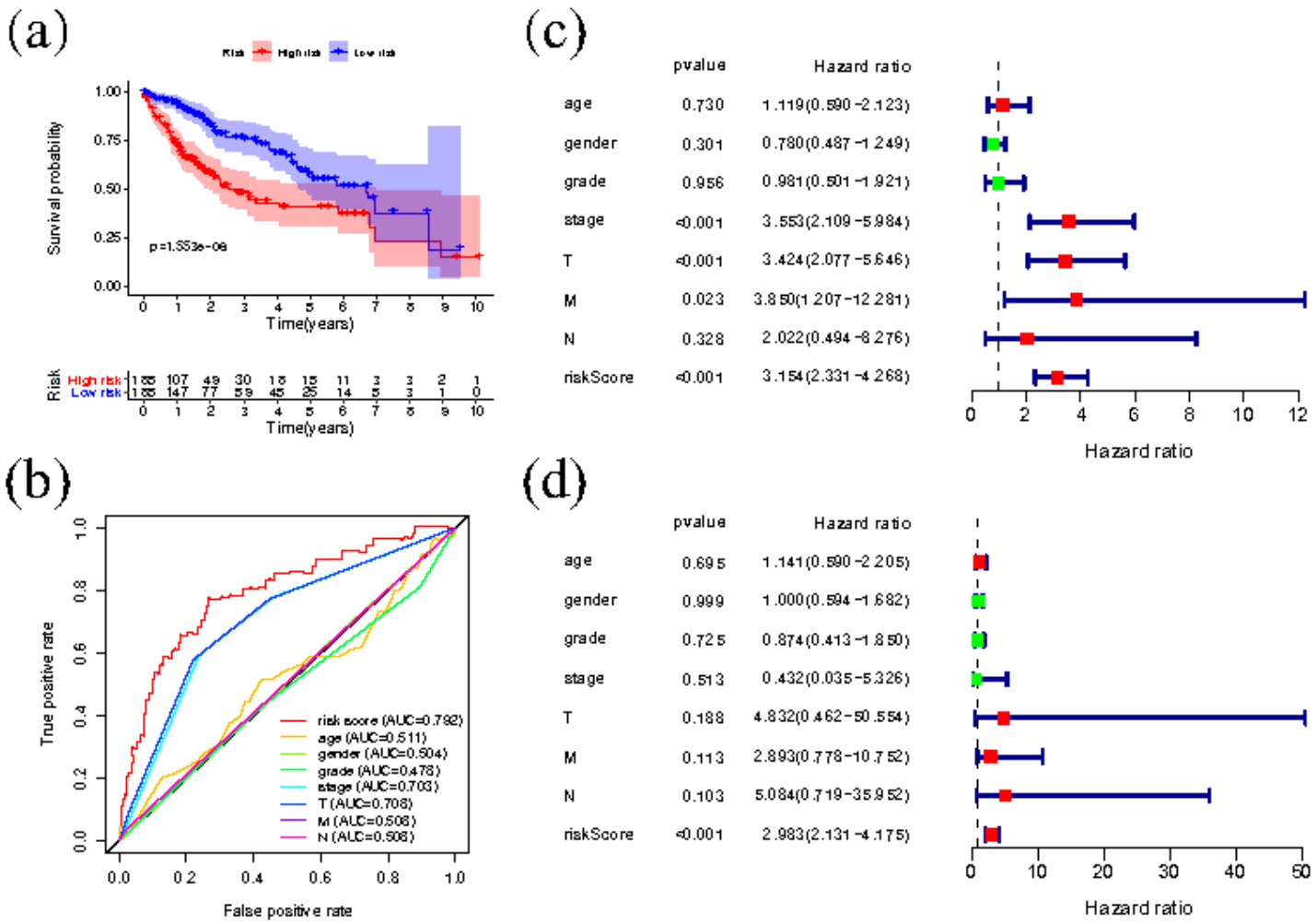

Figure 5

Predictive performance of prognostic model. (a) KM survival curve. (b) ROC curve of multiple indicators. (c) univariate hazard regression analysis. (d) multivariate hazard regression analysis.

