Development and Validation an Epigenetic Modification-related Signals for Diagnosis and Prognosis of Hepatocellular Carcinoma

Diguang Wen
Chongqing Medical University
Sheng Qiu
Chongqing Medical University
Zuojin Liu (300376@hospital.cqmu.edu.cn)
Chongqing Medical University https://orcid.org/0000-0003-1440-5693

Research article

Keywords: Bioinformatics, Hepatocellular carcinoma, Epigenetic modification, Diagnosis, Prognosis

Posted Date: January 5th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-137969/v1

Abstract

Background: Increasing evidence has indicated that abnormal epigenetic modification such as RNAm6a modification, histone modification, DNA methylation modification, RNA binding proteins and transcription factors, is correlated with Hepatocarcinogenesis. However, it is unknown how epigenetic modification associated genes contribute to the occurrence and clinical outcome of hepatocellular carcinoma (HCC). Thus, we constructed epigenetic modification associated model that may enhance the diagnosis and prognosis of HCC.

METHODS: In this study, we focused on the clinical values of epigenetic modification associated genes for HCC. Our gene expression data were collected from TCGA and a HCC datasets from GEO dataset in order to ensure the reliability of data. Their function was analyzed by bioinformatics methods. We used lasso regression, SUV, logistic regression and cox regression to construct the diagnosis and prognosis models. We also constructed a nomogram for the practicability of the above-mentioned prognosis model. The above results have been verified in an independent liver cancer dataset from ICGC database. Furthermore, we carried out pan cancer analysis to verify the specificity of the above model.

RESULT: A large number of epigenetic modification associated genes were significantly different in HCC and normal liver tissues. The gene signatures showed good performance for predicting the occurrence and survival of HCC patients verified by DCA and ROC curve.

CONCLUSION: Gene signatures based on epigenetic modification associated genes can be used to identify the occurrence and prognosis of liver cancer.

1. Introducation

Hepatocellular carcinoma (HCC) is one of the most common and fatal malignancies in the world and the number of HCC cases increases gradually by about 4% every year [14]. Though early HCC can be treated by tumor resection and liver transplantation, median survival time of this disease is only a few months [15]. The poor prognosis and short lifetime of HCC, to some extent, is dependent on late diagnosis and lack of effective treatment plan due to early liver cancer has no specific symptoms.

The diagnosis of HCC mainly depends on biopsy and imaging evidence [16, 17]. However, due to the technical difficulty, the above diagnostic indicators may be affected by subjective factors, resulting in false positive or false negative rates [18]. The development of new diagnostic technology can better assist the traditional diagnostic methods and help to improve the detection rate of early liver cancer patients. Accurate judgment of patient's prognosis is helpful for guiding clinical decision-making and implementation of precision medicine. Currently, the prognosis is mainly judged by BCLC and TNM stages, which is far from enough to predict the outcome of patients. Therefore, it is necessary to explore effective diagnostic and prognostic biomarkers to help optimize the treatment system of HCC. In the past few decades, researchers have come to realize more and more tumorigenesis mechanism. One of the breakthroughs is the participation of epigenetics process in the development of cancer [19].

Epigenetics is a dynamic and heritable modification of independent DNA sequences [20]. Abnormal epigenetic changes can destroy the expression balance of oncogenes and tumor suppressor genes, and promote tumorigenesis. Common epigenetic modifications include DNA methylation, RNAm6a methylation, histone acetylation and etc. which are considered as main mechanisms of regulation during cancer progression. Previous studies have focused on the functional exploration of single epigenetic related genes, but lack of extensive exploration. At the same time, the value of these genes in the diagnosis and prognosis of liver cancer is not fully clear. In this study, we novel collected 5 kinds of epigenetic related-genes (ERGs), a total of 2397 genes, including RNAm6a modification related genes, histone modification related genes, DNA methylation modification related genes, RNA binding proteins and transcription factors [5–10][5–10]. In order to explore them in HCC, we analyzed the differentially expressed epigenetic-related genes in HCC by WGCNA analysis and constructed an apparent regulatory network. To optimize the treatment system of HCC, we integrated epigenetic-related genes to build a diagnosis and prognosis model, and compared the signal differences between high-risk group and low-risk group by GSEA and GSVA analysis.

2. Materials And Methods

2.1 Data acquisition and processing

The mRNA expression and patient clinical data were downloaded from GEO database (GSE14520), TCGA database (TCGA-LIHC dataset), GTEx database (GTEX-liver dataset) and ICGC database (ICGC-JP dataset) [1–4]. ERGs consist of m6A-related gene, histone modification-related gene, RNA binding protein, transcription factor and DNA methylase were collected based on the previous literatures and databases [5–10]. Cytoscape software was used to construct the regulatory network of ERGs and target genes with a criteria of the correlation coefficient is greater than 0.7.

2.2 Functional analysis based on the WGCNA

‘WGCNA’ R package used to Weighted correlation network analysis (WGCNA), and a proper soft-threshold was chosen to cluster genes with co-expression similarity to the same module [11]. Clinical data were combined with the above modules to seek to clinical meaningfully genes clusters. We conducted Go and KEGG analysis with a criteria of P value < 0.05 and q value < 0.05 using R package ‘enrichplot’.

2.3 Construction of the diagnosis and prognostic model

Support Vector Machine (SVM) depended on ‘Random Forest’ R package and LASSO regression was conducted in order to screen diagnostic markers. Diagnostic models were built by logistic regression. Univariate Cox regression analysis was conducted to identify Survival-related genes which the criterion was P < 0.001. Lasso-Multivariate Cox regression analysis was used to established prognosis model. Kaplan-Meier survival curves were conducted to compare survival-time differences between high-risk and

low-risk groups. We draw the receiver operating characteristic curve (ROC) and decision curve analysis (DCA) for showing model accuracy [12, 13]. The model stability was shown using calibration curve.

2.4 Functional prediction analysis between high-PERs and low-PERs

Based on the median prognostic epigenetic risk score (PERs) in the above TCGA-LIHC dataset, the samples were divided into high-PERs group and low-PERs group. GSEA_4.0.1 software was used for exploring biological function of PERs for 2 groups rely on Hallmarkes gene set. Gene Set Variation Analysis (GSVA) used 'GSVA' R package and Hallmarkes gene set in order to intersect with the results.

2.5 Independent risk factor analysis

Univariate and multivariate Cox regression analyses was performed to identify independent risk factors from the above PERs and other clinicopathological factors such as TNM stages, age, gender.

