

# Expression of Coiled-Coil Domain Containing Family mRNA and prognostic value in hepatocellular carcinoma

**Sina Zhang**

The First Affiliated Hospital of Wenzhou Medical University <https://orcid.org/0000-0001-8063-5804>

**Chen Jin**

Wenzhou Medical University

**Yan Yang**

Wenzhou Medical University

**Haoqi Li**

Wenzhou Medical University

**Jun Ma**

Wenzhou Medical University

**Gang Chen**

The First Affiliated Hospital of Wenzhou Medical University

**Yi Wang** (✉ [wang.yi@wmu.edu.cn](mailto:wang.yi@wmu.edu.cn))

School of Public Health and Management, Wenzhou Medical University <https://orcid.org/0000-0002-0558-7491>

---

## Primary research

**Keywords:** Hepatocellular carcinoma, TCGA, ICGC, Coiled-Coil Domain Containing Family, Prognosis, Gene signature

**Posted Date:** January 29th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-153938/v1>

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

[Read Full License](#)

---

# Abstract

**Background:** The CCDC family plays a significant role in the development and progression of malignant tumors. However, the relationship between CCDC family members and HCC progression is incompletely known. This study used bioinformatics analysis to investigate the expression as well as clinical prognostic value of CCDC family members in HCC and to predict the role of CCDCs family in the development and progression of HCC.

**Methods:** This study utilized the data from two platforms databases to explore the diagnostic value and prognostic significance of CCDC family members by Cox proportional hazards regression analysis, Kaplan-Meier curve and log-rank test, ROC and nomogram diagnostic and prognostic analysis methods. GSEA and tumor microenvironment analysis were employed to investigate the underlying mechanisms and cell-cell interactions of CCDCs family in the development and progression of HCC. The relationship between mutational signatures and CCDCs family were evaluated in HCC patients with somatic mutation.

**Results:** Five CCDC family members (CCDC34, CCDC137, CCDC77, CCDC93 and CCDC21) mRNA expression showed significantly higher in HCC tissues than in normal tissues and high expression levels of these genes predicted poor prognosis in HCC patients. The combined effect analysis of five CCDCs family prognostic markers suggests that the prognosis difference for CCDC family members combination was more significant than that for any individual CCDC family genes. We then developed a risk score model that could predict the prognosis of HCC, and nomogram gene expression was visualized with the probability of predicting the prognosis of HCC by clinical factors. GSEA revealed that, while five CCDCs family combined high expression was associated with increased cell cycle progression and low expression was associated with complement activation pathway. Mutation analysis showed that the combined high expression group had a higher TP53 mutation rate than the combined low expression group, and the high expression group showed higher TMB, which was associated with a better prognosis than high TMB.

**Conclusions:** Our data suggest that the expression of CCDC34, CCDC137, CCDC77, CCDC93 and CCDC21 may be potential prognostic markers in HCC and in combination have a strong interaction and better predictive value for HCC prognosis.

## Background

Liver cancer is one of the most common malignancies worldwide and hepatocellular carcinoma (HCC) is the most frequent type of global cancer mortality rates [1]. HCC is highly malignant, invasive and metastatic, and has a poor prognosis, posing a serious threat to human health [2]. Although there is an increasing number of tumor markers used for cancer diagnosis and prognosis, there is not yet a highly discriminative, sensitive, and specific tumor marker that can be used to evaluate HCC condition, efficacy, and prognosis [3, 4]. Accordingly, there is an increasingly urgent need to identify useful biomarkers that can accurately display the biological features and prognostic outcomes in patients with HCC. Coiled-coil

domain containing (CCDC) protein family is a homo- or oligomeric sequence protein composed of two or more coiled coil domains. Because of CCDCs family special coiled coil structure, it makes its protein prone to spatial conformational changes, thereby exerting multiple cell biological functions. Recently, an increasing number of studies have identified that CCDC proteins play important roles in the development and progression of various malignancies, including thyroid, pancreatic, and breast cancer [5–7]. Moreover, the specific molecular mechanisms of CCDCs family in tumor malignant progression and distal metastasis are being further explored. Studies have reported that abnormal expression of CCDCs family can promote tumor cell motility and migration by regulating the rearrangement of the cytoskeleton and activating the activation of intracellular signaling phospholipase C with mitogen activated proteases [8]. However, the potential application value and mechanism of CCDCs family in prognosis prediction of HCC patients are not fully understood. The main objective of this study was to investigate this linkage by collecting data from public databases and performing a series of bioinformatics analyses.

## Materials And Methods

### Data acquisition and collation

The RNA-sequencing recorded based on FPKM and corresponding clinical data of 371 HCC patients were downloaded from The Cancer Genome Atlas (TCGA, <https://portal.gdc.cancer.gov/repository>) database. The RNA-Seq mRNA expression data and clinical pathological data of 243 samples with HCC patients were available in the International Cancer Genome Consortium (ICGC: <https://dcc.icgc.org/projects/LIRI-JP>) database. Then, coiled-coil domain containing genes (CCDCGs) were retrieved from GeneCards website (<https://www.genecards.org/>) and provided in supplementary table.

### Identification of differentially expressed CCDCGs in HCC patient's prognosis

In both TCGA dataset and ICGC dataset, the expression profile data were  $\log_2(\text{FPKM} + 1)$  transformed to identify the differentially expressed genes between tumor tissues and normal tissues, respectively. The false discovery rate (FDR) less than 0.05 and the  $|\log_2\text{FC}|$  greater than 1 were set as cutoff values. The union of the two datasets was finally taken to screen for genes with differential expression of CCDCGs (DECCDCGs). The DECCDCGs in the two datasets was analyzed with the R studio software (version 3.6.3), using the “limma” package. Univariate Cox analysis of HCC patient's overall survival (OS) of TCGA dataset and ICGC dataset were performed to screen DECCDCGs with prognostic value. To evaluate the prognostic value of CCDCGs, both datasets were screened for differential expression of genes associated with HCC prognosis with a *P*-value less than 0.05.

### Construction and validation of a prognostically relevant DECCDCGs family signature

Using data from the TCGA dataset, a Cox regression model for OS was constructed and further validated in the ICGC dataset. The prognosis related DECCDCGs of the TCGA dataset were subjected to multivariate Cox regression analysis and lasso Cox regression, respectively, to screen candidate genes. After model building, the signature genes involved in the model are further validated in various ways: (1) HCC patients of the TCGA dataset with their corresponding calculated risk score were divided into low-risk and high-risk prognostic groups according to the median value. Survival analyses were performed using Kaplan-Meier curve and log-rank test in TCGA dataset at low- and high-risk. The model was then validated in the ICGC dataset with HCC patient's data. (2) Time-dependent receiver-operating characteristic (ROC) analysis was performed using the "survivalRoc" R package in the TCGA and ICGC datasets, and AUC values were calculated for 1-, 3-, and 5-year survival. Univariate and multivariate Cox analyses were also used to define independent risk variables. Meanwhile, stratified analysis was performed according to clinicopathological characteristics.

