Identification and Validation of a Reliable Prognostic Eleven-Genes Signature for Hepatocellular Carcinoma

Yuliang Li (liyuliang688@sina.com)
Southwest Hospital, Third Military Medical University (Army Medical University) https://orcid.org/0000-0003-1603-8479

Zhirui Liu
Third Military Medical University Southwest Hospital

Qian Wang
Third Military Medical University Southwest Hospital

Primary research

Keywords: Hepatocellular carcinoma, Differentially expressed genes, Robust rank aggregation, Overall survival, Risk index, Nomogram

Posted Date: September 2nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-70353/v1

Abstract

Background: Hepatocellular carcinoma (HCC) is a common malignant tumor with high mortality and mortality. Although advances in early diagnosis, disease management and treatment of HCC, the outcomes remain unsatisfactory. This study aimed to identify the reliable prognostic biomarkers based integrated bioinformatics analysis to predict and improve the survival of HCC patients.

Methods: The gene expression or transcriptome profiles and survival of HCC were acquired from the Gene Expression Omnibus database (GEO) and the Cancer Genome Atlas (TCGA) database. Differentially expressed genes (DEGs) were screened out by the limma or edgeR package in the R software. Univariate, LASSO and multivariate Cox regression analyses were conducted to explore survival-related signature. Subsequently, a prognostic model and nomogram composed of prognostic signature were constructed for assessing overall survival (OS). Kaplan-Meier analysis, receiver operating characteristic (ROC) curve and stratified analysis were performed to confirm the prognostic performance of the prognostic model.

Results: Compared with nontumor samples, 451 reliable DEGs were identified using the robust rank aggregation and overlap validation. Eleven survival-related DEGs were selected for the construction of a risk evaluation model, which could efficiently distinguish high-risk patients from low-risk patients and even be feasible in the subgroups of stages and age. Further analyses suggested the positive and independent prognostic performance of the model compared to other clinical characteristics ($P < 0.05$, ROC > 0.7). Finally, a prognostic nomogram composed of the model was constructed for assessing the overall survival, and Harrell's concordance index and calibration curves demonstrated its efficient predictive performance.

Conclusion: The predictive model and nomogram will contribute directly to further clinical applications in the individualized survival prediction, the improvement of treatment strategies and more accurate management for patients with HCC.

Background

Hepatocellular carcinoma (HCC) is one of the lethal cancers and the second-leading cause of cancer-related mortality [1]. In the last decades, advances in the treatment of unresectable liver cancer and the earlier diagnoses have made it possible for patients with HCC to live longer than in the past. Still, HCC is a malignant tumor with a worse prognosis due to its difficulty in early diagnosis, metastasis and dissemination. The 1-year survival rate of these patients is 44% [2, 3], and the 5-year survival rate is only 18% [4]. Poor outcomes may be caused by the rapid progression, early metastasis, and lack of typical clinical symptoms or suitable biomarkers for HCC. Thus, the early evaluation of individual outcomes is extremely urgent for improving HCC treatment and prognosis. Frequently, clinicopathological characteristics such as the pathologic stage, histologic grade, vascular invasion and tumor-node-metastasis status have been utilized to assess HCC outcomes in practice. Patients with higher-stage, vascular invasion or metastasis have a worse overall survival (OS). However, the assessment is often based on pathological biopsy; It is difficult to predict the progression due to invasive surgery. Therefore,

reliable prognostic biomarkers are needed to predict the outcome of treatment and guide personalized treatment of HCC.

Advances in microarrays, high-throughput sequencing techniques and bioinformatics have demonstrated that prognostic gene signatures be capable of predicting HCC patients' OS. Unfortunately, bias, errors as well as noises could be obtained due to the limited sample, single technological platforms or inappropriate analysis method. Integrated analysis of different microarrays and/or high-throughput sequencing data using effective analysis methods could obtain reliably differentially expressed genes (DEGs) and further identify potential molecular biomarkers. Recently, the robust rank aggregation (RRA) has been specifically designed to compare several ranked gene lists and identified commonly overlapping genes to reduce or get rid of these mistakes [5]. Numerous studies adopted the RRA method to select and filter differentially expressed microRNA or mRNA profiles on multiple datasets [6–8]. However, there is a little study on identifying DEGs using the RRA method in HCC, especially on screening prognostic gene signatures.

In this study, 451 robust DEGs were identified by an integrated analysis of the data from the GEO and TCGA database using the RRA and overlap validation. Among them, eleven survival-related genes (AURKB, CDCA3, CDCA8, CENPA, ECT2, KIF20A, NEK2, PRC1, SPC25, TOP2A and TPX2) were selected by univariate, least absolute shrinkage and selection operator (LASSO) and multivariate Cox regression analysis. Based on their expression level and survival data from the TCGA HCC cohort, a prognostic risk model was proposed. The model was positively validated on the subgroup (age and pathologic stage) and the entire cohort ($P < 0.05$, ROC > 0.7). Subsequently, a user-friendly nomogram incorporating the signature and age was proposed for further potential application in clinical work. In general, the signature and nomogram may accurately evaluate the overall survival of HCC and will contribute directly to the individualized treatment and management for patients with HCC.

Materials And Methods

Data set collection and data processing

The gene expression microarray profiles of HCC were filtered and downloaded from the NCBI-GEO database (https://www.ncbi.nlm.nih.gov/geo/). To reduce the influence of individual heterogeneity and obtain more reliable results, those datasets with adequate sample size as well as paired tissues were screened and selected. The detailed screening criteria were as follows: (1) these data were derived from liver tumor tissues (TT) and adjacent nontumor tissues (NT) of HCC patients; (2) datasets were raw or standardized; (3) the sample sizes of NT and TT both were more than 35; (4) datasets were published later than 2010. A total of ten datasets were finally selected and their detailed information was shown in Additional file 1: Table S1. Meanwhile, transcriptome profiles with count-per-million calculated on an Illumina HiSeq RNA-seq platform and survival of HCC from the TCGA database (https://portal.gdc.cancer.gov/) were downloaded, containing 50 adjacent nontumor liver tissues and 374 HCC tissues. When multiple probes corresponded to one defined gene, the average expression level was

regarded as its final expression. The quantile-normalization method was applied for normalizing gene expression intensities.

Identification of reliable DEGs

The open statistical R software (version 3.6.3, https://r-project.org/) and Bioconductor packages (http://www.bioconductor.org/) were utilized to excavate and screen DEGs between NT and TT. The original series matrix files of datasets and the annotation documents of platforms were downloaded, and the background correction and normalization were carried out by Perl software (version 5.26.3, https://www.perl.org/). Afterward, the identification DEGs between NT and TT in each GEO dataset was performed using the "*limma*" package with the cut-off criteria of the adjusted *P*-value < 0.05 and |log 2 fold-change (logFC)| > 1. To obtain robust DEGs of these GEO datasets, the RRA method was employed. This method employs a probabilistic and nonparametric-based model for aggregation and is capable of identifying genes that were ranked consistently better than expected by chance and robust to outliers, noise and errors of results with the "*RobustRankAggreg*" package [5]. First, genes were ordered by their fold-change value. Then, the aggregation based on the ranks of genes in different GEO datasets was carried out. Finally, to reduce and avoid false-positive results, the *P*-value was subjected to Bonferroni's correction. Using the core algorithm of the "*RobustRankAggreg*" package, the screening criteria of |logFC| > 1.0 and corrected *P*-value < 0.05 were regards as statistically significant for the robust DEGs.

The transcriptome profile of HCC was analyzed by the "*edgeR*" package, and the data analysis was consistent with that of GEO datasets except for the filtering approach in low expression genes. To obtain reliable results, DEGs held in common between GEO and TCGA were filtered out through Venn diagram analysis.

PPI network construction, hub gene selection and analysis

First of all, a functional protein association network was constructed in the STRING database (version 11.0, https://string-db.org/). The parameter of interactions among DEGs was set the highest confidence > 0.9. Then, Cytoscape software (version 3.7.2) was used to visualize and analyze the interaction network. The cytoHubba plug-in of Cytoscape was utilized to explore vital nodes and sub-networks in a given network using different built-in topological algorithms. Any overlap in the top list of genes from different ranking algorithms will be defined as hub genes. Subsequently, the expression and survival analysis of these hub genes were separately examined in Gene Expression Profiling Interactive Analysis (GEPIA, http://gepia.cancer-pku.cn/). At the cut-off criteria of *P*-value < 0.05 and |log 2 fold-change (logFC)| > 1, hub genes were included in subsequent analyses.

GO and KEGG pathway enrichment analysis

To investigate and understand the biological functional roles of these hub genes, a user-friendly bioinformatics analysis tool, FunRich (version 3.1.3, http://www.funrich.org/) was applied for the biological process (BP), cellular component (CC), and molecular function (MF) of Gene Ontology (GO)

and KEGG pathway enrichment analysis. The screening criteria of *P*-value < 0.05 was considered as statistical significance.

Construction of the risk score system

Univariate, LASSO and multivariate Cox proportional regression analyses were utilized to explore the correlation between the expression level of each hub gene and HCC patient survival. The genes with $P < 0.05$ were considered statistically significant in the univariate Cox regression analysis and included in subsequent analyses. Next, LASSO-penalized Cox regression analysis was applied to narrow the range of target genes in the selected panel using 10-fold cross-validation by the "*glmnet*" package in R software. Then, a stepwise multivariate Cox regression analysis was performed to evaluate the contribution of a single gene as an independent prognostic factor and screen the genes tightly associated with survival for the best-fit model. Finally, a prognostic gene signature was established based on a linear combination of the expression levels of the best-fit OS-related genes multiplied with their corresponding regression coefficients (β) derived from the stepwise multivariate Cox regression. The prognostic gene signature was constructed as follows:

Prognosis risk index =

Where n is the number of genes, β_i is the corresponding regression coefficient from the stepwise multivariate Cox regression analysis and Exp_i is the expression value of gene.