2.6 Statistical analysis

SPSS and R software was used for statistical analysis. If there is no special note, p value < 0.05 was considered statistically significant. Two tailed student's test was used for testing the differences of different groups.

3. Results

3.1 Identification hub ERGs in HCC

The analysis flow chart is Fig. 1. We integrated the 2397 ERGs mentioned above. First, based on the combination of TCGA-LIHC and GTEX-liver sequence dataset due to TCGA-LIHC dataset has few samples of normal liver tissue, we used differential gene analysis. The cutoff criterion was FC > 1.5 and FDR < 0.01. Considering the advantages and disadvantages of sequencing and microarray in evaluating gene expression, we further verify the above results in the largest and comprehensive online HCC microarray dataset GSE14520, with a cut-off value of FDR < 0.01 which identified 493 differentially expressed ERGs (Fig. 2A). In order to further obtain clinically significant ERGs in HCC, we applied WGCNA to analyze above genes which get five modules in HCC which among them, the blue, turquoise and yellow modules are closely related to patients' DFI, PFI and OS based on the widely validated TCGA clinical data which collected 353 hub- ERGs (Fig. 2B).

3.2 KEGG and GO analysis, and built epigenetic factors regulatory network

Kyoto Encyclopedia of Gene and Genome (KEGG) and gene ontology (Go) analysis which can identify gene function with high throughput based on previous gene annotation, was performed to explore the function of 353 hub-ERGs. The KEGG pathways were mainly involved in ribosome, spliceosome and RNA transport (Fig. 2C). In addition, the GO analyses showed that biological process (BP) terms were mainly enriched in the regulation of RNA splicing, ribonucleoprotein complex biogenesis, and RNA splicing, cell component (CC) terms were mainly involved in ribosomal subunit, ribosome and spliceosomal complex, as well as molecular function terms were significantly included structural constituent of ribosome, catalytic activity, acting on RNA and ribonuclease activity (Fig. 2D). Furthermore, in order to understand the regulatory targets of ERGs, we identified the differentially expressed genes in HCC according to the previous differential analysis conditions, and constructed a differential apparent regulatory network according to the correlation coefficient greater than 0.7 (Fig. 2E).

3.3 ERGs for diagnostic of HCC

To evaluate the diagnostic value of hub-ERGs, we strictly filter the range of difference analysis, and the cutoff value FC is greater than 2 based on the TCGA-GTEX dataset which acquired 12 ERGs (Fig. 3A). For the accuracy and refinement of the model, we performed support vector machines (SVM) and least absolute shrinkage and selection operator (LSAAO) analysis, and six genes were selected from the intersection to construct a diagnostic epigenetics risk score (dERS) model using multivariate logistic regression analysis. The model formula is =(TOP2A*2.631)+(EEF1A2*0.519)+(SNRPB*3.975)+(RNASE4*-0.482)+(NR4A1*-0.631)+(RPS29*-1.112) (Fig. 3B-D). In TCGA-GTEX dataset, the receiver operating characteristic curve (ROC) showed excellent AUC value (0.996), specificity (0.971) and sensitivity (0.973) (Fig. 3E). We also verified the above diagnostic model in an independent dataset ICGC-JP (Fig. 3F).

3.4 ERGs for prognostic of HCC

In order to explore the value of the above hub ERGs in HCC, we conducted a univariate Cox regression analysis to identify 23 survival-related genes ($P < 0.001$). Next we established a prognostic epigenetic risk score (PERs) which consist of ten genes by LASSO-COX regression analysis (Fig. 4A). PERs was calculated for patients in TCGA-LIHC dataset and divided patients into high-PERs and low-PIRs groups according the 50% cutoff point. The high-PIRs group has shorter survival than the low-PERs group (Fig. 4B). The AUC value for 1-, 3- and 5- year OS were 0.861, 0.765, 0.698 (Fig. 4C). In addition, we performed also Decision Curve Analysis (DCA) which showed that PERs can bring good benefits to patients (Fig. 4D-F). Calibration curve showed also excellent stability (Fig. 4G). The above survival-related model was also verified in an independent dataset, ICGC-JP dataset which the result was consistent with TCGA-LIHC dataset (Fig. 4H.I).

3.5 Clinical characteristics analysis for PERs

For the convenience of clinical practicability, the nomogram was drawn to predict the 1-year, 3-year, 5-year survival (Fig. 5A). Through univariate and multivariate regression analysis, we identified PERs as an independent prognostic factor for HCC (Fig. 5B.C). The difference of clinical characteristics between high-

PERs and low-PERs group were analyzed in the above all datasets which the accuracy of the model is further verified from the inside of the dataset. Based on the TCGA-LIHC dataset, the differential signaling pathways between high-PERs and low-PERs groups by using the GSEA and GSVA algorithms to get the intersection (Fig. 5D.E.S1) [21, 22], which offer a understand for the difference mechanism between high-PERs and low-PERs groups.

3.6 Pan cancer analysis

One of the advantages of bioinformatics analysis is that it can obtain a variety of tumor data for analysis. In order to explore the applicability of these prognostic mod-els in other tumors, we collected 14 tumor (BLCA, BRCA, COAD, HNSC, KIRC, KIRP, LGG, LUAD, LUSC, OV, PRAD, SKCM, STAD, UCEC) data with sample size greater than 300 in TCGA database. Through univariate Cox regression analysis and KM survival curve, we found that only a few tumors such as KIRP were suitable for the above survival model. The results showed the specificity of the model in liver cancer.

4. Discussion

Although the level of medical diagnosis and treatment has been improved in recent years, the accuracy of diagnosis and survival rate of prognosis of HCC is still poor [23]. It is one of the hot research directions for future research to identify clinically significant genes, predict their functions and explore their prognostic value based on bioinformatics. At present, there is lack of biomarker with effective and high accuracy for diagnostic and prognostic for HCC based on biomolecule [24]. However, in the past, bioinformatics research often focused on single database or only focused on prognostic value, which had some limitations. In recent decades, scientific workers have realized many aspects about epigenetic modification regulating gene expression to interfere with tumor progression. DNA level methylation, RNAm6a methylation, and histone-related modification and so on are the hot spots of tumor research. Previous studies mainly focused on the impact of single epigenetic related genes on tumor prognosis and function. Based on bioinformatics analysis, the above genes can be widely recognized for the diagnosis and prognosis of liver cancer, and functional prediction can be carried out, which can provide help for the later experimental research. In addition, the regulation of gene expression by epigenetic regulators is closely related to transcription factors and RNA binding proteins, thus we collected 5 kinds of ERGs including RNAm6a modification related genes, histone modification related genes, DNA methylation modification related genes, RNA binding proteins and transcription factors. And based on these genes, we have successfully constructed a stable and reliable diagnosis and prognosis model for HCC which was verified by an independent data-set.