## Building and validating a predictive nomogram

The independent prognostic factors were further integrated, and nomograms were established to examine the 1-, 3-, and 5-year predictive abilities of HCC. The nomogram was evaluated and validated by plotting the calibration curve of the nomogram, observing the relationship between nomogram predicted probabilities and observed rates. The nomograms including all independent prognostic factors were subsequently compared using ROC curves.

## Gene set enrichment analysis (GSEA) and HCC tumor microenvironment analysis

GSEA (<http://software.broadinstitute.org/gsea/index.jsp>) was used to determine the enrichment of certain gene ranks in pre-defined biological processes. According to the prognostic model, the HCC samples of TCGA dataset were divided into high-risk group and low-risk group. The potential biological functions of key genes were defined by comparing the enriched biological processes in the two groups. Molecular characterization data (<http://software.broadinstitute.org/gsea/msigdb/index.jsp> MSigDB) c2.cp.kegg.v6.0.symbols.gmt and c5.bp.v7.1.symbols.gmt was selected as the reference gene set in the GSEA software. The threshold of significance was set as  $|NES| > 1$ ,  $FDR < 0.25$ , and  $P \text{ value} < 0.05$ . Stromal score and immune score were performed by estimate algorithm on downloaded TCGA HCC gene expression data.

## Mutation analysis

TCGA data was downloaded from the Genomic Data Commons Data Portal (GDC portal, <https://portal.gdc.cancer.gov/>). The TCGA database included 350 HCC patients were analyzed. The quantity and quality of gene mutations were analyzed in low-group and high-group by using the Maftools

package of R software. tumour mutational burden (TMB) was defined as the number of somatic, coding, base substitution, and indel mutations per megabase (Mb) of genome examined. The exome size was estimated as 38 Mb. For studies reporting mutation number from whole exome sequencing, the normalized TMB = (whole exome non-synonymous mutations)/(38 Mb).

## **Result**

### **Collation of samples and clinical data**

HCC and normal samples in TCGA and ICGC databases were downloaded with 370 tumour tissues and 50 normal tissues in TCGA dataset and 240 tumour tissues and 202 normal tissues in ICGC dataset. Samples without clinical information were excluded for analysis of clinical and survival prognosis. Detailed clinical information of the HCC patients included in this study was shown in Table 1.

Table 1  
Summary of clinical characteristics of patients with HCC each dataset

	TCGA	ICGC	<i>p</i> .overall
	<i>N</i> =365	<i>N</i> =231	
time	596 [344;1091]	780 [510;1095]	0.002
status			<0.001
Aliver	235 (64.4%)	189 (81.8%)	
Dead	130 (35.6%)	42 (18.2%)	
Age:			<0.001
elder	200 (54.8%)	182 (78.8%)	
young	165 (45.2%)	49 (21.2%)	
Gender:			0.130
Female	246 (67.4%)	170 (73.6%)	
Male	119 (32.6%)	61 (26.4%)	
Stage:			0.184
stage I-II	230 (63.0%)	141 (61.0%)	
stage III-IV	130 (35.6%)	90 (39.0%)	
grade:			0.629
G1-G2	248 (67.9%)	152 (65.8%)	
G3-G4	93 (25.5%)	59 (25.5%)	

## Identification of prognostic DECCDCGs in HCC

TCGA data matrix, 93 differentially correlated genes (FDR < 0.05) were obtained, of which 19 were differentially expressed genes that satisfied the  $|\log_2FC| > 1$  condition (Fig. 1a). In the ICGC data matrix, 85 differentially correlated genes (FDR < 0.05) were obtained, of which 19 genes were differentially expressed that satisfied the  $|\log_2FC| > 1$  condition. (Fig. 1b). Twenty-one DECCDCGs were identified at least one dataset in the two datasets ( $|\log_2FC| > 1$  and FDR < 0.05) (Fig. 1c). Seventeen prognostic DECCDCGs were identified in the TCGA database (Fig. 1d), and eight prognostic DECCDCGs were identified in the ICGC database (Fig. 1e). Eight prognostic DECCDCGs were shared between TCGA and ICGC databases (Fig. 1f).

# Construction and validation of a prognostically relevant DECCDCGs signature

The lasso Cox method and the multivariate Cox regression method were applied to select candidate genes using the univariate Cox regression results for the 8 genes in the TCGA dataset. It was found that five OS related prognostic DECCDCGs (OS-DECCDCGs) were obtained by lasso Cox method (Fig. 2a). After extracting coefficients from the results, an individualized risk score was calculated from the coefficient weighted expression levels of the five OS-DECCDCGs as follows:  $e^{(0.094 * \text{expression level of CCDC34} + 0.344 * \text{expression level of CCDC137} + 0.258 * \text{expression level of CCDC77} + 0.047 * \text{expression level of CCDC93} + 0.301 * \text{expression level of CCDC21})}$ . Three OS-DECCDCGs were derived by Cox multivariate analysis. An individualized risk score was calculated from the coefficient weighted expression levels of the 3 OS-DECCDCGs as follows:  $0.4426 * \text{expression level of CCDC137} + 0.3979 * \text{expression level of CCDC21} + 0.3936 * \text{expression level of CCDC77}$ . Based on these prognostic genes, a prognostic risk signature for OS was established. To evaluate the prognostic ability of the risk model, the ICGC dataset was used for model validation. Two model comparison analysis found that the predictive accuracy of the two models were comparable in the TCGA dataset (Fig. 2b), but the model derived by the lasso Cox method in the ICGC dataset had higher predictive accuracy than the model derived by multivariate Cox regression analysis (Fig. 2c).

Based on the five OS-DECCDCGs models, a risk score was calculated and ranked for patients with HCC in the TCGA dataset. Based on the median risk score, the OS prognosis of patients in the TCGA dataset was divided into high- and low-risk groups. Kaplan-Meier and log-rank analysis demonstrated a significant difference in OS between the two risk groups in the TCGA and ICGC datasets ( $P < 0.0001$ ) (Fig. 2d, g). The risk score, survival status, and the five OS-DECCDCGs expression profiles of HCC patients are shown in the TCGA and ICGC datasets (Fig. 2e, h). For OS, the area under the ROC curve (AUC) at 1-, 3- and 5- years were 0.760, 0.693, 0.642 in the TCGA group respectively and 0.755, 0.779, 0.694 in the ICGC group respectively (Fig. 2f, i). To evaluate whether the model was independent of other clinical variables in patients with HCC, univariate and multivariate Cox regression analyses including clinical factors and risk scores were performed. The results showed that the five OS-DECCDCGs combined model remained significantly associated with overall survival after adjusting for tumor grade, clinical stage, age and gender (Table2). The HR for overall survival in the high-risk versus low-risk group was 2.31 (95% CI = 1.56-3.4;  $P < 0.001$ ) in the TCGA dataset and the HR in the ICGC dataset was 2.53 (95% CI = 1.19-5.38;  $P = 0.016$ ). The overall survival of patients with HCC was compared between low- and high-risk groups in the TCGA dataset and ICGC dataset patients, and among each subgroup of the TCGA dataset, patients with a prognostic score classified as low risk had a better OS than patients with a high-risk score ( $P < 0.05$ ). In all subgroups of ICGC, except for early stage and tumor histologic grade G3-G4 stage, the prognostic score was also significantly higher for patients in the low-risk group compared with the high-risk group ( $P < 0.01$ ) (Table 3).