Using the median risk index as the cutoff value, the 364 HCC patients with survival data were divided into high- and low-risk groups. Kaplan-Meier analysis and time-dependent receiver operating characteristic (ROC) curve analysis were utilized to assess the distinguishing and predictive performance of the signature. Additionally, a stratified analysis was performed to evaluate whether the signature was efficient and practicable in the subgroups of the pathologic stage and age.

Identification of independent prognostic parameters

To investigate their independent prognostic value, univariate and multivariate Cox regression analyses were performed on the prognostic gene signature and the clinicopathological characteristics, including age, gender, pathologic stage, histologic grade, and vascular tumor invasion. Parameters with $P < 0.05$ based on univariate Cox regression analysis were considered statistically significant and further included in subsequent multivariate Cox regression analysis.

Predictive nomogram construction and validation

To predict the probability of 1-, 3- and 5-year OS for each patient with HCC, all independent prognostic parameters were included in the construction of a nomogram with the "*rms*" package in R software. Subsequently, Harrell's concordance index (C-index) and a calibration plot comparing observed and predicted OS were utilized to assess the performance of the nomogram. The C-index was calculated to

evaluate nomogram discrimination using a bootstrap method with 1000 resamplings. The calibration curve was plotted to compare observed and predicted OS.

Oncomine database analysis for prognostic signature

The expression levels of the prognostic signature in the prognosis risk model were further analyzed using the Oncomine database (https://www.oncomine.org), a web-based cancer microarray data mining platform. The mRNA expression folds in HCC tissue compared to respective normal tissue were obtained and compared. *P*-value < 0.05 and top 10% gene rank as the threshold.

Results

Expression analysis of DEGs in hepatocellular carcinoma

In the present study, a multistep analysis was performed to investigate DEGs and their significant biological functions on by integrated bioinformatics method in HCC (Figure 1). First, ten microarray profiles (GSE14520, GSE22058, GSE25097, GSE36376, GSE57957, GSE45267, GSE76427, GSE76297, GSE84005 and GSE121248) from the GEO datasets were selected and downloaded. There were a total of 2220 HCC samples, including 1016 adjacent nontumor 1204 tumor samples. After gene expression data processing, averaging and normalizing, DEGs among each GEO datasets were filtered and screened using the "*limma*" package with corrected *P*-value < 0.05 and |logFC| > 1. Based on the filter criteria, 779, 1723, 1675, 444, 373, 1062, 436, 451, 1025 and 580 DEGs from the above GEO datasets were obtained, respectively. The volcano plots of DEGs among each dataset are shown in Figure 2a. To explore robust DEGs in multiple GEO dataset, the RRA method was applied for integrating and carried out the meta-analysis of listed ranked genes. This method is a probabilistic and rank-based method assigned *P*-value as a significant score to each gene in the aggregated list, which suggests how superior it is ranked when compared to an expecting random ordering model. Finally, 518 significantly robust DEGs, which corresponded to 390 down-regulated and 128 up-regulated genes. The gene expression heatmap of the top 20 DEGs is shown in Figure 2b.

In the TCGA HCC cohort, 9077 DEGs were identified and filtered by the "*edgeR*" package, which comprised 7518 up-regulated and 1559 down-regulated genes. Among these DEGs, 451 DEGs held in common between GEO and TCGA HCC cohort were filtered out through Venn diagram analysis according to the thresholds set (Figure 2c), which corresponded to 126 up-regulated genes (logFC > 1.0) and 325 down-regulated genes (logFC < -1.0).

PPI network construction, hub gene selection and analysis

To explore the interactions and central genes associated with HCC, a PPI network was constructed using the STRING database and visualized by Cytoscape software. A total of 154 downregulated and 95 upregulated DEGs were filtered into the PPI network, which contained 249 nodes and 1274 edges. Then, Maximal Clique Centrality (MCC), Density of Maximum Neighborhood Component (DMNC), Maximum

Neighborhood Component (MNC) and Degree in the cytoHubba plug-in were employed for screening significant hub genes. After Venn diagram analysis of the top 50 genes screened by the four ranking algorithms, 41 genes were overlapped and identified as hub genes (Additional file 2: Figure S1).

External validation and survival analysis of hub genes

Since there are fewer normal liver tissues in the TCGA database, all 41 genes were subsequently investigated and analyzed in the GEPIA database, including 369 cancerous and 160 normal tissues in Genotype-Tissue Expression (GTEx) and TCGA projects. Based on the cutoffs (|log2FC| > 1 and $p < 0.01$), 36 of the 41 hub genes were significantly upregulated in HCC tissues compared to normal liver tissues (Figure 3). Subsequently, overall survival (OS) and disease-free survival (RFS) analysis of 36 hub genes were further performed using Kaplan-Meier analysis in the GEPIA database (364 HCC). The results illustrated that HCC patients with high expression (> median expression value) of these genes displayed worse OS and RFS ($p < 0.05$; Additional file 3: Table S2).

Function and pathway analysis of hub genes

To elucidate the functional characteristics of the verified 36 hub genes, the enrichment analysis of GO and KEGG pathway was performed on FunRich software. The GO enrichment analysis results were divided into three functional categories (BP, CC and MF) (Additional file 4: Table S3). For BP, these 36 hub genes were mainly enriched in spindle assembly, cell cycle, chromosome segregation, cell communication and signal transduction (Figure 4a). For CC, they were particularly enriched in kinetochore, nucleus, microtubule, chromosome and condensed chromosome kinetochore (Figure 4b). For MF, they were notably enriched in protein serine/threonine kinase activity, kinase binding, protein binding, motor activity and ATP binding (Figure 4c). According to KEGG pathway enrichment analysis, they were significantly enriched in the cell cycle, PLK1 signaling events, Polo-like kinase signaling events in the cell cycle, Aurora B signaling and M phase (Figure 4d, Additional file 4: Table S3).

Construction of the gene-related associated risk score system

After the exclusion of the patients with incomplete survival data, 364 HCC patients remained in this study. Univariate Cox analysis was performed to explore the relationship between the overall survival and the expression level of each hub gene. As a result, the high expression of 36 hub genes was significantly correlated with worse overall survival ($P < 0.05$). Then, LASSO-penalized Cox analysis was used to remove confounding factors and cut the number of genes by 10-fold cross-validation producing the best lambda to minimize the biases and errors. Based on the LASSO-penalized Cox regression, 18 genes (ASPM, AURKB, CCNA2, CDCA3, CDCA8, CENPA, CENPF, CENPK, ECT2, KIF20A, KIF4A, NEK2, PBK, PRC1, RACGAP1, RRM2, SPC25, TOP2A and TPX2) closely correlated with survival were selected for the development of a risk prediction model (Figure 5a, 5b). Finally, a stepwise multivariate Cox regression analysis was performed, and 11 genes were finally chosen to establish an evaluation model. The model was characterized as: prognosis index = (-0.2348 * expression level of AURKB) + (-0.4974 * expression level of CDCA3) + (0.6346 * expression level of CDCA8) + (0.3109 * expression level of CENPA) + (0.2525

* expression level of ECT2) + (0.3963 * expression level of KIF20A) + (-0.2562 * expression level of NEK2) + (-0.8553 * expression level of PRC1) + (0.3235 * expression level of SPC25) + (-0.3274 * expression level of TOP2A) + (0.6281 * expression level of TPX2). All of these genes, including CDCA8, CENPA, ECT2, KIF20A, SPC25 and TPX2, displayed positive coefficients in the formula, indicating high-risk signatures for them. The risk score of each HCC patient was calculated, and the patients were divided into the high-risk (n=182) or low-risk (n=182) group based on the median risk score as the cutoff value. The distributions of the eleven-gene-based risk scores, OS statuses and the expression profiles of 11 genes in the TCGA HCC cohort were displayed in Figure 5c. Intuitively, the number of deaths was notably taller in the high-risk group, and the heatmap suggested that all 11 genes were expressed at higher levels in the high-risk group compared to the low-risk group. Kaplan-Meier analysis of the entire dataset (n=364) clearly showed that the HCC patients in the high-risk group had a worse prognosis than those in the low-risk group (median OS: 3.11 vs 6.73 years, *P* = 3.21E-04) (Figure 5d). Subsequently, the prognostic capacity of the model was evaluated by calculating the receiver operating characteristic (ROC) area under a curve. The areas under the curves (AUCs) of the ROC curve of the whole cohort based on this predictive model were 0.755, 0.708 and 0.729 for the 1-, 3- and 5-year survival time, respectively (Figure 5e), suggesting that the predictive model had a high specificity and sensitivity.

Next, the model was further assessed in subgroups. The 374 patients with HCC were divided into the stage-Ⅰ subgroup (n=168), stage-Ⅱ subgroup (n=84), stage-Ⅲ subgroup (n=83) based on their pathologic stage. Except for the stage-Ⅲ subgroup, which did not have a plenitudinous sample, patients in each subgroup were divided into the high-risk or low-risk group based on the above cutoff value. As shown in Figure 6a-c, patients in the high-risk group had significantly shorter survival time than those in the low-risk group. When a stratified analysis was executed based on age, the test in the <65-year-old or ≥65-year-old subgroup illustrated the same results (Figure 6d, 6e). Thus, the model was certainly reliable and practicable in the prognosis evaluation.