In this charter, in order to ensure the accuracy of the analysis, we identified 493 differentially expressed ERGs according to the combination analysis of three on-line datasets. WGCNA analysis can cluster disease genes according to their intrinsic expression correlation, and identify key gene modules by combining with clinical phenotypes [11]. TCGA database is the most authoritative tumor database, including a variety of histological data and clinical information. OS, DFI and PFI are recommended for survival related analysis of TCGA-LIHC dataset. Thus, using WGCNA analysis, a total of 5 co-expression

modules were separated and we choose the blue, turquoise and yellow modules for next study because they are closely related to survival which acquired 353 hub-IRGs. KEGG and Go analysis can fully understand gene set function, compared with individual study on gene function [25, 26][25, 26]. We performed KEGG and GO analysis, and the results showed that the above genes were mainly enrichment in RNA splicing, ribonucleoprotein complex biogenesis and etc [27, 28]. Previous studies have shown that these physiological activities are closely related to tumorigenesis. Furthermore, we also constructed the regulatory network of ERGs and target genes, which can preliminarily and comprehensively speculate the regulatory relationship of these genes.

Next, we investigated the diagnostic and prognostic value of the above hub-ERGs. Support vector machine can not only be used to build classification model, but also be used to screen important variables in order to build a simpler and more accurate model [29]. Based on the TCGA-GTEX dataset, we found important variables through SVM and LASSO analysis, and established a diagnostic model consist of six ERGs for HCC by logistic regression and which specificity 0.971, sensitivity 0.973. We also validated the above diagnostic model in an independent dataset, ICGC-LIHC dataset which specificity 0.882, sensitivity 0.871. There are a few differences which may be attributing to the difference of the source of the included groups. The results showed the reliability and stability of the diagnosis model. Furthermore, we established a prognostic risk model including ten ERGs by lasso-Cox analysis. The AUC value for 1-, 3- and 5- year OS were 0.861, 0.765, 0.698 in TCGA-LIHC dataset and 0.846, 0.860, 0.829 in ICGC-JP dataset. These results support the reliability of the prognostic model. For the reliability of function prediction, we also carried out two methods (GSEA and GSVA analysis) to identify different signal pathways between high-risk group and low-risk group. The results showed that mTOR, PI3K, Wnt and other signaling pathways were enriched in high-risk group which it may be a potential therapeutic target [30, 31].

There are significant metabolic abnormalities of fatty acids in tumors. The significant increase of lipid droplets in tumor cells can promote the transformation of tumor mesenchymal. Reducing the accumulation of fatty acids in cells can reduce the migration and invasion of tumor cells [32]. Extracellular fatty acid metabolites also participate in the formation of pro-tumor microenvironment [33]. Wnt signaling pathway is one of the most important pathways in tumor EMT [34]. Previous studies have shown that tumor cell EMT is associated with poor prognosis and immunosuppression [35]. MTOR signaling pathway is the switch of cell energy metabolism [36]. The progression of solid tumors is often accompanied by metabolic reprogramming and stress [37]. MTOR mediates tumor proliferation, migration, invasion and autophagy by regulating tumor energy metabolism [36]. These studies suggest that the differential expression of multiple signaling pathways in high-PERs and low-PERs groups may be involved in poor prognosis of HCC and may be used as potential therapeutic targets which also confirmed the reliability of our prognosis model from within the gene set.

5. Conclusion

we first comprehensively analyzed the ERGs in liver cancer. Compared with the previous bioinformatics studies which only construct the prognosis model, we discuss the influence of gene set on diagnosis and prognosis. These results can offer a widely understand for HCC development which can provide help for future experimental research. More importantly, we novel constructed two biomarkers to guide the management of liver cancer. Although the major limitation of this study is lack of verification of experimental data, we tried our best to ensure the reliability of data through multiple data sets or multiple algorithms. All in all, this study remains helpful and valuable for the treatment and diagnosis of HCC.

Abbreviations

HCC: Hepatocellular carcinoma; TCGA: The Cancer Genome Atlas; GEO: Gene expression omnibus; KEGG: Kyoto Encyclopedia of Genes and Genomes; GSEA: Gene Set Enrichment Analyses; GSVA: Gene Set Variation Analysis; ICGC: International Cancer Genome Consortium

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication:

Not applicable

Availability of data and materials:

All data of this study can be obtained from TCGA, GTEX, ICGC and geo databases

Competing interest:

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

Funding:

This project was supported by the National Science Foundation of China (No. 81170442, 81470899, 81702357).

Authors' contributions:

WDG, QS and **LZJ** performed research and wrote the first draft. **WDG and QS** collected and analyzed the data. All authors contributed to the design and interpretation of the study and to further drafts. **LZJ** is the guarantor.

Acknowledgements:

None.

References

1. Barrett T, Edgar R. Reannotation of array probes at NCBI's GEO database. Nat Methods 2008; 5: 117.
2. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013; 45: 580-585.
3. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell 2018; 173: 400-416.e11.
4. Cortés-Ciriano I, Lee JJ, Xi R, Jain D, Jung YL, Yang L, et al. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. Nat Genet 2020; 52: 331-341.
5. Liu T, Ortiz JA, Taing L, Meyer CA, Lee B, Zhang Y, et al. Cistrome: an integrative platform for transcriptional regulation studies. Genome Biol 2011; 12: R83.
6. Williams K, Christensen J, Helin K. DNA methylation: TET proteins-guardians of CpG islands. EMBO Rep 2011; 13: 28-35.
7. Khare SP, Habib F, Sharma R, Gadewal N, Gupta S, Galande S. HIstome--a relational knowledgebase of human histone proteins and histone modifying enzymes. Nucleic Acids Res 2012; 40: D337-342.
8. Gujar H, Weisenberger DJ, Liang G. The Roles of Human DNA Methyltransferases and Their Isoforms in Shaping the Epigenome. Genes (Basel) 2019; 10:
9. Chen M, Wong CM. The emerging roles of N6-methyladenosine (m6A) deregulation in liver carcinogenesis. Mol Cancer 2020; 19: 44.
10. Li K, Guo ZW, Zhai XM, Yang XX, Wu YS, Liu TC. RBPTD: a database of cancer-related RNA-binding proteins in humans. Database (Oxford) 2020; 2020:
11. Zhao W, Langfelder P, Fuller T, Dong J, Li A, Hovarth S. Weighted gene coexpression network analysis: state of the art. J Biopharm Stat 2010; 20: 281-300.
12. Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. Phys Med Biol 2018; 63: 07TR01.
13. Van Calster B, Wynants L, Verbeek J, Verbakel JY, Christodoulou E, Vickers AJ, et al. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. Eur Urol 2018; 74: 796-804.
14. Tang A, Hallouch O, Chernyak V, Kamaya A, Sirlin CB. Epidemiology of hepatocellular carcinoma: target population for surveillance and diagnosis. Abdom Radiol (NY) 2018; 43: 13-25.
15. Yang JD, Hainaut P, Gores GJ, Amadou A, Plymoth A, Roberts LR. A global view of hepatocellular carcinoma: trends, risk, prevention and management. Nat Rev Gastroenterol Hepatol 2019; 16: 589-604.