Table 2  
Univariate and multivariate Cox regression analyses in TCGA dataset and ICGC dataset

Characteristics	Univariate analysis		<i>P</i>	Multivariate analysis		<i>P</i>
	HR	95%CI		HR	95%CI	
TCGA dataset						
Age (>60 vs. ≤60)	1.21	0.85-1.72	0.281			
Gender (Male vs. Female)	0.82	0.57-1.16	0.263			
Grade (G3&G4 vs. G1&G2)	1.12	0.89-1.42	0.339	1.23	0.93-1.63	0.143
Stage (Stage I&II vs. Stage III&IV)	1.66	1.36-2.04	<0.001	1.52	1.23-1.88	<0.001
RiskScore (high vs. low)	2.64	1.84-3.79	<0.001	2.31	1.56-3.4	<0.001
ICGC dataset						
Age (>60 vs. ≤60)	0.88	0.43-1.8	0.731			
Gender (Male vs. Female)	0.5	0.27-0.94	0.031	0.35	0.17-0.74	0.005
Grade (G3&G4 vs. G1&G2)	2.57	1.35-4.92	0.004	1.73	0.86-3.51	0.126
Stage (Stage I&II vs. Stage III&IV)	2.49	1.35-4.6	0.003	2.74	1.33-5.65	0.006
RiskScore (high vs. low)	3.82	1.91-7.65	<0.001	2.53	1.19-5.38	0.016

Table 3  
The relationship between the risk model of gene combination and each clinicopathological feature

Clinicopathologic characteristics	classification	low risk (death count)	high risk (death count)	Log-rank $\chi^2$
TCGA dataset				
Age (year)	≤60	85(22)	80(32)	9.390**
	>60	98(26)	102(50)	19.874***
Gender	Male	60(23)	59(28)	7.439**
	Female	123(25)	123(54)	24.828***
Stage	Stage I-II	139(28)	115(40)	16.133***
	Stage III-IV	32(14)	55(34)	5.057*
Grade	G1-G2	133(36)	97(42)	18.457***
	G3-G4	47(10)	83(38)	8.480**
ICGC dataset				
Age (year)	≤60	25(2)	24(8)	7.989**
	>60	91(9)	91(23)	9.916**
Gender	Male	29(3)	32(13)	7.926**
	Female	87(8)	83(18)	8.446**
Stage	Stage I-II	82(8)	59(10)	2.054
	Stage III-IV	34(3)	56(21)	11.166**
Grade	G1-G2	89(8)	63(12)	5.228*
	G3-G4	19(3)	40(14)	3.167
*P<0.05, **P<0.01, ***P<0.001				

## Predictive nomogram construction and performance assessment

In addition, a nomogram was established to predict OS in patients with two independent prognostic factors, including stage and the risk score (Fig. 3a). The calibration curve indicated that the nomogram was able to accurately predict 1-, 3- and 5-year OS (Fig. 3b-d). The combined model shows good accuracy and stability in clinical outcome prediction either for 1-year (AUC=0.778), 3-year (AUC=0.737) and 5-year (AUC=0.701) OS of HCC patients, when compared with single clinical factors (Fig. 3e-g). The results suggested that the nomogram constructed using the combined model might be the best nomogram for predicting the survival of HCC patients compared with the nomogram established using single prognostic factors, which might be helpful for clinical management.

## Gene set enrichment analysis (GSEA) and tumor microenvironment analysis

In addition, GSEA analyzed transcriptional information of patients stratified by risk score into high- and low-risk groups. Both KEGG pathway (Fig. 4a) and GO biological pathway (Fig. 4b) enrichment found that cell cycle was enriched in the high-risk group. Pathways enriched in the low-risk group were complement activation pathway related, suggesting that complement regulatory processes may play a protective role in low-risk patients. The stromal score and immune score based on the ESTIMATE algorithm based on the TCGA HCC gene expression data showed that in the TCGA dataset, patients with HCC were divided into high- and low-risk groups by stromal and immune scores, and the analysis found no differences in immune scores between the high- and low-risk groups ( $p=0.124$ ), but the high-risk group had lower stromal scores than the low-risk group ( $p < 0.05$ ) (Fig. 4c). Patients with a lower stromal score had a worse prognosis (Fig. 4d), and these results suggest that complement may influence tumor growth by altering the tumor microenvironment.

## Mutational profile of CCDCGs related risk signature

TCGA HCC patients assessed the relationship between mutational signatures and signatures using available somatic mutation data. The most frequently mutated genes in the high- and low-risk groups are shown (Fig. 5a is high-risk group and Fig. 5b is low-risk group). TP53, GRIA1 and other genes were found to be differentially mutated genes in high and low risk groups (Fig. 5c), and TP53 was found to be mutated significantly more frequently in high-risk group than in low-risk group (Fig. 5d). High risk patients were found to have a higher TMB in the TCGA somatic mutation data (Fig. 5e), and the higher the mutation rate was found to have a worse prognosis (Fig. 5f).

## Discussion

Coiled coils are a prevalent protein domain, and proteins with coiled coil structure include structural proteins, membrane proteins, enzymes, and transcription factors [9]. The spatial folding of coiled coil domain containing (CCDC) is variable, leading to different spatial conformations, enabling many different

molecular biological functions, including regulation of gene expression, cell division, membrane fusion, and controlled release of drugs [10]. An increasing number of studies in recent years have begun to focus on the role of the CCDC family of proteins in tumors, with studies reporting that abnormal expression of CCDC proteins may promote tumor cell motility and migration by regulating rearrangements of the cytoskeleton [8, 11]. Previous studies have confirmed that the domains of the CCDC gene family are associated with aberrant CCDC associated protein expression for a variety of malignancies, such as non-small cell lung cancer [12], gastric cancer [13], colorectal cancer [14] and hepatocellular carcinoma [15]. In this study, the association between CCDCs and HCC was explored in both TCGA and ICGC databases. Our results showed that high expression levels of CCDC34, CCDC137, CCDC77, CCDC93 and CCDC21 were associated with poor prognosis of HCC patients in TCGA and ICGC databases. In addition, the combination of the five OS-DECCDCGs had a good predictive value in patients with HCC. Therefore, the combination of the five OS-DECCDCGs identified in this study may serve as a potential prognostic biomarker for HCC.

Currently, several CCDCs have been reported to be involved in intracellular signaling, transcription of genetic signals, molecular recognition and cell cycle regulation, cell differentiation, apoptosis and other biological processes [16]. Coiled coil domain containing 34 (CCDC34), also known as NY-REN-41, comprises 373 amino acids and is located on chromosome 11p14.1. Previous studies have found that CCDC34 is overexpressed in several human malignancies, including renal cell carcinoma [17], non-small cell lung cancer (NSCLC) [18], bladder cancer [19], pancreatic adenocarcinoma [20], esophageal squamous cell carcinoma [21], colorectal cancer [22] and cervical cancer [23]. Recently, it was shown [24], that CCDC34 overexpression in liver cancer tissues is associated with poor prognosis of HCC patients. This study confirmed by in vitro and in vivo experiments that knockdown of CCDC34 could effectively inhibit the proliferation and metastasis of HCC cells and found that inhibition of CCDC34 could affect the activation of protein kinase B (PKB or Akt) as well as the epithelial mesenchymal transition (EMT) process. The remaining four CCDCs (CCDC137, CCDC77, CCDC93 and CCDC21) have not been studied experimentally in relation to tumour development, but from the regression models constructed for the five genes as well as the expression profiles of the genes in HCC tissues, they are closely related to HCC and may be involved in the molecular mechanisms of HCC pathogenesis, and are expected to be target genes for future studies.