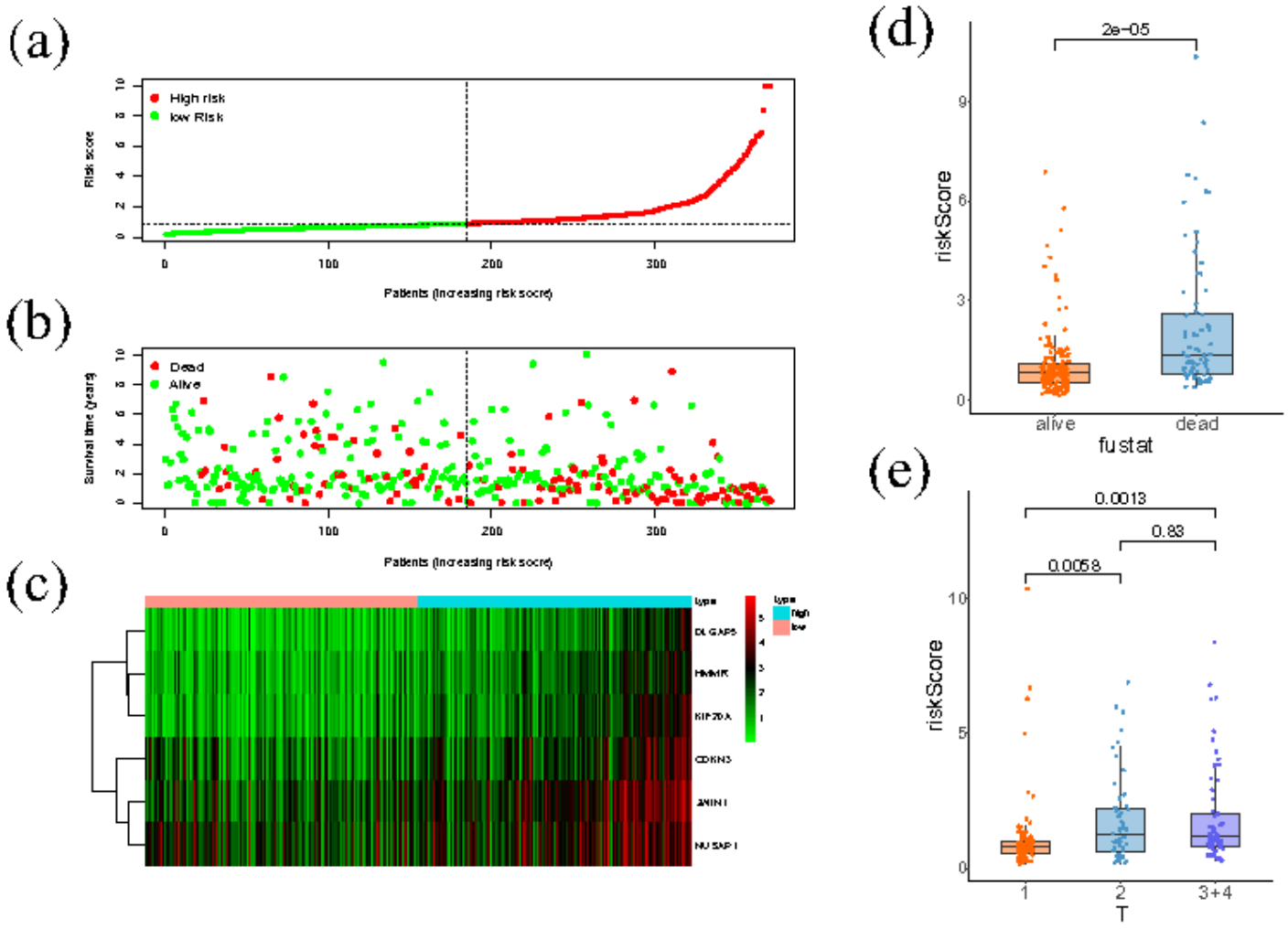

Figure 6

Risk score analysis. (a) Samples were sorted according to risk score from low to high. (b) Correlation between survical time and risk score. (c) Heatmap of six genes expression involved in prognosis model. (d) Correlation between risk score and fustate. (e) Box plot of risk score relative to tumor size.