Independence of the prognostic signature from other clinical characteristics

Next, univariate and multivariate Cox regression analyses were utilized to explore whether the prognostic performance of the signature was independent of those of conventional clinical risk factors in 283 HCC patients with full clinical information. Univariate Cox regression analysis suggested that the signature had independent prognostic value (*P*< 0.05), while age, gender, pathologic stage, histologic grade and vascular tumor invasion did not closely correlate with the OS and could be not extremely effective of independent prognostic signature (Table 1). Considering that age nearly reached the statistical significance, it and the prognostic model were incorporated into the multivariate Cox analysis. The results indicated that the eleven-gene signature could be an independent prognostic indicator (Table 1).

Table 1 Univariate and multivariate Cox regression analysis of 11-mRNA signature and clinical risk factors in TCGA HCC dataset

Characteristic	Univariate analysis	Multivariate analysis		
HR (95% CI)	*P*-Value	HR (95% CI)	*P*-Value	
Age (≥65 vs. <65)	1.422 (0.961-2.105)	0.078	1.459 (0.985-2.161)	0.059
Gender (Male vs. Female)	0.787(0.527-1.175)	0.242		
Pathologic_stage (III-IV vs. I-II)	1.010 (0.632-1.615)	0.966		
Histologic_grade (G3-G4 vs. G1-G2)	1.041 (0.696-1.555)	0.846		
Vascular_tumor_invasion (macro vs. micro vs. none)	0.825 (0.591-1.153)	0.260		
Risk_level (high vs. low)	2.316 (1.531-3.503)	6.99E-05	2.342 (1.548-3.544)	5.61E-05

Abbreviations: HR, hazard ratio; CI, confidence interval.

Development and validation of a predictive nomogram

To build an efficient quantitative method for predicting the survival probability of HCC patients, a user-friendly nomogram included two factors (age and prognostic model) that were generated (Figure 7a). The nomogram allowed the clinicians to calculate the 1-, 3- and 5-year OS probability of each HCC patient easily. Subsequently, the discrimination and calibration abilities of the nomogram were evaluated by using a concordance index (C-index) and calibration plots. The C-index was 0.707 (95% confidence interval: 0.660 - 0.754) using bootstrap with 1000 resamplings, suggesting that this nomogram has an excellent discrimination ability. The 1-, 3- and 5-year OS probabilities were visualized by calibration plots (Figure 7b), suggesting that the probabilities generated by this nomogram were closely approximated the actual survival situation.

Expression validation and KEGG analysis of the prognostic signature

The eleven genes in the prognostic model were further analyzed individually in the Oncomine database. As illustrated in Additional file 6: Figure S2, the mRNA expression levels of AURKB, CDCA3, CDCA8, CENPA, ECT2, KIF20A, NEK2, PRC1, SPC25, TOP2A and TPX2 were notably upregulated in HCC tissues compared with those in nontumor liver tissues ($p < 0.05$), which was consistent with our findings. To further elucidate and understand the biological functions of the eleven genes, KEGG pathway enrichment analysis was performed again. Results showed that these genes were markedly enriched in the biological process related to the regulation of cell proliferation, such as “cell cycle”, “PLK1 signaling pathway” and “Aurora B signaling pathway” (Additional file 5: Table S4).

Discussion

Hepatocellular carcinoma is one of the deadly malignant cancers worldwide with > 700,000 new cases per year [3], which is associated with various risk factors, such as hepatitis B virus, hepatitis C virus, alcoholic cirrhosis, nonalcoholic fatty liver diseases and certain genes. It is estimated that Asia will possess the largest number of HCC patients in the world by 2030 [9]. Although advances in the diagnosis and treatment, 1-, 3- and 5-year overall survivals of HCC are still low and unsatisfactory. Accurate prediction of overall survival is vital for individualized therapy approach selection and prognosis improvement. Conventional indicators such as pathologic stage, histologic grade, vascular invasion and tumor-node-metastasis status currently assist in predicting prognosis to a certain extent. However, the clinical outcomes of patients with the same stage or status often differ, suggesting that the current assessment is not sufficient for the heterogeneity of HCC and identification of novel prognostic biomarkers and construction of more accurate prognostic models are urgently required. The integration of prognostic signature and significant clinical parameters may contribute to a better prognosis prediction than a single biomarker or clinical factor. Recently, prognostic signatures based on DEGs have caused wide public concern and showed great potential in prognosis prediction of cancer and novel therapeutic targets.

In the present study, 451 robust DEGs were identified through an integrated analysis of multiple microarray and transcriptome profiles from different analysis platforms. Among them, eleven genes (AURKB, CDCA3, CDCA8, CENPA, ECT2, KIF20A, NEK2, PRC1, SPC25, TOP2A and TPX2) closely correlated with survival were selected by univariate, LASSO and multivariate Cox analysis. Based on their expression level and corresponding regression coefficients, a prognostic risk model was proposed. The verification of this model in the entire cohort and subgroup (age and pathologic stage) showed that patients in high-risk groups had significantly worse overall survival. The AUCs of the ROC curve of the whole dataset based on this model were 0.755, 0.708 and 0.729 at 1, 3 and 5 years of OS, respectively. For further application of this model in clinical, a user-friendly nomogram integrated the eleven-gene signature with age was generated and can be used to assess the survival probability of individual patients. The C-index and calibration plots indicated that the predicted survival is very close to the actual survival situation, suggesting that this model had an excellent performance in survival prediction. Based on the nomogram, the clinician can set an appropriate treatment plan to achieve the individualized treatment of HCC patients.

With the growing interest in personalized medicine, many prognostic risk assessments have been identified and found to boost survival predictions in a wide variety of cancers. Unfortunately, DEGs with no biological functions could be obtained due to the limited sample sizes and impertinent methods such as plain intersecting the DEGs lists from different technological platforms. In terms of these difficulties, ten microarray profiles from the GEO database and a high-throughput sequencing profile from the TCGA database were integrated to identify DEGs in this study. Undoubtedly, the candidate genes filtered in this study are more reliable than those identified in previous studies. Among reliable DEGs, the eleven genes

included in the prognostic model were closely related to the proliferation and invasion of the tumor cell and further analyzed individually.

Aurora kinase B (AURKB) is the core enzyme of the chromosomal passenger complex (CPC), which is responsible for the accurate regulation of chromosomal segregation, cytokinesis, mitotic checkpoint and protein localization to the centromere and kinetochore, and plays a vital role in the correction of microtubule-kinetochore attachments [10–12]. Its uncontrolled expression can promote aneuploidy and tumor development. Increasing evidence revealed that AURKB was highly expressed in a variety of malignant cancers. Tanaka *et al.* reported that high expression level AURKB in HCC and may be an effective predictor of aggressive HCC recurrence after curative hepatectomy [10]. Benten *et al*. reported that inhibition of AURKB by a novel aurora kinase inhibitor (PHA-739358) significantly suppressed the growth of hepatocellular carcinoma in vitro and a xenograft mouse model [13], implying AURKB could be a potential therapeutic target for HCC. Another study showed a selective inhibitor of AURKB (AZD1152) suppressed histone H3 phosphorylation and induced cellular apoptosis in twelve human HCC cell lines; AZD1152 can decelerate tumor growth and increase survival in an orthotopic hepatoma model [14].

Cell division cycle associated 3 (CDCA3), a component of Skp1-cullin-F-box, acts as a regulator of mitosis. Timo *et al*. demonstrated CDCA3, as a driver gene in carcinogenesis, exerted crucial functions in HCC by inducing cell cycle progression [15]. Cell division cycle associated 8 (CDCA8) is one key regulatory component of CPC, which plays a crucial regulatory role in mitosis and cell division [16]. The high expression level of CDCA8 was positively correlated to tumor cell proliferation, invasion, and poor prognosis in a variety of cancers, such as gastric [17] and lung cancer [18].

Centromere protein A (CENPA), a histone H3 variant of centromeric nucleosomes, plays a vital role in cell cycle regulation and genetic stability. Previous research has shown that it can promote HCC cell proliferation both in vitro and in vivo [19]. RNAi-mediated its depletion inhibited HCC cell growth, blocked cell cycle progression at the G1 phase, and assisted apoptosis of HCC cell [20, 21].

Epithelial cell transforming sequence 2 (ECT2) belongs to the Ras GTPases superfamily,[22] is a classical oncogene originally identified in 1991 and highly expressed in various cancers [23]. It performs crucial regulatory roles in cell proliferation, oncogenesis, tumorigenesis, and metastasis of HCC cells [24]. Chen *et al*. demonstrated that ECT2 was significantly upregulated in HCC and strongly responsible for promoting early recurrence of HCC. Knockdown of ECT2 will notably suppress Rho GTPases activities, enhance apoptosis, attenuate oncogenicity, and restrain the metastatic ability of HCC cells [24]. Another study reported that the down-regulation of ECT2 by miR-190-5p can significantly inhibit the metastasis of HCC [25]. These results indicated that ECT2 may be a potential therapeutic target for HCC.

Kinesin family member 20A (KIF20A), a microtubule-associated motor protein, is an essential mitotic kinesin required for cytokinesis, which can directly interact with Rab6 small GTPase and participate in the dynamics of the Golgi apparatus [26]. Increasing evidence revealed that KIF20A may play an important role in the development and progression of various cancers [27]. Shi *et al*. reported that knockdown of

KIF20A could markedly inhibit the growth of HCC cells in vitro and the growth of HCC xenografts in vivo [28].