Studies also performed GSEA analysis to explore the potential mechanism of the high- and low-risk groups of CCDCs combination in HCC prognosis. The results showed that cell cycle related pathways were significantly enriched in the high-risk group and complement activation pathways were significantly enriched in the low-risk group, indicating that complement regulatory processes may play a protective role in low-risk patients. TP53 is one of the most frequently mutated genes in HCC. Considerable studies have revealed the role of TP53 as a biomarker for certain molecular features and a prognostic factor for unfavorable survival in HCC [25, 26]. In this study, the mutation rates of TP53 in high- and low-risk groups were compared, and it was found that high-risk group had a higher TP53 mutation rate than low-risk group, high-risk group showed higher TMB, lower TMB had a better prognosis compared with high TMB, it can be considered that HCC patients with lower TMB provide a potential benefit in prognosis.

## Conclusions

In this study, we investigated the relationship between the expression of individual CCDCGs and the OS of HCC patients. In addition, we investigated the joint effects of the five OS-DECCDCGs, constructed an effective CCDCGs-related risk signature model and constructed a nomogram to predict the prognosis of HCC patients. Our model combination was statistically significant in predicting the prognosis of HCC. In addition, we also investigated the potential mechanism of CCDCGs in the prognosis of HCC patients using GSEA method. These results provide insight into the CCDCGs for cancer clinical outcomes that may serve as biomarkers for predicting the prognosis of HCC patients. There are also some limitations in our study: basic experiments are needed to validate the CCDCGs prognostic signature; further studies with larger sample sizes are needed in the future for our prognostic signature and its relationship with the CCDCGs.

## Abbreviations

HCC: hepatocellular carcinoma; CCDC: Coiled-coil domain-containing; TCGA: The Cancer Genome Atlas; ICGC: International Cancer Genome Consortium; CCDCGs: coiled-coil domain containing genes; FDR: false discovery rate; DECCDCGs: differential expression of CCDCGs; OS: overall survival; OS-DECCDCG: OS related prognostic DECCDCGs; ROC: receiver-operating characteristic; GSEA: Gene set enrichment analysis; C-index: Concordance index; TMB: tumor mutational burden.

## Declarations

## Acknowledgements

We would like to thank all the authors listed for their contribution to the present study.

## Authors' contributions

SNZ, CJ and YW provided the scientific idea and designed the study. SNZ, CJ, YW, and GC performed all data analysis, designed the figures, and drafted the manuscript. YY, HQL, and JM performed preliminary data processing and modified the manuscript. All authors read and approved the final manuscript.

## Funding

This work was supported in part by the Natural Science Foundation of Zhejiang Province, China (LY17H160047); Public Projects of Zhejiang Province, China (2016C37127 and 2018C37114); National Natural Science Foundation of China (81470868, 81772628 and 81703310); Basic Projects of Wenzhou Science and Technology Bureau, China (Y20190206).

# Availability of data and materials

All data analyzed and displayed in the present manuscript are available from the corresponding author upon reasonable request.

# Ethics approval and consent to participate

Not applicable.

# Consent for publication

Not applicable.

# Competing interests

The authors declare that they have no competing interests.

# Authors' information

1 Key Laboratory of Diagnosis and Treatment of Severe Hepato-Pancreatic Diseases of Zhejiang Province, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325000, China; 2 School of Public Health, Inner Mongolia Medical University, Hohhot 010110, China; 3 Department of Epidemiology and Biostatistics, School of Public Health and Management, Wenzhou Medical University, Wenzhou 325000, China; 4 Department of Hepatobiliary Surgery, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325035, China