16. Russo FP, Imondi A, Lynch EN, Farinati F. When and how should we perform a biopsy for HCC in patients with liver cirrhosis in 2018? A review. Dig Liver Dis 2018; 50: 640-646.
17. Yang Y, Liu C, Qi L, Zhao T, Feng Y, Ai X, et al. Diagnosis of Pre-HCC Disease by Hepatobiliary-Specific Contrast-Enhanced Magnetic Resonance Imaging: A Review. Dig Dis Sci 2020; 65: 2492-2502.
18. Ayuso C, Rimola J, Vilana R, Burrel M, Darnell A, García-Criado Á, et al. Diagnosis and staging of hepatocellular carcinoma (HCC): current guidelines. Eur J Radiol 2018; 101: 72-81.
19. Nacev BA, Jones KB, Intlekofer AM, Yu J, Allis CD, Tap WD, et al. The epigenomics of sarcoma. Nat Rev Cancer 2020; 20: 608-623.
20. Lu Y, Chan YT, Tan HY, Li S, Wang N, Feng Y. Epigenetic regulation in human cancer: the potential role of epi-drug in cancer therapy. Mol Cancer 2020; 19: 79.
21. Lappano R, Talia M, Cirillo F, Rigiracciolo DC, Scordamaglia D, Guzzi R, et al. The IL1β-IL1R signaling is involved in the stimulatory effects triggered by hypoxia in breast cancer cells and cancer-associated fibroblasts (CAFs). J Exp Clin Cancer Res 2020; 39: 153.
22. Wang C, Yang Y, Yin L, Wei N, Hong T, Sun Z, et al. Novel Potential Biomarkers Associated With Epithelial to Mesenchymal Transition and Bladder Cancer Prognosis Identified by Integrated Bioinformatic Analysis. Front Oncol 2020; 10: 931.
23. Huang A, Yang XR, Chung WY, Dennison AR, Zhou J. Targeted therapy for hepatocellular carcinoma. Signal Transduct Target Ther 2020; 5: 146.
24. Baudi I, Inoue T, Tanaka Y. Novel Biomarkers of Hepatitis B and Hepatocellular Carcinoma: Clinical Significance of HBcrAg and M2BPGi. Int J Mol Sci 2020; 21:
25. Chen L, Zhang YH, Lu G, Huang T, Cai YD. Analysis of cancer-related lncRNAs using gene ontology and KEGG pathways. Artif Intell Med 2017; 76: 27-36.
26. Zheng F, Zhang W, Chu X, Dai Y, Li J, Zhao H, et al. Genome sequencing of strain Cellulosimicrobium sp. TH-20 with ginseng biotransformation ability. 3 Biotech 2017; 7: 237.
27. Domínguez-Sánchez MS, Sáez C, Japón MA, Aguilera A, Luna R. Differential expression of THOC1 and ALY mRNP biogenesis/export factors in human cancers. BMC Cancer 2011; 11: 77.
28. Kawamura N, Nimura K, Saga K, Ishibashi A, Kitamura K, Nagano H, et al. SF3B2-Mediated RNA Splicing Drives Human Prostate Cancer Progression. Cancer Res 2019; 79: 5204-5217.
29. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. Cancer Genomics Proteomics 2018; 15: 41-51.
30. Li Y, Shen J, Cheng CS, Gao H, Zhao J, Chen L. Overexpression of pyruvate dehydrogenase phosphatase 1 promotes the progression of pancreatic adenocarcinoma by regulating energy-related AMPK/mTOR signaling. Cell Biosci 2020; 10: 95.
31. Tang B, Zhang Y, Wang W, Qi G, Shimamoto F. PARP6 suppresses the proliferation and metastasis of hepatocellular carcinoma by degrading XRCC6 to regulate the Wnt/β-catenin pathway. Am J Cancer Res 2020; 10: 2100-2113.

32. Corbet C, Bastien E, Santiago de Jesus JP, Dierge E, Martherus R, Vander Linden C, et al. TGFβ2-induced formation of lipid droplets supports acidosis-driven EMT and the metastatic spreading of cancer cells. Nat Commun 2020; 11: 454.
33. Coutzac C, Jouniaux JM, Paci A, Schmidt J, Mallardo D, Seck A, et al. Systemic short chain fatty acids limit antitumor effect of CTLA-4 blockade in hosts with cancer. Nat Commun 2020; 11: 2168.
34. Pan J, Fang S, Tian H, Zhou C, Zhao X, Tian H, et al. lncRNA JPX/miR-33a-5p/Twist1 axis regulates tumorigenesis and metastasis of lung cancer by activating Wnt/β-catenin signaling. Mol Cancer 2020; 19: 9.
35. Kudo-Saito C, Shirako H, Takeuchi T, Kawakami Y. Cancer metastasis is accelerated through immunosuppression during Snail-induced EMT of cancer cells. Cancer Cell 2009; 15: 195-206.
36. Kim J, Guan KL. mTOR as a central hub of nutrient signalling and cell growth. Nat Cell Biol 2019; 21: 63-71.
37. Fendt SM, Frezza C, Erez A. Targeting Metabolic Plasticity and Flexibility Dynamics for Cancer Therapy. Cancer Discov 2020; 10: 1797-1807.

Table

Table 1

Clinical characteristics between high and low risk group.