NIMA-related kinase 2 (NEK2) located in the centrosome, a member of the NIMA family of serine/threonine kinases, is involved in the cell cycle and mitosis as a vital oncogene. Elevation of NEK2 will contribute to premature centrosome splitting, concomitant with centrosomal abnormalities, monopolar spindles and aneuploidy [29]. Previous studies have reported that it was overexpressed in numerous cancer cell lines and associated closely with tumor size, portal vein invasion and poor tumor differentiation [30]. Zhang *et al*. reported that over-expression of NEK2 could promote the proliferation of HepG2 cells by activating the mitogen-activated protein kinase (MAPK) signaling pathway [31] and interfering or silencing NEK2 expression can suppress HepG2 cells' cell growth and invasion, promote apoptosis and cycle arrest [32]. Lai *et al*. demonstrated that aberrant NEK2 could accelerate the progression of the cell cycle and promote cell proliferation via the activation of the Wnt/β-catenin signaling pathway [33]. Another study found that overexpression of NEK2 in hepatoma cells could enhance drug resistance, promote metastasis and angiogenesis via the PI3K/AKT-NF-κB signaling pathway [34]. These studies suggested that the aberrant expression of NEK2 was significantly relative to the occurrence and progression of HCC and imperfect survival.

Protein regulator of cytokinesis-1 (PRC1), a member of the microtubule-associated protein family, is involved in cytokinesis and carcinogenesis. Increasing evidence revealed that PRC1 as an oncogenic factor in tumorigenesis of HCC [35] and lung adenocarcinoma [36] by promoting cancer proliferation, stemness, metastasis and tumorigenesis. Chen *et al*. demonstrated that PRC1 may promote an early recurrence of HCC in association with the Wnt/β-catenin signaling pathway [35]. Zhan *et al*. confirmed that silencing or decreasing PRC1 expression can inhibit the proliferation and invasion of cancer cells in vitro and in vivo [36]. Taken together, these studies indicated that silencing of PRC1 could suppress tumorigenesis in a variety of cancer cells, and further demonstrated that PRC1 may become a novel potential target for gene therapy of tumors.

Spindle pole body component 25 (SPC25), as a component of the nuclear division cycle 80 (Ndc80) complex, plays an essential role in chromosomal segregation by regulating the combination of microtubule and kinetochore protein. Several studies have reported that the dysregulation of SPC25 was closely related to the oncogenic process and malignant phenotypes of cancers [37]. Additionally, SPC25 plays a vital part in carcinogenesis, cancer cell growth, and metastasis. Several studies have indicated that SPC25 was markedly upregulated in cancers and involved in cancer cell proliferation and metastasis. Chen *et al*. demonstrated that SPC25 knockdown could significantly inhibit the proliferation, invasion, migration and adhesion of HCC cells in vitro via the p53 signaling pathway [38].

Topoisomerase II alpha (TOP2A) encodes a DNA topoisomerase that controls and alters the topologic states of DNA during transcription and determines the tumor cell response to chemotherapeutics. Recently, several studies confirmed that intracellular topoisomerase levels could determine the

chemosensitivity of tumor cells,[39–41] suggesting that TOP2A may be a potential target for enhancing drug response in cancer therapy.

Targeting protein for Xenopus kinesin-like protein 2 (TPX2) is a nuclear proliferation microtubule-associated protein and impacts spindle assembly in mammalian cells [42]. Aberrant expression of TPX2 induces the amplification of centrosome and causes DNA polyploidy. Recently, several studies have reported that TPX2 promotes tumorigenesis and metastasis and is significantly upregulated in numerous types of malignant tumors. And increasing evidence demonstrated that TPX2 expression is associated with tumor migration, invasion and poor prognosis in HCC [43, 44]. Reiko et al. confirmed that the siRNA-mediated knockdown of TPX2 can inhibit the proliferation of HCC cells, and the growth of HCC xenografts transplanted into immunodeficient mice [45]. Liu *et al*. found that siRNA-mediated knockdown of TPX2 suppressed HCC cell invasion via inactivating AKT signaling and down-regulation of MMP2 and MMP9 expression in SMMC-7721 and HepG2 cells [46]. Another study reported that the siRNA-mediated knockdown of TPX2 may suppress the growth of HCC by regulating the PI3K/AKT signaling pathway [47]. These studies indicated that TPX2 may be a promising target for the treatment of HCC.

The predictive model is based on the expression levels of the above-mentioned genes. The eleven-gene signature was integrated with patients' age to build a user-friendly nomogram, which enables clinicians to assess patients' prognosis and facilitate the customized treatment for high-risk patients. To our knowledge, the eleven-gene signature described herein and the nomogram has not been studied previously. These findings will contribute directly to the understanding of the molecular mechanism of HCC and the prediction of prognosis. Moreover, the gene signature may be a potential therapeutic molecular target. However, this study also had certain limitations as follows: 1) the model was mainly validated in the TCGA-HCC project. To reduce the risk of overfitting, other independent cohorts with plenty of clinical data were needed. 2) the clinical information was required from the TCGA database and most of the patients are white and Asian. Caution must be taken when extrapolating our findings to other ethnicities.

Conclusion

In summary, significant and reliable survival-related DEGs were identified by using multiple gene expression profiles from the GEO and TCGA datasets, and a nomogram composed of an eleven-gene signature was established. Based on the prognosis risk index and patients' age, the clinician can change the patient's treatment plan to realize a more individualized approach for high-risk patients and effective measures can be taken to prevent or slow down the recurrence and deterioration of HCC in high-risk populations. The findings contribute directly to the survival prediction and treatment management for patients with HCC.

Abbreviations

HCC: Hepatocellular carcinoma; GEO: Gene Expression Omnibus database; TCGA: Cancer Genome Atlas; DEGs: Differentially expressed genes; LASSO: least absolute shrinkage and selection operator; RRA: Robust rank aggregation; KEGG: Kyoto Encyclopedia of Genes and Genomes; GO: Gene Ontology; GEPIA: Gene Expression Profiling Interactive Analysis; OS: overall survival; ROC: receiver operating characteristic; C-index: Harrell's concordance index; AUC: Areas under the curves.

Declarations

Competing interests

All authors declare that they have no competing interests.

Funding

No.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Funding information

National Natural Science Foundation of China, Grant/Award Numbers: 81802117; Talent Training of Army Medical University, Grant/Award Numbers: XZ-2019-505-012.

Authors' contributions

Y.L. Li contributed to the conception and design of this study, the acquisition, analysis and interpretation of the data and draughting this manuscript. Z.R. Liu contributed to the analysis and interpretation of the data. Q. Wang contributed to the critically revised manuscript. All authors have approved the final manuscript.

Acknowledgment

This study was supported by the National Natural Science Foundation of China (No. 81802117) and the Talent Training of Army Medical University (XZ-2019-505-012).

References

1. Heimbach JK, Kulik LM, Finn RS, Sirlin CB, Abecassis MM, Roberts LR, et al. AASLD guidelines for the treatment of hepatocellular carcinoma. Hepatology. 2018;67(1):358–80.
2. Yim SY, Seo YS, Jung CH, Kim TH, Lee JM, Kim ES, et al. The management and prognosis of patients with hepatocellular carcinoma: what has changed in 20 years? Liver Int. 2016;36(3):445–53.
3. Wang Z, Teng D, Li Y, Hu Z, Liu L, Zheng H. A six-gene-based prognostic signature for hepatocellular carcinoma overall survival prediction. Life Sci. 2018;203:83–91.
4. Villanueva A. Hepatocellular Carcinoma. N Engl J Med. 2019;380(15):1450–62.
5. Raivo Kolde S, Laur P, Adler, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics. 2012;28(4):573–80.
6. Liu L, He C, Zhou Q, Wang G, Lv Z, Liu J. Identification of key genes and pathways of thyroid cancer by integrated bioinformatics analysis. J Cell Physiol. 2019;234(12):23647–57.
7. 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. 2019;11(13):4736–56.
8. Sun G, Li Y, Peng Y, Lu D, Zhang F, Cui X, et al Identification of differentially expressed genes and biological characteristics of colorectal cancer by integrated bioinformatics analysis. J Cell Physiol. 2019.
9. Estes C, Anstee QM, Arias-Loste MT, Bantel H, Bellentani S, Caballeria J, et al. Modeling NAFLD disease burden in China, France, Germany, Italy, Japan, Spain, United Kingdom, and United States for the period 2016–2030. J Hepatol. 2018;69(4):896–904.
10. Tanaka S, Arii S, Yasen M, Mogushi K, Su NT, Zhao C, et al. Aurora kinase B is a predictive factor for the aggressive recurrence of hepatocellular carcinoma after curative hepatectomy. Br J Surg. 2008;95(5):611–9.
11. Vader G, Medema RH, Lens SM. The chromosomal passenger complex: guiding Aurora-B through mitosis. J Cell Biol. 2006;173(6):833–7.
12. Carloni V, Lulli M, Madiai S, Mello T, Hall A, Luong TV, et al. CHK2 overexpression and mislocalisation within mitotic structures enhances chromosomal instability and hepatocellular carcinoma progression. Gut. 2018;67(2):348–61.
13. Benten D, Keller G, Quaas A, Schrader J, Gontarewicz A, Balabanov S, et al. Aurora kinase inhibitor PHA-739358 suppresses growth of hepatocellular carcinoma in vitro and in a xenograft mouse model. Neoplasia. 2009;11(9):934–44.
14. Aihara A, Tanaka S, Yasen M, Matsumura S, Mitsunori Y, Murakata A, et al. The selective Aurora B kinase inhibitor AZD1152 as a novel treatment for hepatocellular carcinoma. J Hepatol. 2010;52(1):63–71.
15. Itzel T, Scholz P, Maass T, Krupp M, Marquardt JU, Strand S, et al. Translating bioinformatics in oncology: guilt-by-profiling analysis and identification of KIF18B and CDCA3 as novel driver genes in carcinogenesis. Bioinformatics. 2015;31(2):216–24.