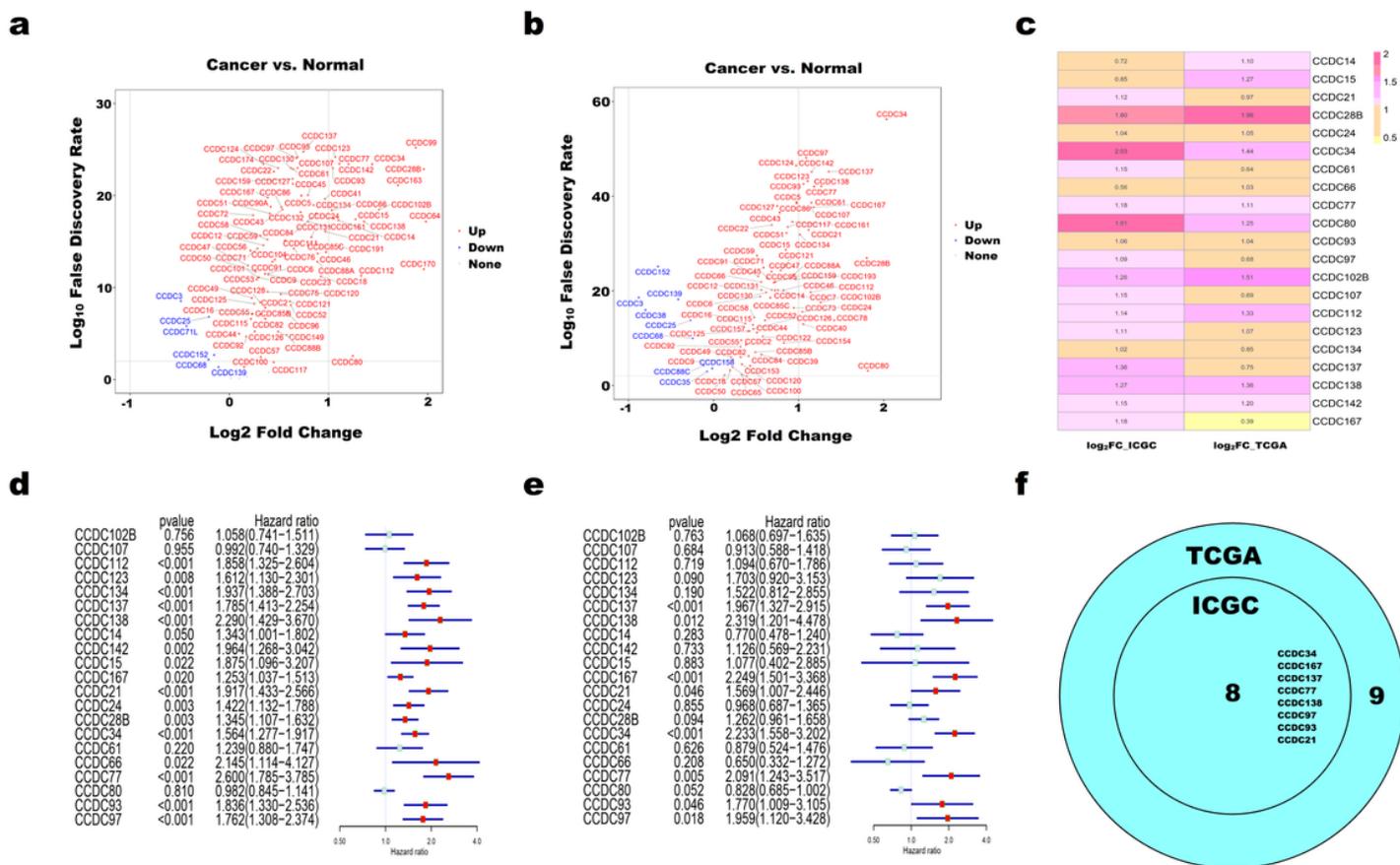
# References

1. Siegel RL, Miller KD, Jemal A: **Cancer statistics, 2020**. *CA Cancer J Clin* 2020, **70**(1):7-30.
2. Cronin KA, Lake AJ, Scott S, Sherman RL, Noone AM, Howlader N, Henley SJ, Anderson RN, Firth AU, Ma J *et al*: **Annual Report to the Nation on the Status of Cancer, part I: National cancer statistics**. *Cancer* 2018, **124**(13):2785-2800.
3. Li L, Wang H: **Heterogeneity of liver cancer and personalized therapy**. *Cancer Lett* 2016, **379**(2):191-197.
4. Dawkins J, Webster RM: **The hepatocellular carcinoma market**. *Nat Rev Drug Discov* 2019, **18**(1):13-14.
5. Yin DT, Xu J, Lei M, Li H, Wang Y, Liu Z, Zhou Y, Xing M: **Characterization of the novel tumor-suppressor gene CCDC67 in papillary thyroid carcinoma**. *Oncotarget* 2016, **7**(5):5830-5841.

6. Tanouchi A, Taniuchi K, Furihata M, Naganuma S, Dabanaka K, Kimura M, Watanabe R, Kohsaki T, Shimizu T, Saito M *et al*: **CCDC88A, a prognostic factor for human pancreatic cancers, promotes the motility and invasiveness of pancreatic cancer cells.** *J Exp Clin Cancer Res* 2016, **35**(1):190.
7. Jiang P, Li Y, Poleshko A, Medvedeva V, Baulina N, Zhang Y, Zhou Y, Slater CM, Pellegrin T, Wasserman J *et al*: **The Protein Encoded by the CCDC170 Breast Cancer Gene Functions to Organize the Golgi-Microtubule Network.** *EBioMedicine* 2017, **22**:28-43.
8. Liu Z, Mai C, Yang H, Zhen Y, Yu X, Hua S, Wu Q, Jiang Q, Zhang Y, Song X *et al*: **Candidate tumour suppressor CCDC19 regulates miR-184 direct targeting of C-Myc thereby suppressing cell growth in non-small cell lung cancers.** *J Cell Mol Med* 2014, **18**(8):1667-1679.
9. McFarlane AA, Orriss GL, Stetefeld J: **The use of coiled-coil proteins in drug delivery systems.** *Eur J Pharmacol* 2009, **625**(1-3):101-107.
10. Burkhard P, Stetefeld J, Strelkov SV: **Coiled coils: a highly versatile protein folding motif.** *Trends Cell Biol* 2001, **11**(2):82-88.
11. Maeyama Y, Otsu M, Kubo S, Yamano T, Imura Y, Onodera M, Kondo S, Sakiyama Y, Ariga T: **Intracellular estrogen receptor-binding fragment-associated antigen 9 exerts in vivo tumor-promoting effects via its coiled-coil region.** *Int J Oncol* 2011, **39**(1):41-49.
12. Jiang GY, Zhang XP, Zhang Y, Xu HT, Wang L, Li QC, Wang EH: **Coiled-coil domain-containing protein 8 inhibits the invasiveness and migration of non-small cell lung cancer cells.** *Hum Pathol* 2016, **56**:64-73.
13. Park SJ, Jang HR, Kim M, Kim JH, Kwon OH, Park JL, Noh SM, Song KS, Kim SY, Kim YH *et al*: **Epigenetic alteration of CCDC67 and its tumor suppressor function in gastric cancer.** *Carcinogenesis* 2012, **33**(8):1494-1501.
14. Jun BY, Kim SW, Jung CK, Cho YK, Lee IS, Choi MG, Choi KY, Oh ST: **Expression of girdin in human colorectal cancer and its association with tumor progression.** *Dis Colon Rectum* 2013, **56**(1):51-57.
15. Hu X, Zhao Y, Wei L, Zhu B, Song D, Wang J, Yu L, Wu J: **CCDC178 promotes hepatocellular carcinoma metastasis through modulation of anoikis.** *Oncogene* 2017, **36**(28):4047-4059.
16. Lupas AN, Gruber M: **The structure of alpha-helical coiled coils.** *Adv Protein Chem* 2005, **70**:37-78.
17. Scanlan MJ, Gordan JD, Williamson B, Stockert E, Bander NH, Jongeneel V, Gure AO, Jager D, Jager E, Knuth A *et al*: **Antigens recognized by autologous antibody in patients with renal-cell carcinoma.** *Int J Cancer* 1999, **83**(4):456-464.
18. Petroziello J, Yamane A, Westendorf L, Thompson M, McDonagh C, Cervený C, Law CL, Wahl A, Carter P: **Suppression subtractive hybridization and expression profiling identifies a unique set of genes overexpressed in non-small-cell lung cancer.** *Oncogene* 2004, **23**(46):7734-7745.
19. Gong Y, Qiu W, Ning X, Yang X, Liu L, Wang Z, Lin J, Li X, Guo Y: **CCDC34 is up-regulated in bladder cancer and regulates bladder cancer cell proliferation, apoptosis and migration.** *Oncotarget* 2015, **6**(28):25856-25867.
20. Qi W, Shao F, Huang Q: **Expression of Coiled-Coil Domain Containing 34 (CCDC34) and its Prognostic Significance in Pancreatic Adenocarcinoma.** *Med Sci Monit* 2017, **23**:6012-6018.