TCGA-LIHC			
Characters	Low risk	High risk	P value
Gender			.201
Female	52	51	
Male	114	113	
Age(years)			.964
≤ 55	52	62	
> 55	114	101	
Grade			< .001
G1 + G2	119	86	
G3 + G4	43	76	
TNM Stage			.002
I-II	126	103	
III-IV	28	51	
T Stage			.007
T1 + T2	132	111	
T3 + T4	31	52	
N Stage			.561
N0	115	114	
N1	1	2	
M Stage			.561
M0	119	118	
M1	1	2	

Figures

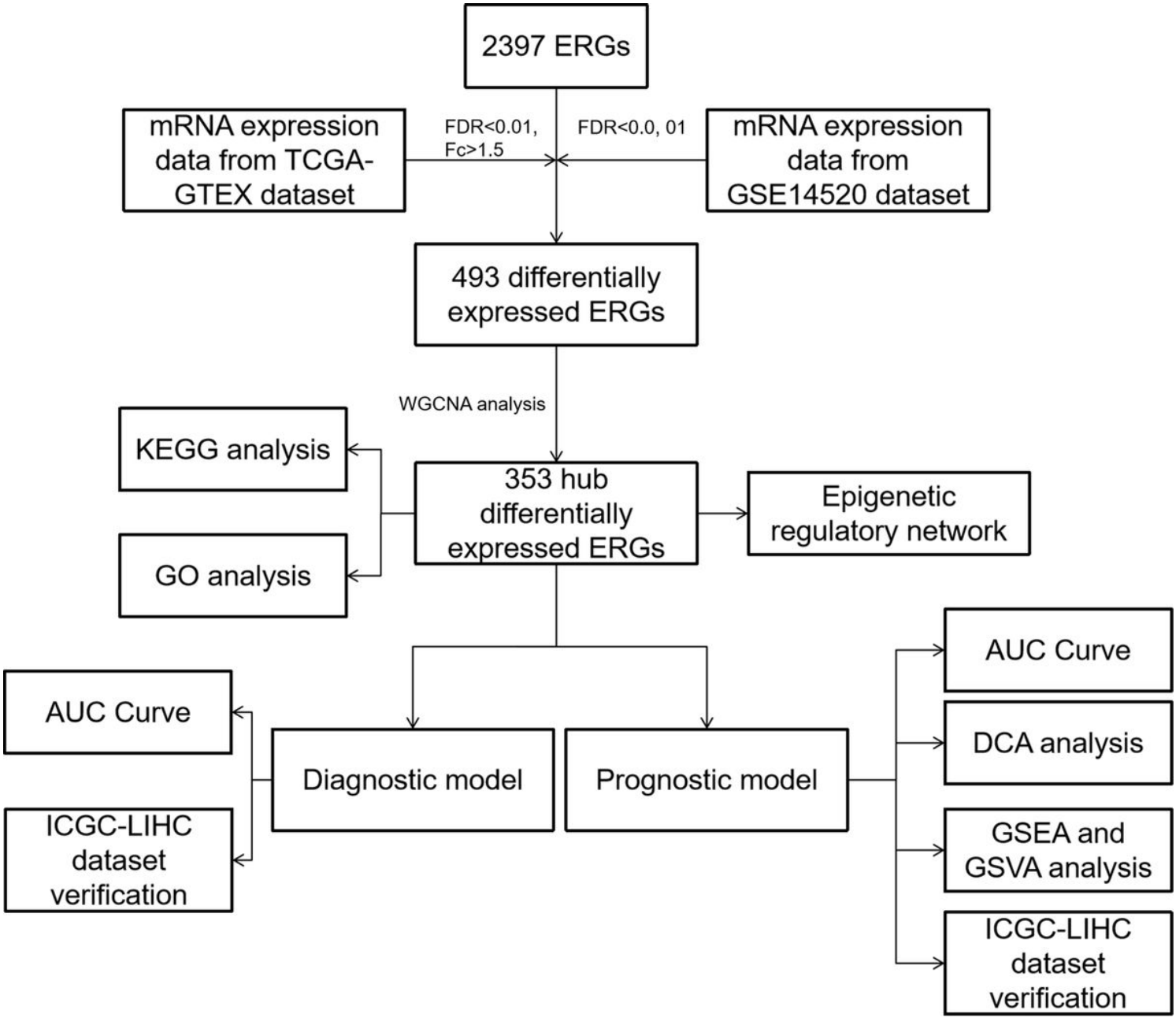

Figure 1

Analysis flow chart of the study.