16. Li L, Li D, Tian F, Cen J, Chen X, Ji Y, et al. Hepatic Loss of Borealin Impairs Postnatal Liver Development, Regeneration, and Hepatocarcinogenesis. J Biol Chem. 2016;291(40):21137–47.
17. Chang JL, Chen TH, Wang CF, Chiang YH, Huang YL, Wong FH, et al. Borealin/Dasra B is a cell cycle-regulated chromosomal passenger protein and its nuclear accumulation is linked to poor prognosis for human gastric cancer. Exp Cell Res. 2006;312(7):962–73.
18. Nguyen MH, Koinuma J, Ueda K, Ito T, Tsuchiya E, Nakamura Y, et al. Phosphorylation and activation of cell division cycle associated 5 by mitogen-activated protein kinase play a crucial role in human lung carcinogenesis. Cancer Res. 2010;70(13):5337–47.
19. Zhang Y, Yang L, Shi J, Lu Y, Chen X, Yang Z. The Oncogenic Role of CENPA in Hepatocellular Carcinoma Development: Evidence from Bioinformatic Analysis. Biomed Res Int. 2020; 2020:3040839.
20. Liu L, Li Y, Zhang S, Yu D, Zhu M. Hepatitis B virus X protein mutant upregulates CENP-A expression in hepatoma cells. Oncol Rep. 2012;27(1):168–73.
21. 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.
22. Vigil D, Cherfils J, Rossman KL, Der CJ. Ras superfamily GEFs and GAPs: validated and tractable targets for cancer therapy? Nat Rev Cancer. 2010;10(12):842–57.
23. Chaisaingmongkol J, Budhu A, Dang H, Rabibhadana S, Pupacdi B, Kwon SM, et al. Common Molecular Subtypes Among Asian Hepatocellular Carcinoma and Cholangiocarcinoma. Cancer Cell. 2017;32(1):57–70 e53.
24. Chen J, Xia H, Zhang X, Karthik S, Pratap SV, Ooi LL, et al. ECT2 regulates the Rho/ERK signalling axis to promote early recurrence in human hepatocellular carcinoma. J Hepatol. 2015;62(6):1287–95.
25. Fang ZQ, Li MC, Zhang YQ, Liu XG. MiR-490-5p inhibits the metastasis of hepatocellular carcinoma by down-regulating E2F2 and ECT2. J Cell Biochem. 2018;119(10):8317–24.
26. Lu M, Huang X, Chen Y, Fu Y, Xu C, Xiang W, et al. Aberrant KIF20A expression might independently predict poor overall survival and recurrence-free survival of hepatocellular carcinoma. Iubmb Life. 2018;70(4):328–35.
27. Gasnereau I, Boissan M, Margall-Ducos G, Couchy G, Wendum D, Bourgain-Guglielmetti F, et al. KIF20A mRNA and its product MKlp2 are increased during hepatocyte proliferation and hepatocarcinogenesis. Am J Pathol. 2012;180(1):131–40.
28. Shi C, Huang D, Lu N, Chen D, Zhang M, Yan Y, et al. Aberrantly activated Gli2-KIF20A axis is crucial for growth of hepatocellular carcinoma and predicts poor prognosis. Oncotarget. 2016;7(18):26206–19.
29. Faragher AJ, Fry AM. Nek2A kinase stimulates centrosome disjunction and is required for formation of bipolar mitotic spindles. Mol Biol Cell. 2003;14(7):2876–89.

30. Cheng Y, Chen X, Ye L, Zhang Y, Liang J, Liu W, et al. The Prognostic Significance of NEK2 in Hepatocellular Carcinoma: Evidence from a Meta-Analysis and Retrospective Cohort Study. Cell Physiol Biochem. 2018;51(6):2746–59.
31. Zhang Y, Wang W, Wang Y, Huang X, Zhang Z, Chen B, et al. NEK2 promotes hepatocellular carcinoma migration and invasion through modulation of the epithelial-mesenchymal transition. Oncol Rep. 2018;39(3):1023–33.
32. Zhang MX, Xu XM, Zhang P, Han NN, Deng JJ, Yu TT, et al. Effect of silencing NEK2 on biological behaviors of HepG2 in human hepatoma cells and MAPK signal pathway. Tumour Biol. 2016;37(2):2023–35.
33. Lai XB, Nie YQ, Huang HL, Li YF, Cao CY, Yang H, et al. NIMA-related kinase 2 regulates hepatocellular carcinoma cell growth and proliferation. Oncol Lett. 2017;13(3):1587–94.
34. Wu SM, Lin SL, Lee KY, Chuang HC, Feng PH, Cheng WL, et al. Hepatoma cell functions modulated by NEK2 are associated with liver cancer progression. Int J Cancer. 2017;140(7):1581–96.
35. Chen J, Rajasekaran M, Xia H, Zhang X, Kong SN, Sekar K, et al. The microtubule-associated protein PRC1 promotes early recurrence of hepatocellular carcinoma in association with the Wnt/beta-catenin signalling pathway. Gut. 2016;65(9):1522–34.
36. Zhan P, Zhang B, Xi GM, Wu Y, Liu HB, Liu YF, et al. PRC1 contributes to tumorigenesis of lung adenocarcinoma in association with the Wnt/beta-catenin signaling pathway. Mol Cancer. 2017;16(1):108.
37. Pathania R, Ramachandran S, Mariappan G, Thakur P, Shi H, Choi JH, et al. Combined Inhibition of DNMT and HDAC Blocks the Tumorigenicity of Cancer Stem-like Cells and Attenuates Mammary Tumor Growth. Cancer Res. 2016;76(11):3224–35.
38. Chen F, Zhang K, Huang Y, Luo F, Hu K, Cai Q. SPC25 may promote proliferation and metastasis of hepatocellular carcinoma via p53. Febs Open Bio. 2020.
39. Burgess DJ, Doles J, Zender L, Xue W, Ma B, McCombie WR, et al. Topoisomerase levels determine chemotherapy response in vitro and in vivo. Proc Natl Acad Sci U S A. 2008;105(26):9053–8.
40. Wang N, Zhu M, Tsao SW, Man K, Zhang Z, Feng Y. MiR-23a-mediated inhibition of topoisomerase 1 expression potentiates cell response to etoposide in human hepatocellular carcinoma. Mol Cancer. 2013;12(1):119.
41. Wong N, Yeo W, Wong WL, Wong NL, Chan KY, Mo FK, et al. TOP2A overexpression in hepatocellular carcinoma correlates with early age onset, shorter patients survival and chemoresistance. Int J Cancer. 2009;124(3):644–52.
42. Gruss OJ, Vernos I. The mechanism of spindle assembly: functions of Ran and its target TPX2. J Cell Biol. 2004;166(7):949–55.
43. Wang F, Zhao W, Gao Y, Zhou J, Li H, Zhang G, et al. CDK5-mediated phosphorylation and stabilization of TPX2 promotes hepatocellular tumorigenesis. J Exp Clin Cancer Res. 2019;38(1):286.
44. Wang LL, Jin XH, Cai MY, Li HG, Chen JW, Wang FW, et al. AGBL2 promotes cancer cell growth through IRGM-regulated autophagy and enhanced Aurora A activity in hepatocellular carcinoma.

Cancer Lett. 2018;414:71–80.

45. Satow R, Shitashige M, Kanai Y, Takeshita F, Ojima H, Jigami T, et al. Combined functional genome survey of therapeutic targets for hepatocellular carcinoma. Clin Cancer Res. 2010;16(9):2518–28.
46. Liu Q, Yang P, Tu K, Zhang H, Zheng X, Yao Y, et al. TPX2 knockdown suppressed hepatocellular carcinoma cell invasion via inactivating AKT signaling and inhibiting MMP2 and MMP9 expression. Chin J Cancer Res. 2014;26(4):410–7.
47. Huang DH, Jian J, Li S, Zhang Y, Liu LZ. TPX2 silencing exerts antitumor effects on hepatocellular carcinoma by regulating the PI3K/AKT signaling pathway. Int J Mol Med. 2019;44(6):2113–22.