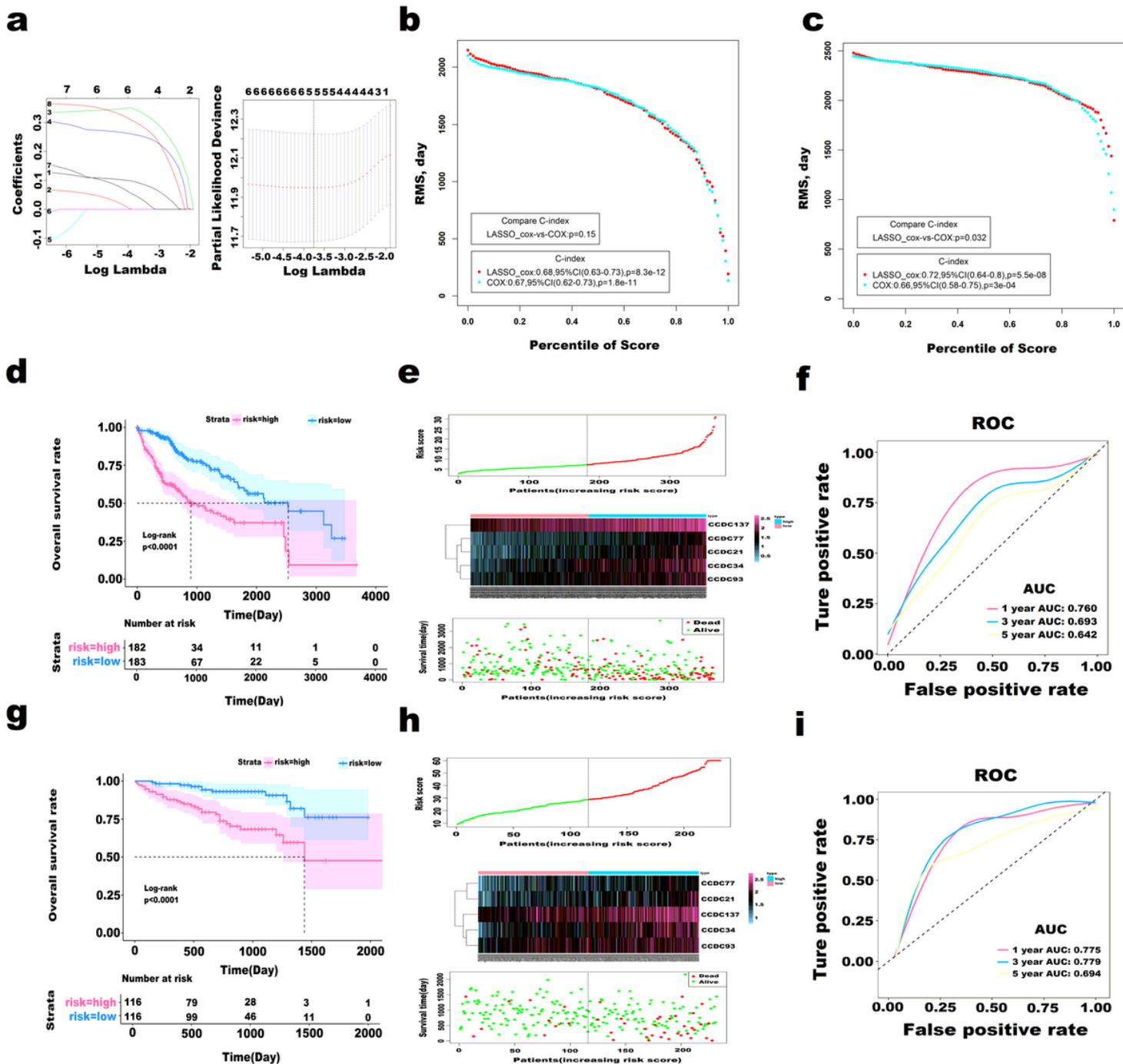
21. Hu DD, Li PC, He YF, Jia W, Hu B: **Overexpression of Coiled-Coil Domain-Containing Protein 34 (CCDC34) and its Correlation with Angiogenesis in Esophageal Squamous Cell Carcinoma.** *Med Sci Monit* 2018, **24**:698-705.
22. Geng W, Liang W, Fan Y, Ye Z, Zhang L: **Overexpression of CCDC34 in colorectal cancer and its involvement in tumor growth, apoptosis and invasion.** *Mol Med Rep* 2018, **17**(1):465-473.
23. Liu LB, Huang J, Zhong JP, Ye GL, Xue L, Zhou MH, Huang G, Li SJ: **High Expression of CCDC34 Is Associated with Poor Survival in Cervical Cancer Patients.** *Med Sci Monit* 2018, **24**:8383-8390.
24. Lin Z, Qu S, Peng W, Yang P, Zhang R, Zhang P, Guo D, Du J, Wu W, Tao K *et al*: **Up-Regulated CCDC34 Contributes to the Proliferation and Metastasis of Hepatocellular Carcinoma.** *Onco Targets Ther* 2020, **13**:51-60.
25. Zhan P, Ji YN, Yu LK: **TP53 mutation is associated with a poor outcome for patients with hepatocellular carcinoma: evidence from a meta-analysis.** *Hepatobiliary Surg Nutr* 2013, **2**(5):260-265.
26. Long J, Wang A, Bai Y, Lin J, Yang X, Wang D, Yang X, Jiang Y, Zhao H: **Development and validation of a TP53-associated immune prognostic model for hepatocellular carcinoma.** *EBioMedicine* 2019, **42**:363-374.