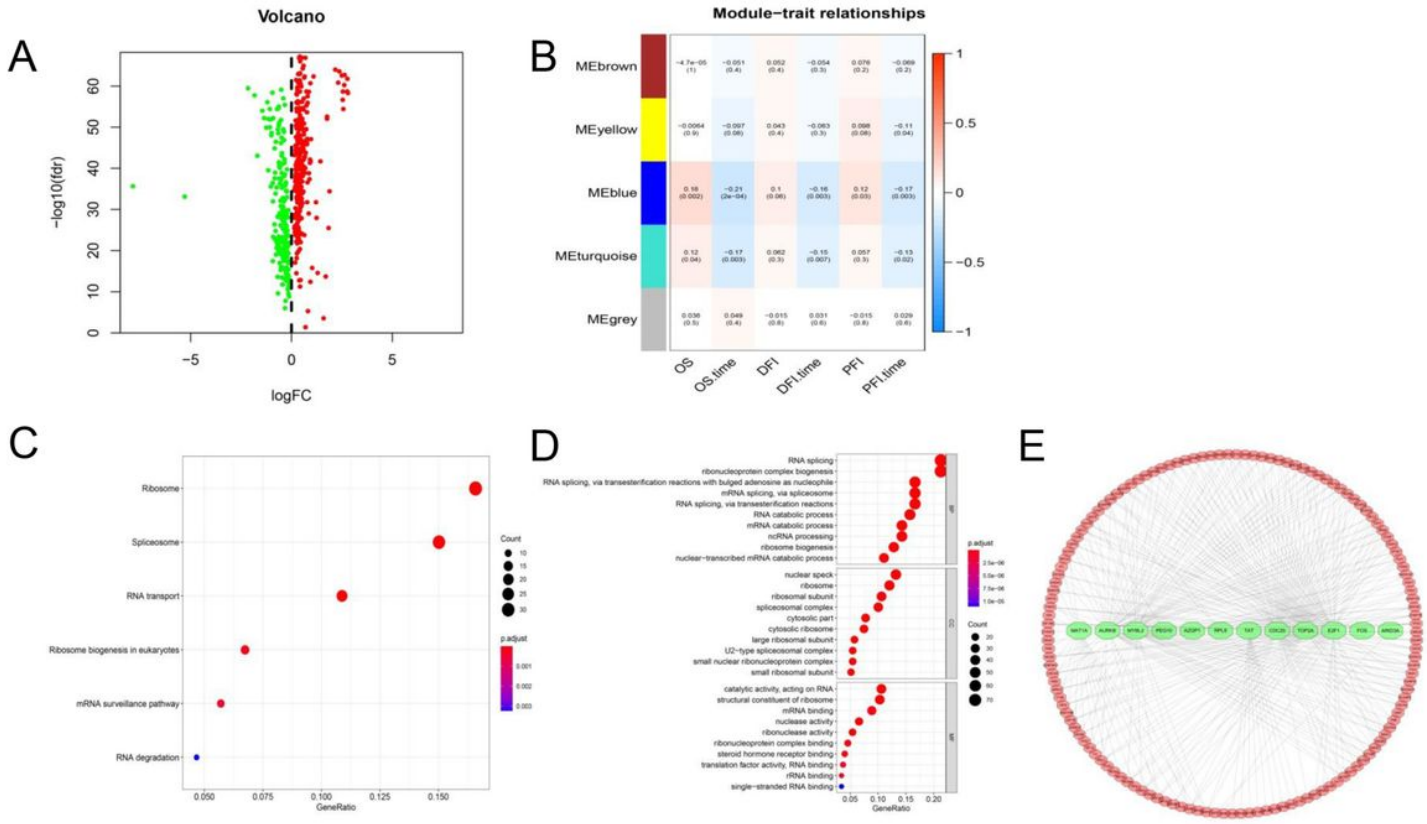

Figure 2

Systematic analysis of epigenetic related genes. (A) Volcano maps for 493 differentially expressed ERGs based on GTEX-TCGA dataset. (B) WGCNA analyzed the 493 ERGs combined with TCGA clinical information. (C) KEGG analysis of 353 hub- ERGs. (D) GO analysis of 353 hub- ERGs. (E) The apparent regulatory network based on the correlation coefficient greater than 0.7.

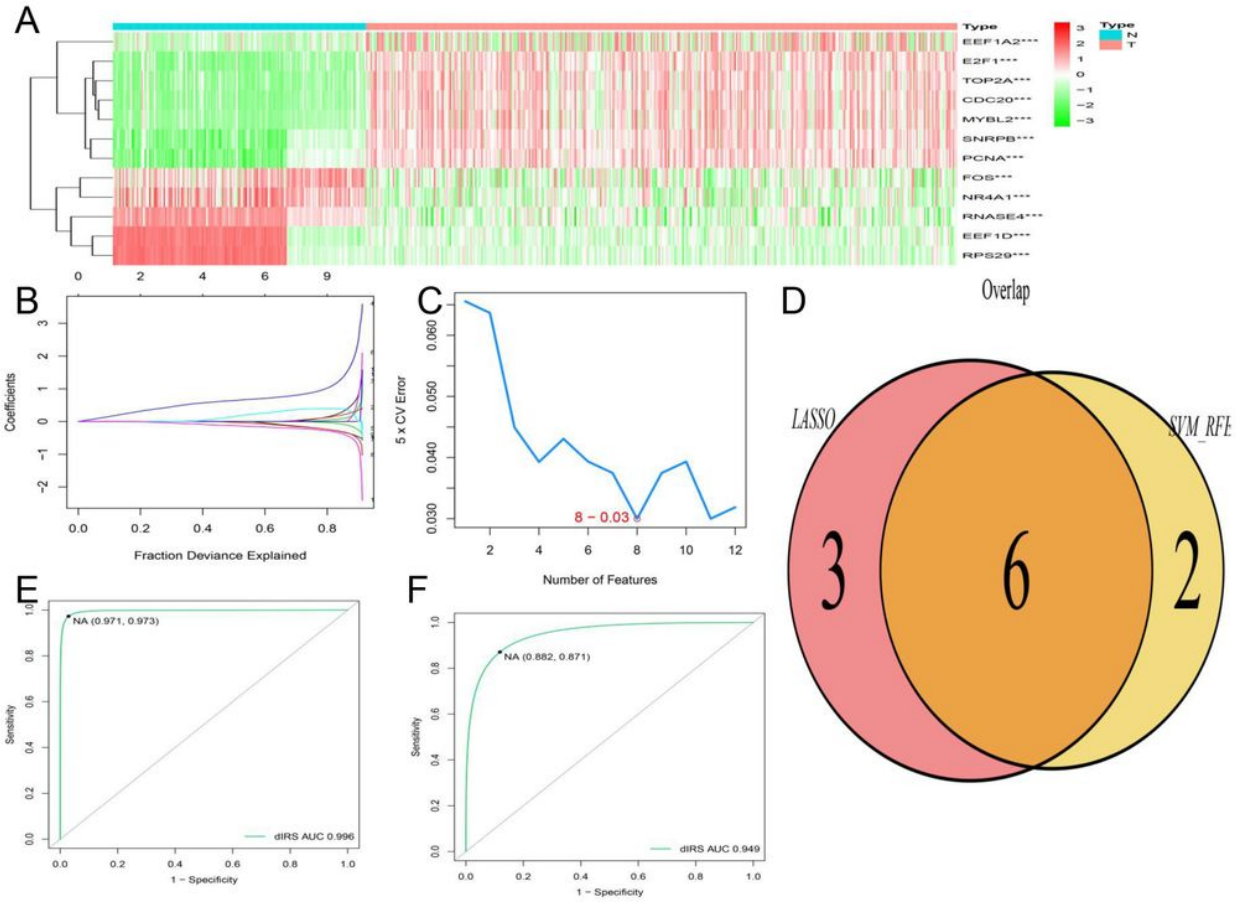

Figure 3

Build and validate diagnostic signal for HCC. (A) 12 significantly different expression ERGs based on the GTEX-TCGA dataset. The cutoff value FC is greater than 2. (B) Lasso regression analysis of the above 12 ERGs. (C) Support vector machine analysis of the above 12 ERGs. (D) The characteristic value is based on the intersection of SVM and LASSO analysis. (E) The ROC curve of the diagnostic signal to show the sensitivity and specificity based on the GTEX-TCGA dataset. (F) The ROC curve of the diagnostic signal to show the sensitivity and specificity based on the ICGC-LIHC dataset.