Figures

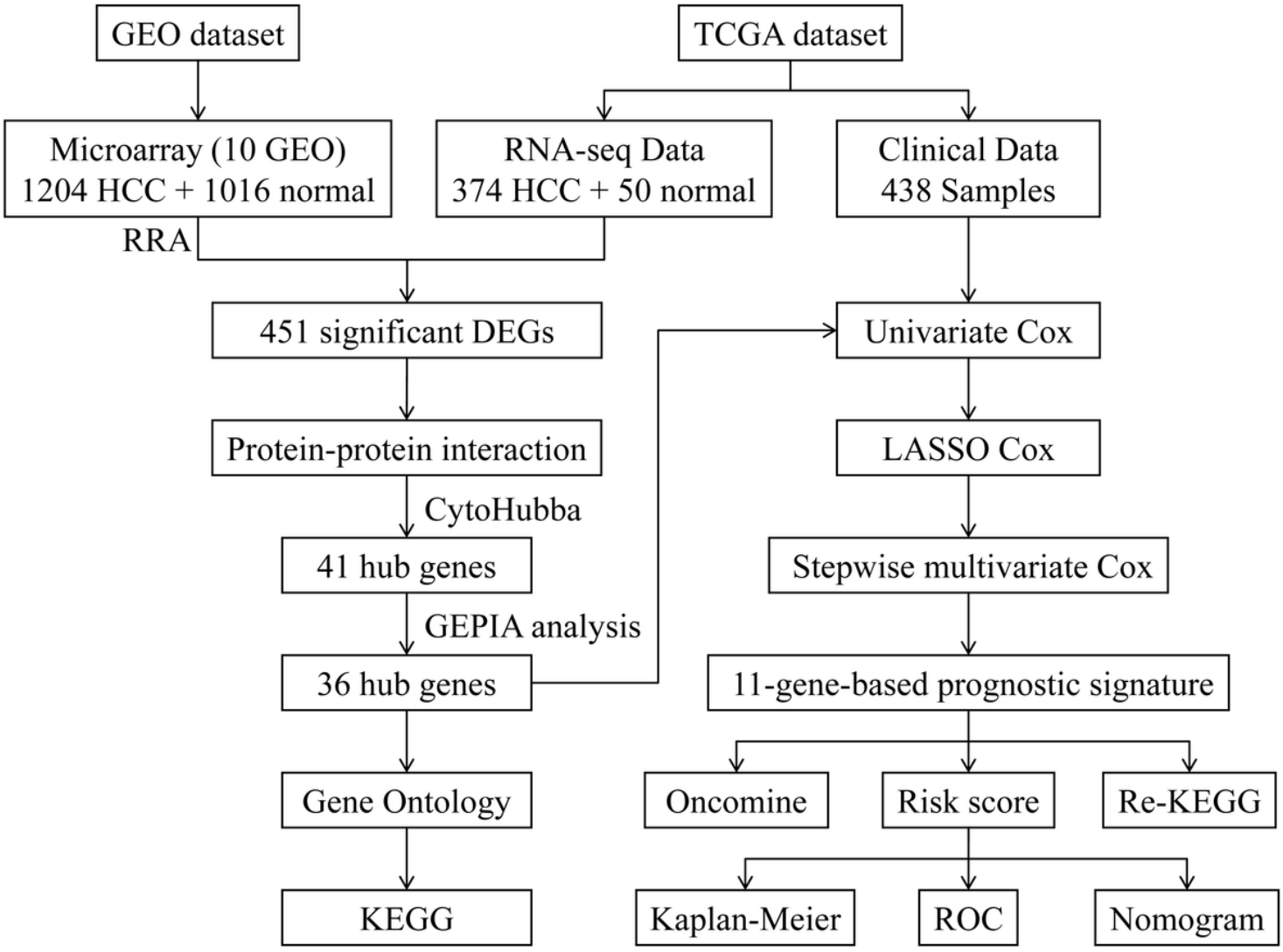

Figure 1

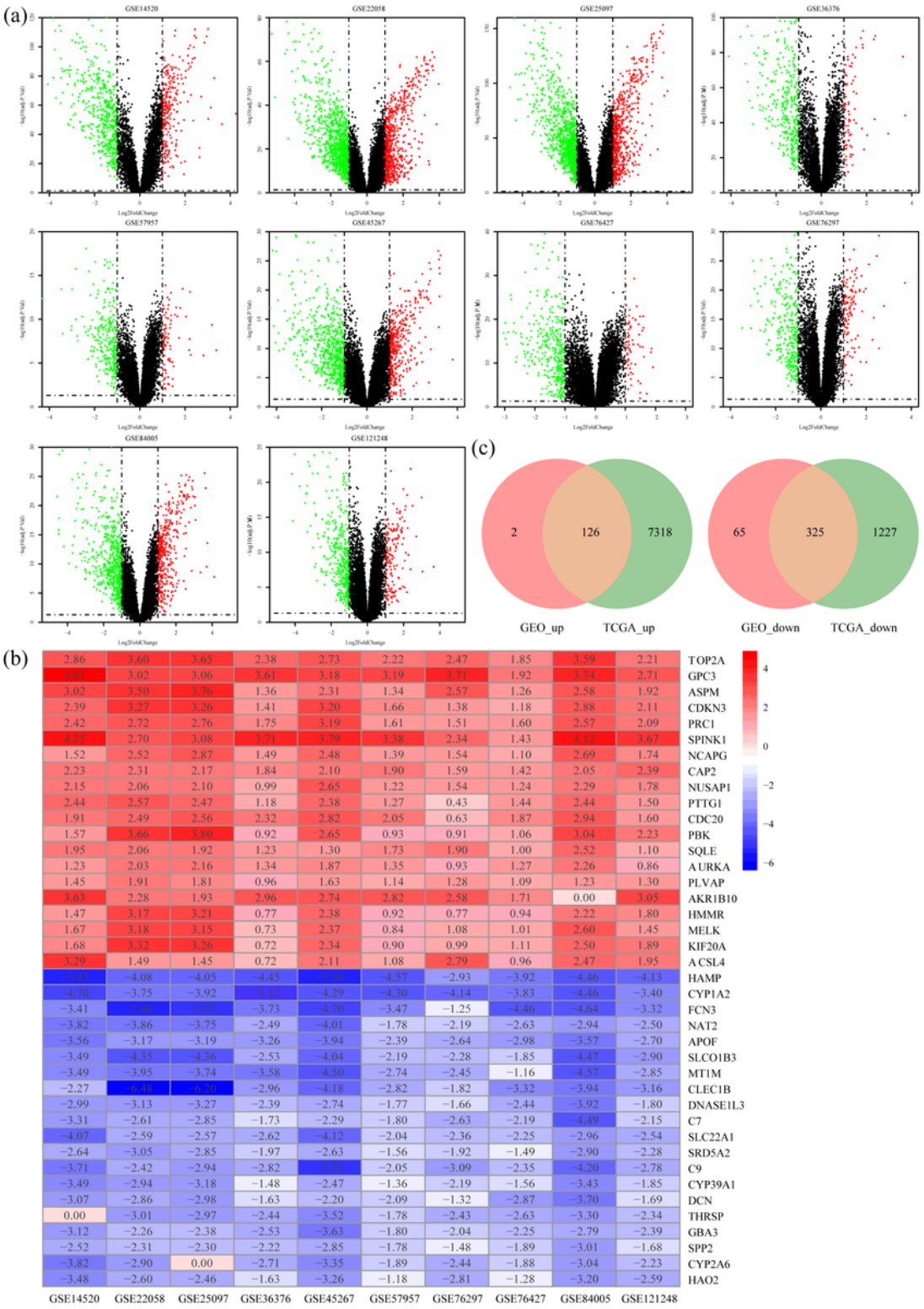
(a)
GSE14520
GSE22058
GSE25097
GSE36376
GSE57957
GSE45267
GSE76427
GSE76297
GSE84005
GSE121248
(c)
2
126
7318
GEO_up
TCGA_up
65
325
1227
GEO_down
TCGA_down
(b)
TOP2A
GPC3
ASPM
CDKN3
PRC1
SPINK1
NCAPG
CAP2
NUSAP1
PTTG1
CDC20
PBK
SQLE
AURKA
PLVAP
AKR1B10
HMMR
MELK
KIF20A
ACSL4
HAMP
CYP1A2
FCN3
NAT2
APOF
SLCO1B3
MT1M
CLEC1B
DNASE1L3
C7
SLC22A1
SRD5A2
C9
CYP39A1
DCN
THRSP
GBA3
SPP2
CYP2A6
HAO2
GSE14520
GSE22058
GSE25097
GSE36376
GSE45267
GSE57957
GSE76297
GSE76427
GSE84005
GSE121248

Figure 2

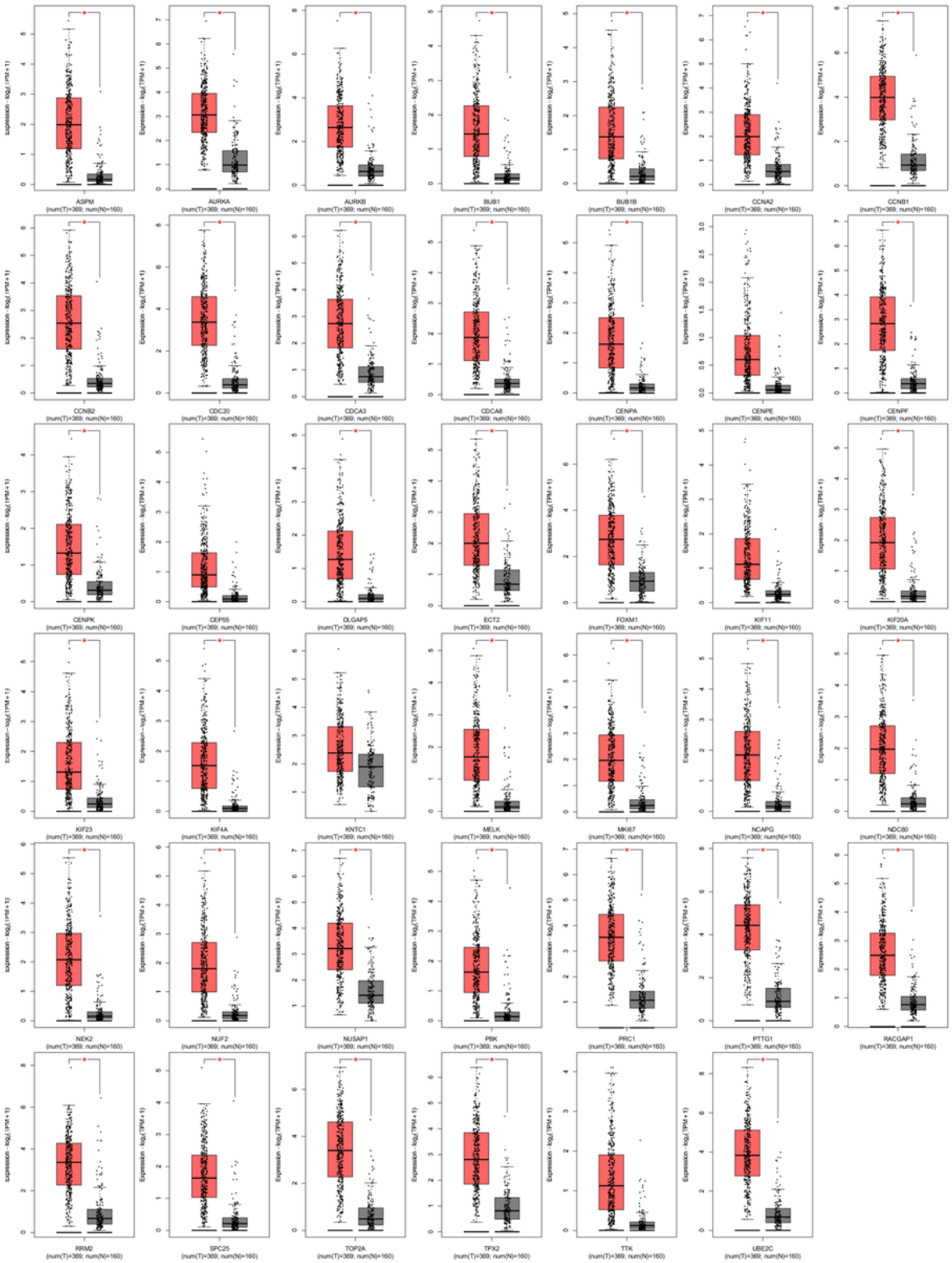
ASPM
AURKA
AURKB
BUB1
BUB1B
CCNA2
CCNB1
CCNB2
CDC20
CDCA3
CDCA8
CENPA
CENPE
CENPF
CENPK
CEP55
DLGAP5
ECT2
FOXM1
KIF11
KIF20A
KIF23
KIF4A
KNTC1
MELK
MKI67
NCAPG
NDC80
NEK2
NUF2
NUSAP1
PBK
PRC1
PTTG1
RACGAP1
RRM2
SPC25
TOP2A
TPX2
TTK
UBE2C