## Figures



## Figure 1

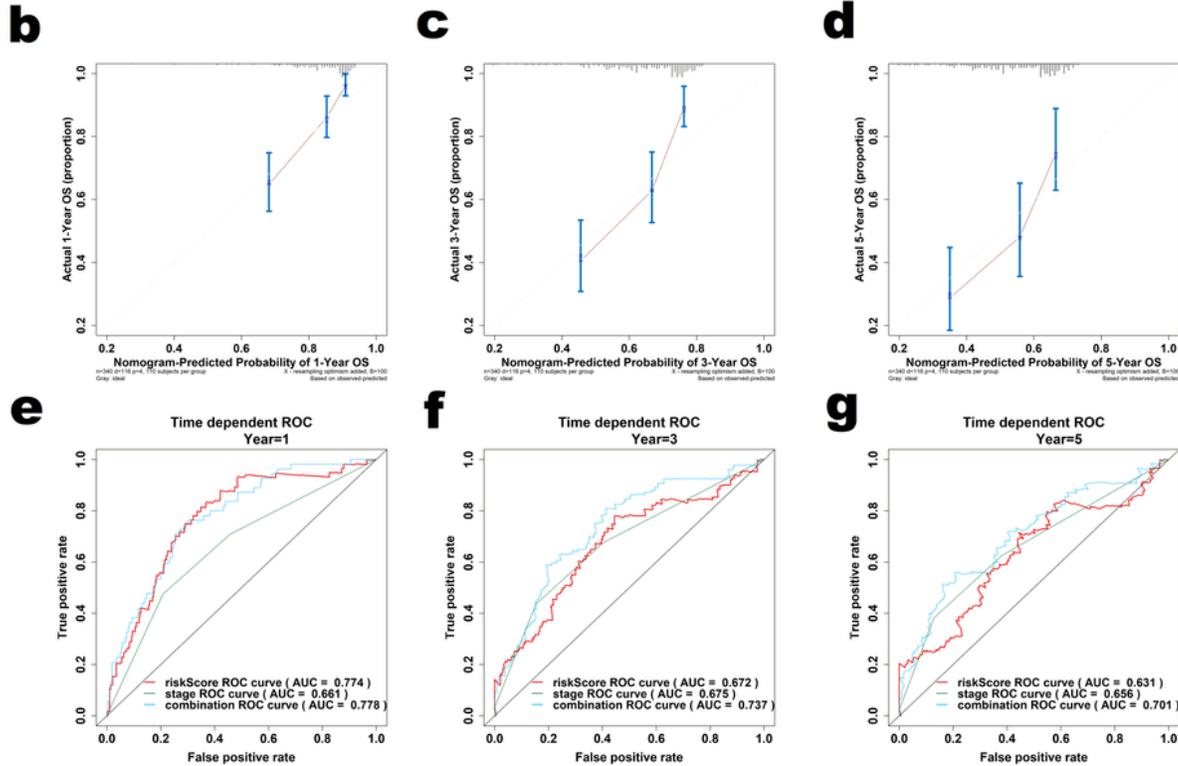
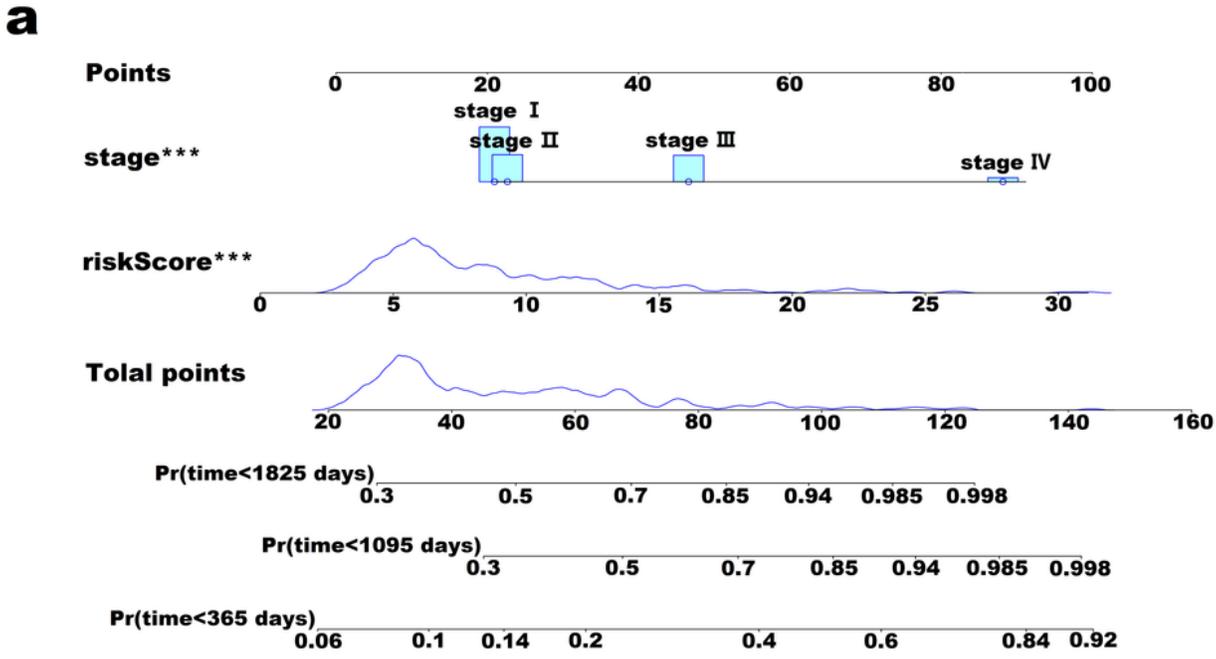
Identification of prognostic DECCDCGs in HCC. a The volcano plot for the 93 DECCDCGs in the TCGA dataset. b The volcano plot for the 85 DECCDCGs in the TCGA dataset. The X-axis represents  $-\log_{10}(\text{FDR})$ . The Y-axis represents  $\log_2$ -transformed fold change. The red dot indicates differently expressed upregulated genes and blue dot indicates differently expressed downregulated genes. c Heatmap demonstrating DECCDCGs in the union of the TCGA and ICGC dataset. Red represents  $|\log_2\text{FC}| > 1$ , yellow represents  $|\log_2\text{FC}| < 1$ . d Results of the univariate Cox regression analyses regarding OS in the TCGA. e Results of the univariate Cox regression analyses regarding OS in the ICGC. f Prognostic DECCDCGs common to TCGA and ICGC databases.



**Figure 2**

Construction and validation of a prognostically relevant DECCDCGs signature. a A prognostic index containing five DECCDCGs was constructed. Partial likelihood deviation map and the lines represent the coefficient of Lasso. Log Lambda represents the calculation method of random sampling method. b The RMS (Restricted mean survival time) curve for CCDCGs pair index COX model and LASSO-cox model was plotted for TCGA. c The RMS curve for CCDC family gene pair index COX model and LASSO-cox model was plotted for ICGC. Each point represents the RMS time of corresponding COX model and LASSO-cox

model risk scores. The RMS curves showed a larger slope in all data sets for LASSO-cox method, indicating the superior estimation of survival with LASSO-cox method. Concordance index (C-index) for COX model and LASSO-cox model was also provided. P value represents the difference between COX model and LASSO-cox model in terms of C-index. d, g The overall survival time of patients in the high-risk group was also shorter than that in the low-risk group in both the TCGA dataset and the ICGC dataset ( $p < 0.05$ ). e, h Predict the grade of the index and population distribution in the TCGA dataset and ICGC dataset, the survival status of different groups of patients, including the heat map of gene expression characteristics. f, i In the TCGA dataset and the ICGC dataset, ROC curve validation of the prognostic value of the prognostic index.



**Figure 3**

Predictive nomogram construction and performance assessment. a For each patient, three lines are drawn upward to determine the points received from the three predictors in the nomogram. The sum of these points is located on the “Total Points” axis. Then a line is drawn downward to determine the possibility of 1-, 3-, and 5-year overall survival of HCC. b-d The calibration plot for internal validation of the nomogram of 1-, 3-, and 5-year overall survival of HCC. The Y-axis represents actual survival, and the

X-axis represents nomogram-predicted survival. e.g The time-dependent ROC curves of the nomograms compared for 1-, 3-, and 5-year OS in HCC, respectively.