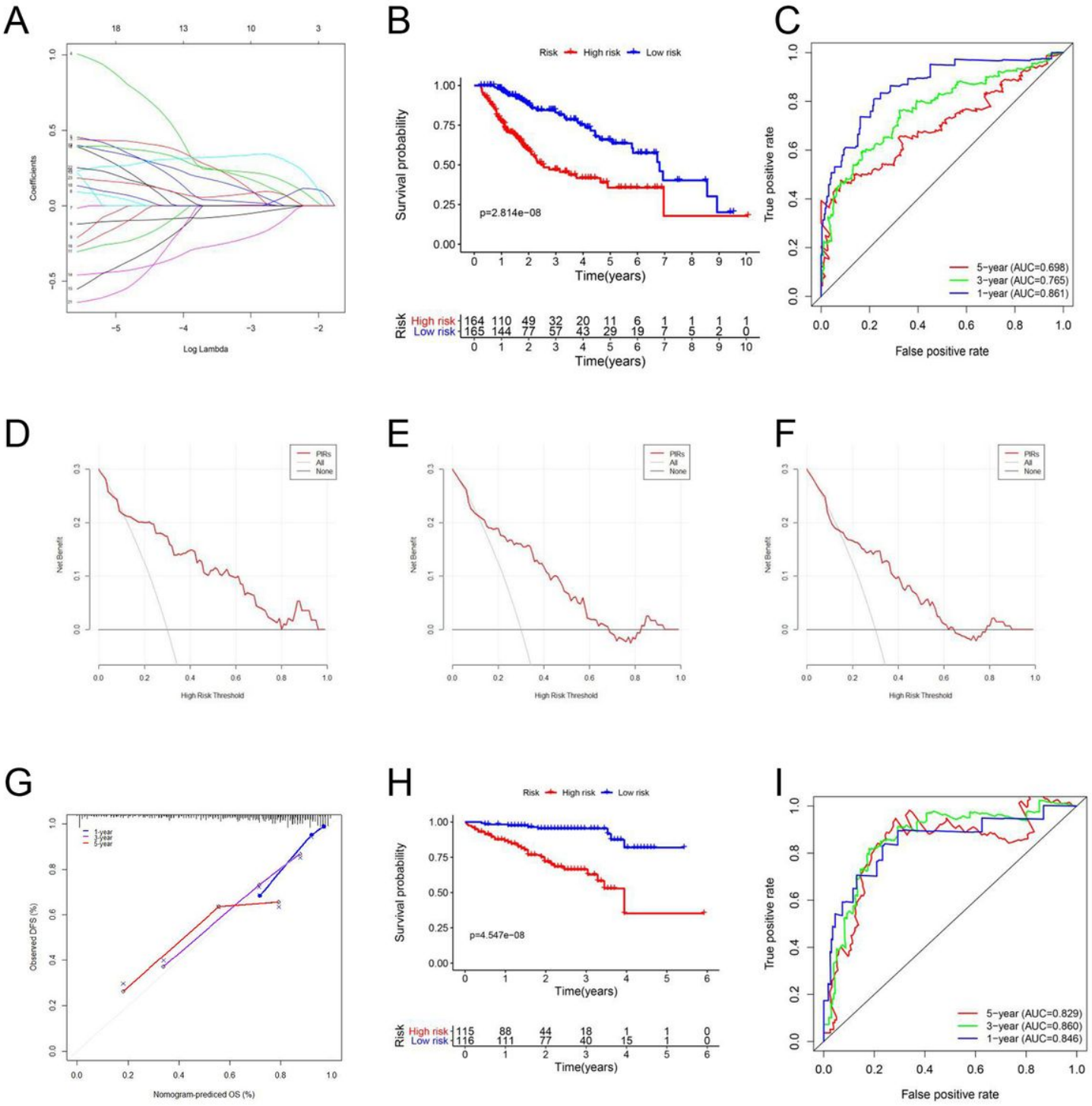

Figure 4

Build and validate prognostic signal for HCC. (A) Lasso regression analysis to identify the characteristic value to constructed diagnostic signal. (B) Survival analysis be-tween high-risk and low-risk group. (C) The area under the ROC curve (AUC) of PIRs for 1-, 3- and 5 –year OS in TCGA dataset. (D) The DCA curve of PERs for 1-, 3- and 5 –year in TCGA dataset. (E) The calibration curve of PERs for 1-, 3- and 5 –year in TCGA dataset. (H) Survival analysis between high-risk and low-risk group in ICGC-LIHC dataset. (I) The AUC of PIRs for 1-, 3- and 5 –year OS in TCGA dataset.