Figure 3

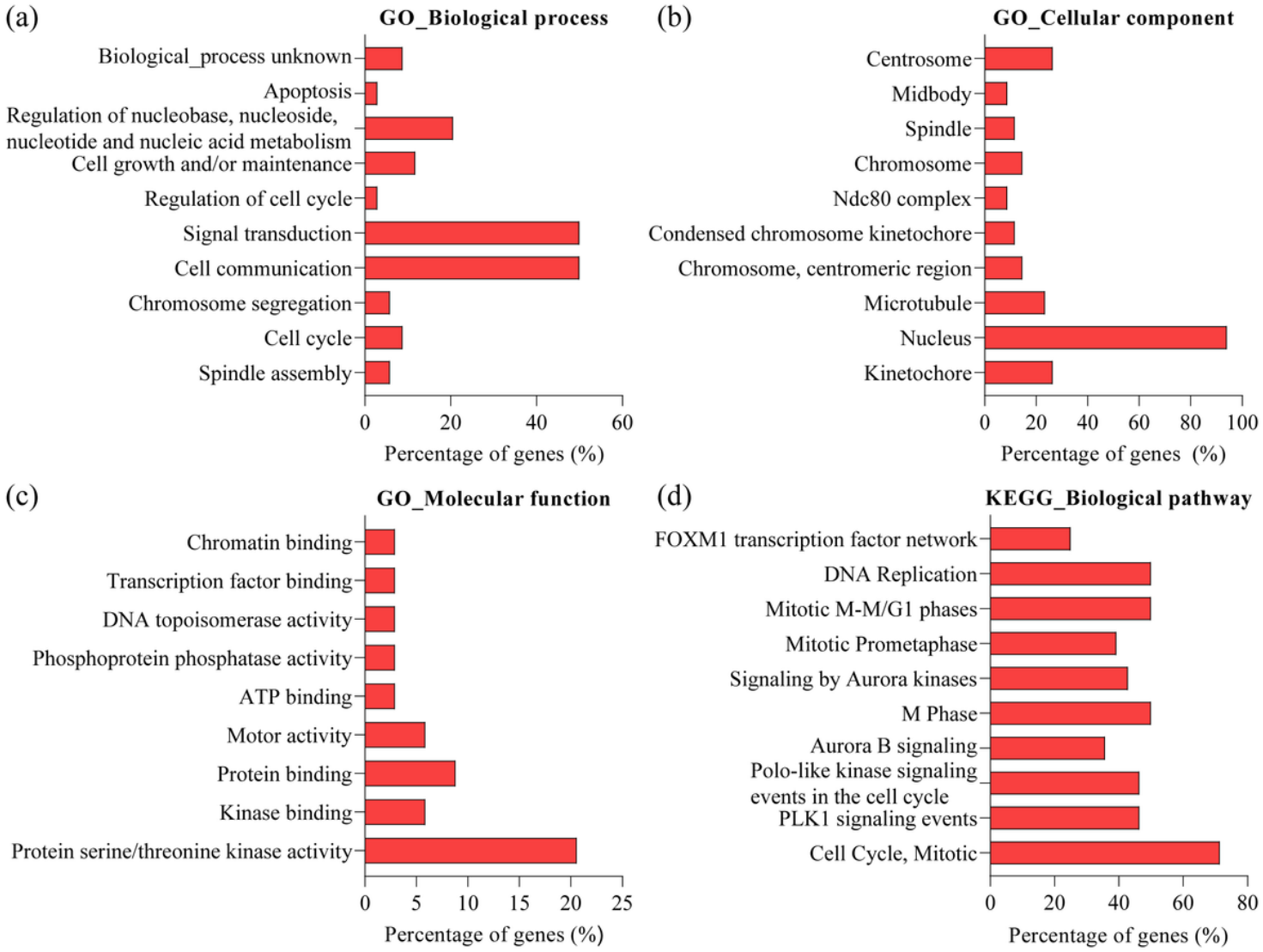
(a)
GO_Biological process
Biological_process unknown
Apoptosis
Regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolism
Cell growth and/or maintenance
Regulation of cell cycle
Signal transduction
Cell communication
Chromosome segregation
Cell cycle
Spindle assembly
0 20 40 60
Percentage of genes (%)
(b)
GO_Cellular component
Centrosome
Midbody
Spindle
Chromosome
Ndc80 complex
Condensed chromosome kinetochore
Chromosome, centromeric region
Microtubule
Nucleus
Kinetochore
0 20 40 60 80 100
Percentage of genes (%)
(c)
GO_Molecular function
Chromatin binding
Transcription factor binding
DNA topoisomerase activity
Phosphoprotein phosphatase activity
ATP binding
Motor activity
Protein binding
Kinase binding
Protein serine/threonine kinase activity
0 5 10 15 20 25
Percentage of genes (%)
(d)
KEGG_Biological pathway
FOXM1 transcription factor network
DNA Replication
Mitotic M-M/G1 phases
Mitotic Prometaphase
Signaling by Aurora kinases
M Phase
Aurora B signaling
Polo-like kinase signaling events in the cell cycle
PLK1 signaling events
Cell Cycle, Mitotic
0 20 40 60 80
Percentage of genes (%)

Figure 4

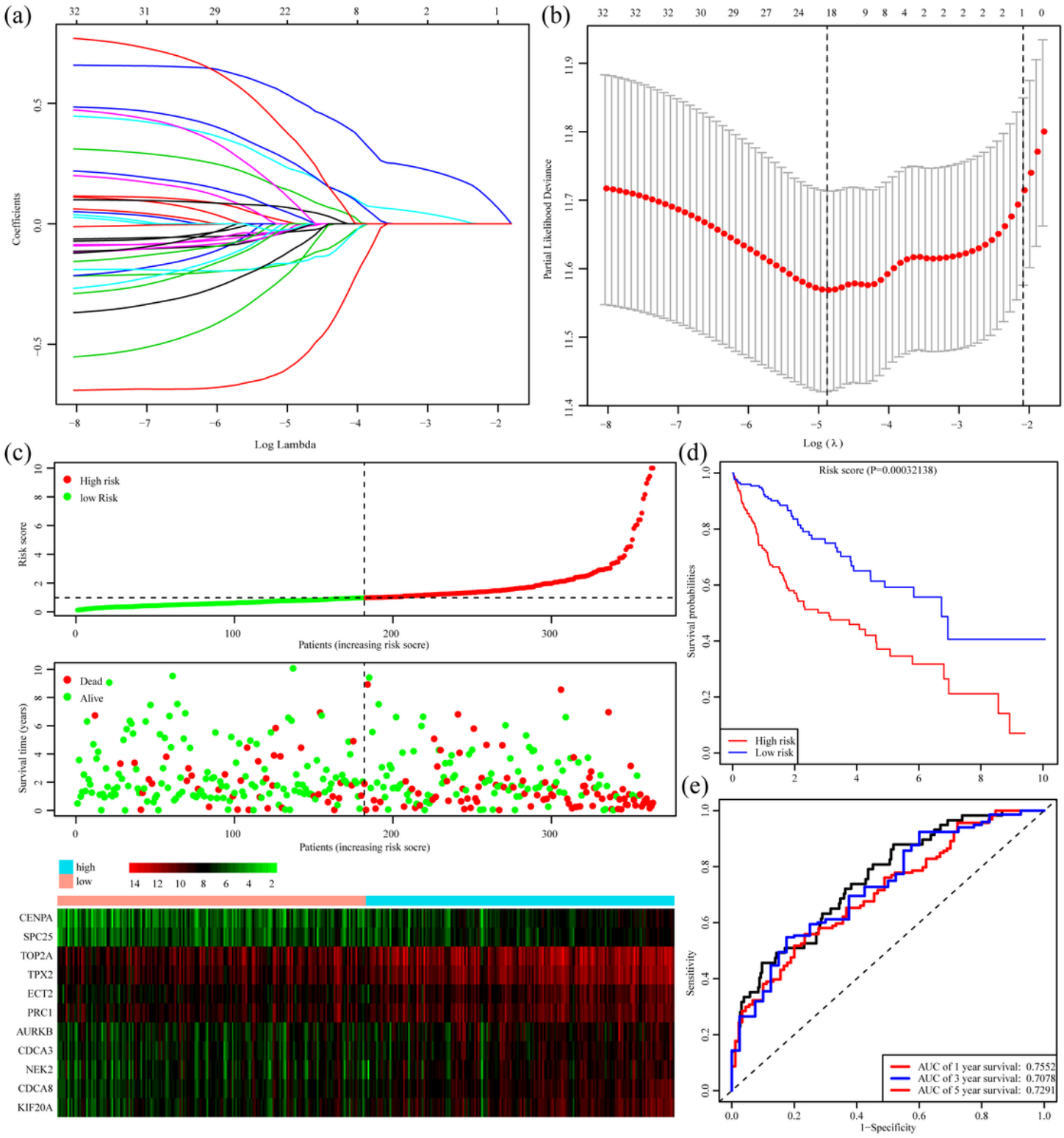
(a)
Coefficients
Log Lambda
(b)
Partial Likelihood Deviance
Log (λ)
(c)
High risk
low Risk
Risk score
Patients (increasing risk socre)
Dead
Alive
Survival time (years)
high
low
CENPA
SPC25
TOP2A
TPX2
ECT2
PRC1
AURKB
CDCA3
NEK2
CDCA8
KIF20A
(d)
Risk score (P=0.00032138)
Survival probabilities
High risk
Low risk
(e)
Sensitivity
1−Specificity
AUC of 1 year survival: 0.7552
AUC of 3 year survival: 0.7078
AUC of 5 year survival: 0.7291