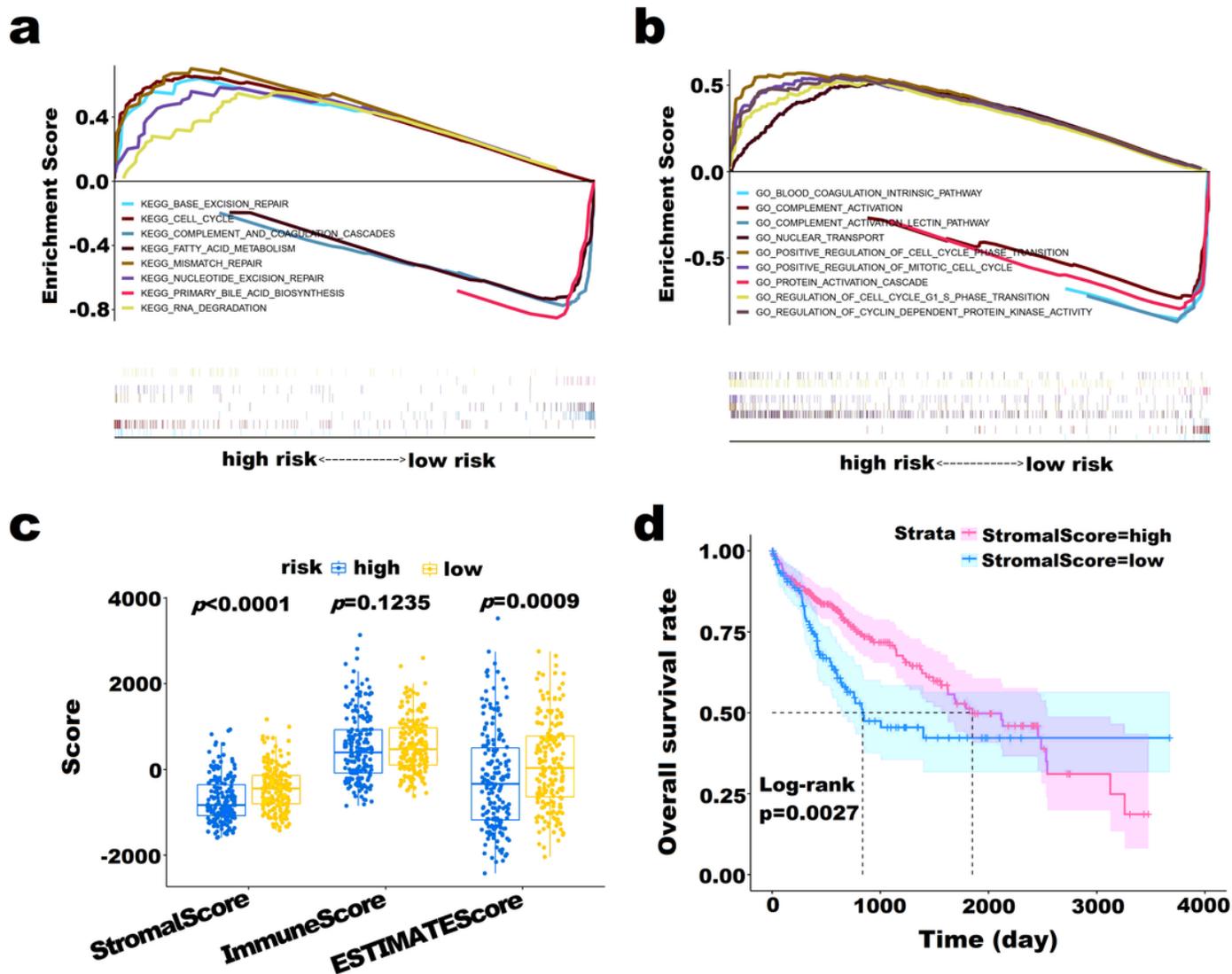


Figure 4

Gene set enrichment analysis (GSEA) and tumor microenvironment analysis. a GSEA showed the significantly enriched KEGG pathways in TCGA. b GSEA showed the significantly enriched GO Biology Process pathways in TCGA. c Relationship between five OS-DECCDCGs risk signature and ESTIMATE's stromal and immune scores. d The association of StromalScore and OS in the TCGA HCC dataset.

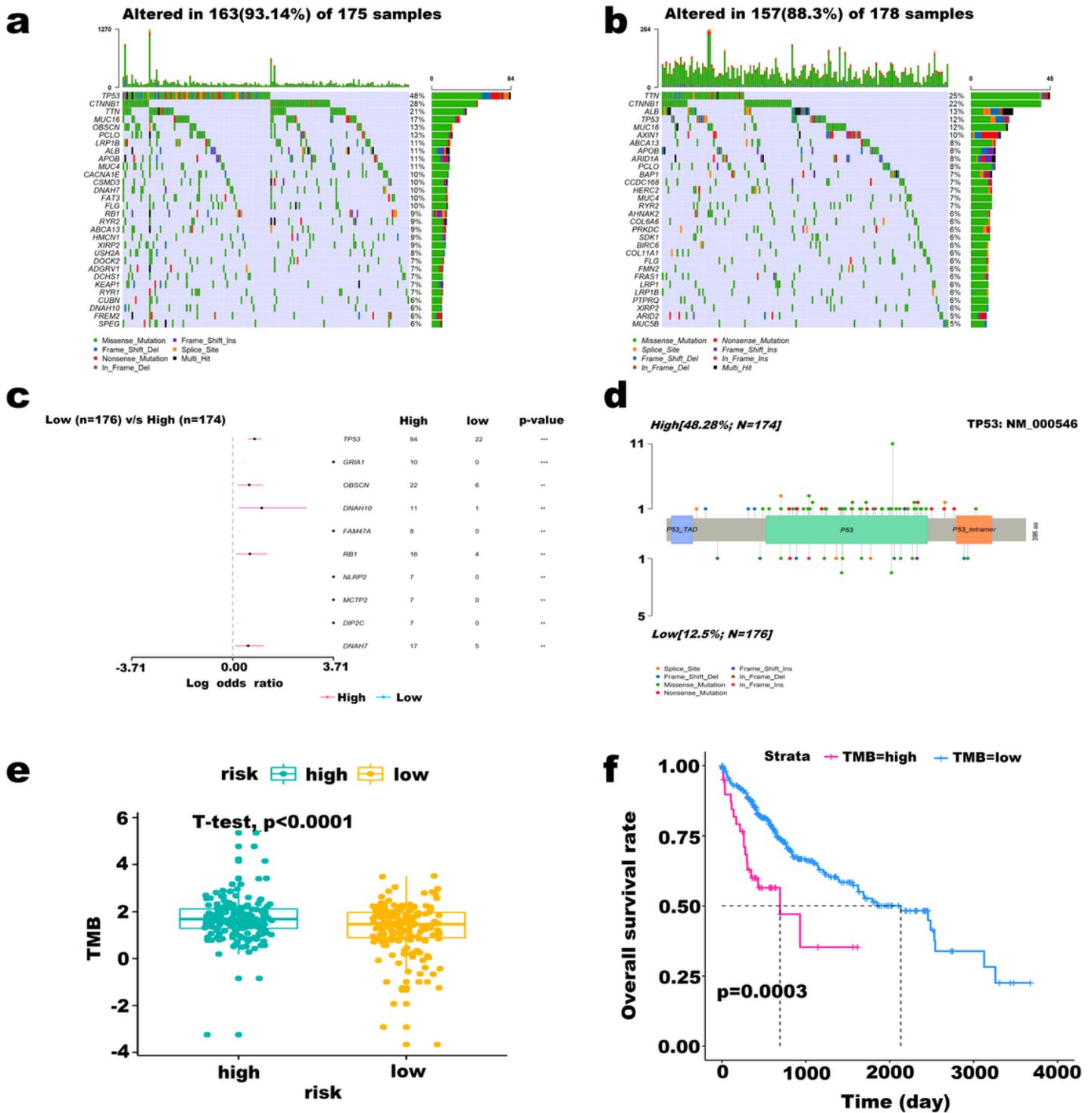


Figure 5

Mutational profile of CCDC family related risk signature. a Mutation profile of high-risk groups. b Mutation profile of low-risk groups. c The forest plot and oncoplot showed the differentially mutated genes between CCDCGs risk signature of the TCGA-HCC dataset. d Lollipop plot compares TP53 mutation sites and types in high- and low-risk groups. e The relationship between the CCDCGs risk signature and TMB in the TCGA HCC dataset. f The association of TMB and OS in the TCGA HCC dataset.

## Supplementary Files

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

- [supplementarytable.docx](#)