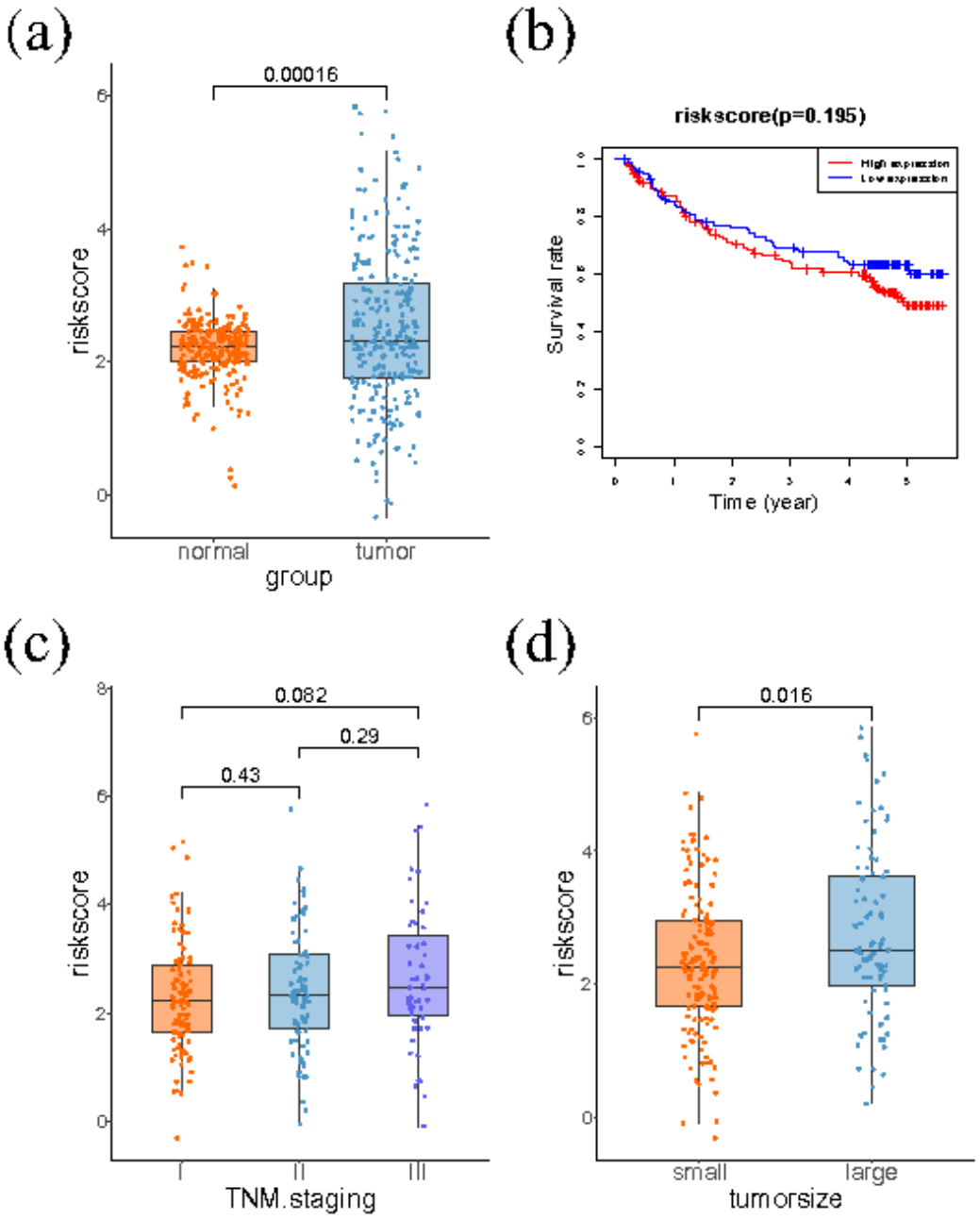

Figure 7

Results of Validation cohort. (a) Comparison of risk score between tumor and normal samples. (b) KM survival curve , the cutoff divided tumor samples into two groups was the median of risk scores. (c) Distribution of risk score relative to tumor stage in HCC. (d) Comparison of risk score between small and large tumors. The diameter of 5 cm was the cutoff value.