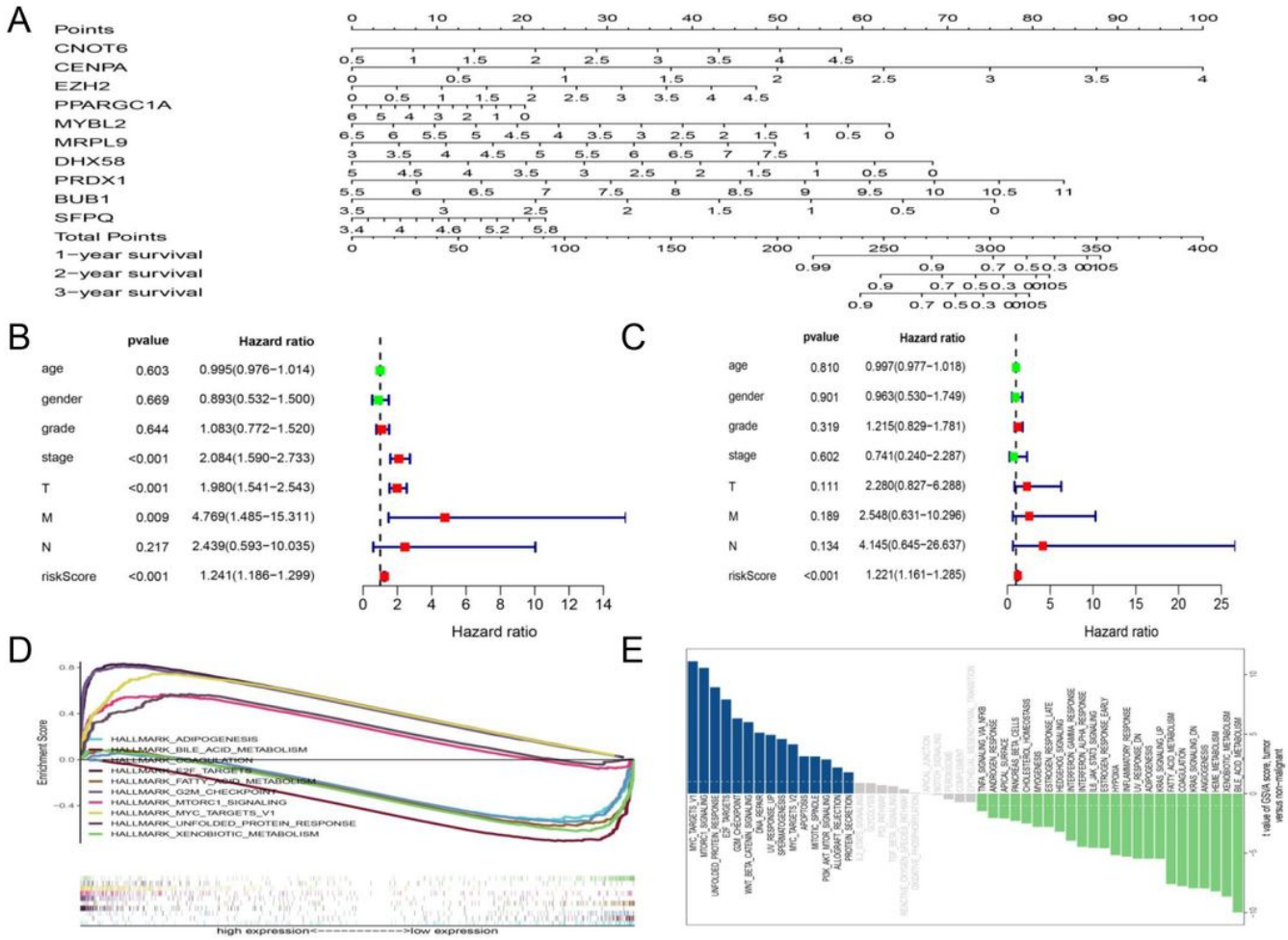

Figure 5

Functional analysis between high-PERs and low-PERs group. (A) The nomograms for predicting 1-, 3-, and 5-year survival rate in TCGA dataset. (B) Univariate Cox regression analysis for PERs in TCGA dataset. (C) multivariate Cox regression analysis for PERs in TCGA dataset. (D) GSEA analysis between high-PERs and low-PERs group in TCGA dataset. (E) GSVA analysis between high-PERs and low-PERs group in TCGA dataset.

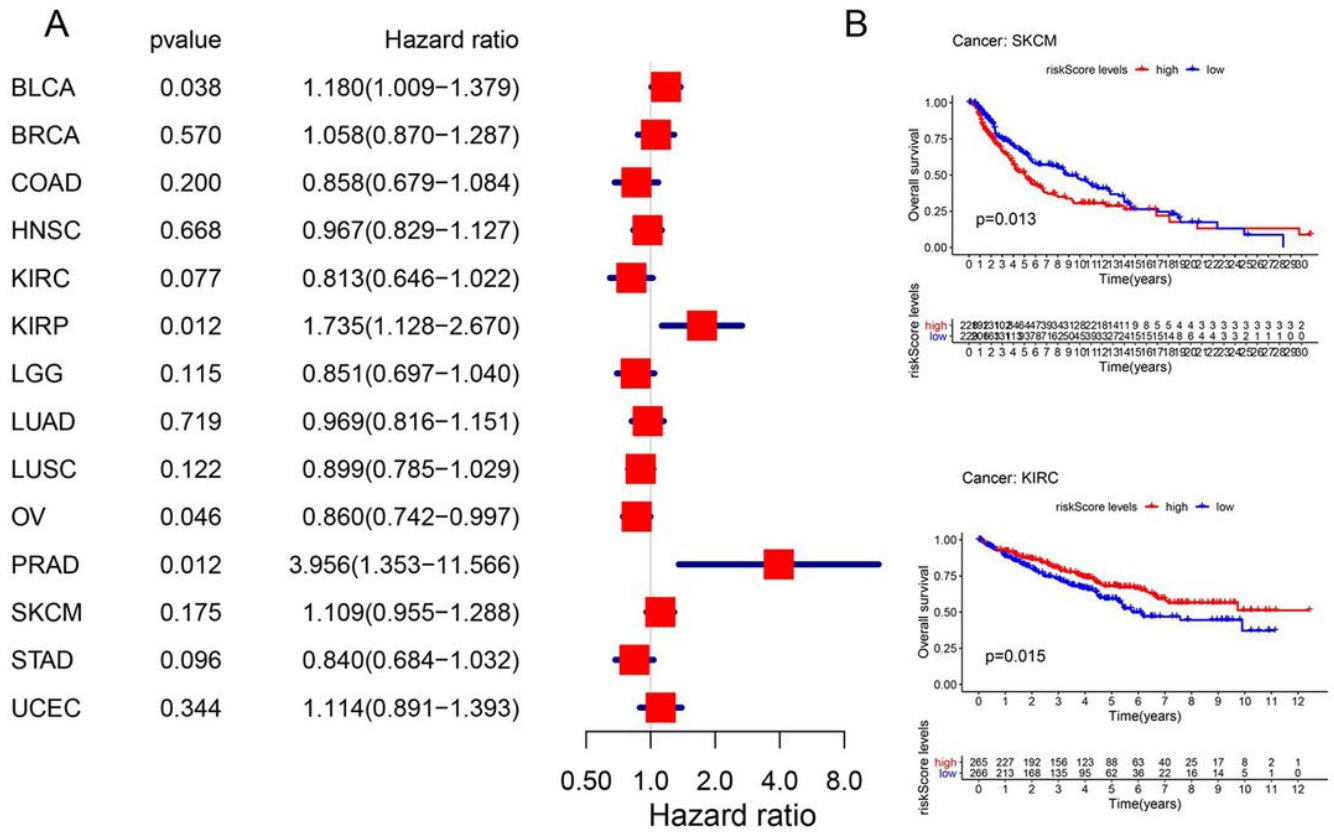

Figure 6

Analysis of Pan cancer using TCGA database. (A) Univariate Cox regression analysis for 14 kinds of tumors. (B) Km survival curve analysis between high-PERs and low-PERs group which P value <0.05.

Supplementary Files

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

- supplementary.doc