Figure 5

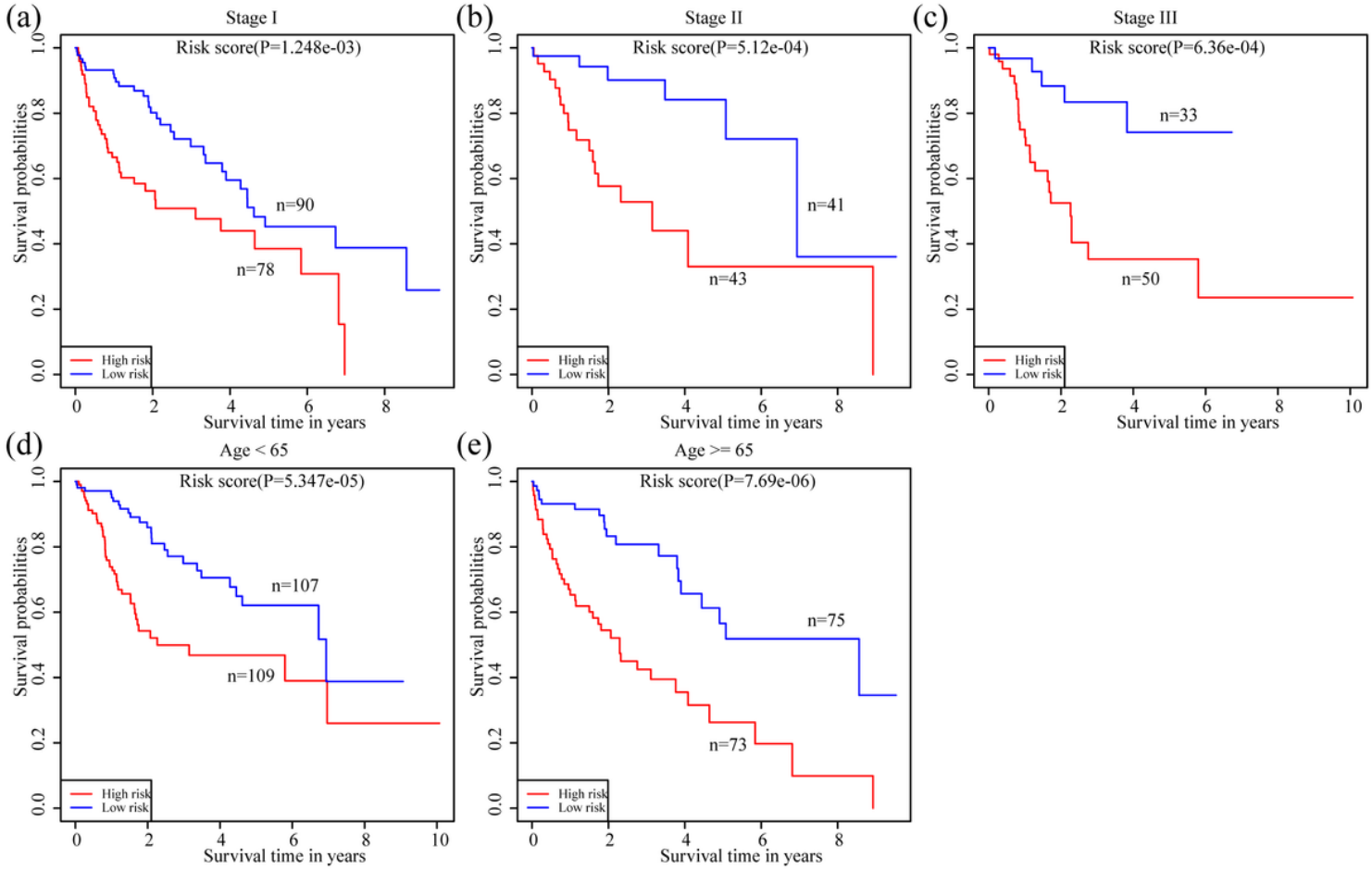
(a)
Stage I
Risk score(P=1.248e-03)
n=90
n=78
(b)
Stage II
Risk score(P=5.12e-04)
n=41
n=43
(c)
Stage III
Risk score(P=6.36e-04)
n=33
n=50
(d)
Age < 65
Risk score(P=5.347e-05)
n=107
n=109
(e)
Age >= 65
Risk score(P=7.69e-06)
n=75
n=73
High risk
Low risk
Survival probabilities
Survival time in years

Figure 6

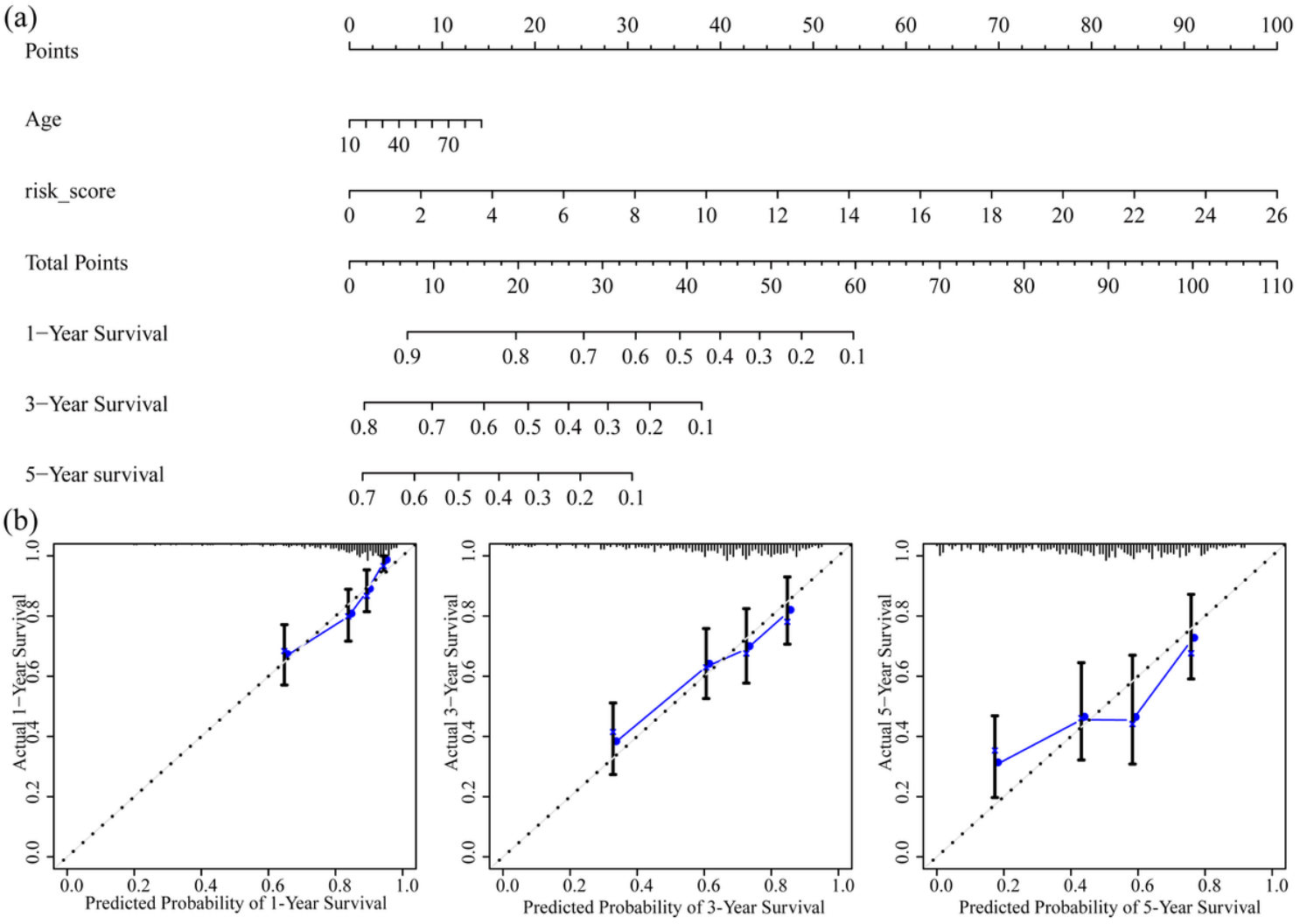

Figure 7

Supplementary Files

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

- Additionalfile6FigureS2.docx
- Additionalfile5TableS4.xls
- Additionalfile4TableS3.xls
- Additionalfile3TableS2.xls
- Additionalfile2FigureS1.docx
- Additionalfile1TableS1.xls