Identification of the Signature Associated With m^{6}A RNA Methylation Regulators and m^{6}A-related Genes and Construction of the Risk Score for Prognostication in Early-stage Lung Adenocarcinoma

Bingzhou Guo
Harbin Normal University
Hongliang Zhang
Chinese PLA General Hospital
Jinliang Wang
Chinese PLA General Hospital
Rilige Wu
Chinese PLA General Hospital
Junyan Zhang
Chinese PLA General Hospital
Qiqin Zhang
Weifang traditional chinese hospital
Lu Xu
Chinese PLA General Hospital
Ming shen
Chinese PLA General Hospital
Zhibo Zhang
Chinese PLA General Hospital
Fangyan Gu
Chinese PLA General Hospital
Weiliang Zeng
Harbin Normal University
Chengliang yin (chengliangyin@163.com)
PLA

Research

Keywords: Lung adenocarcinoma, m6A RNA methylation regulators, m6A-related genes, prognostic signature

Posted Date: December 15th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-125302/v1

Abstract

Background: N^6-methyladenosine (m^6A) RNA modification play critical roles in tumorigenesis because it can change gene expression and even the function in multiple levels including the regulation of degradation, subcellular localization, splicing and local conformation changes of RNA transcripts. In this study, we aim to conduct comprehensive investigation on m6A RNA methylation regulators and m6A-related genes and their association with prognosis in early-stage Lung adenocarcinoma (LUAD).

Methods: The relevant datasets which were used to analyze 21 m6A RNA methylation regulators and 887 m6A-related genes in m6Avar were downloaded from Gene Expression Omnibus database (GEO) and The Cancer Genome Atlas (TCGA) databases. Univariate cox regression analysis, random survival forest analysis, Kaplan-Meier anylysis, STRING and multivariate cox analysis were carried out on the datasets, and a risk prognostic model based on five feature genes was constructed.

Results: Respectively, we treated GSE31210 (n=226) as training set, GSE50081 (n=128) and TCGA data (n=461) as test set. By performing univariable cox regression and random survival forest algorithm in the training group, five prognosis-related genes including ***DENND1A, KBTBD6, KIF4A, BMPER,*** and ***YTHDC2*** were screened out, which could divide LUAD patients into low-risk group and high-risk group (log rank $P < 0.001$). The predictive efficacy of these genes was confirmed in the test group GSE50081 (log rank $P < 0.01$) and TCGA datasets (log rank $P < 0.001$). Cox analysis showed that this five-gene signature was an independent risk factor in LUAD. Further, genes in the signature were also external validated using online database. YTHDC2 played vital role of readers in m^6A methylation.

Conclusion: The findings of this study suggested that associated with m^6A-related genes and m^6A RNA methylation regulators, five-gene signature was reliable prognostic indicator for LUAD patients, indicating a clinical application prospect to serve as a potential therapeutic target.

Backgroud

According to Global Cancer Statistics 2018, there will be approximately 18.1 million new cancer cases and 9.6 million deaths worldwide(1). Researchers around the world have been studing to improve medical technology to provide more sensitive diagnosis and operative treatment of tumors. However, due to the complexity of tumor formation mechanisms, it is far from enough to understand the nature of cancer at the genetic level in the traditional sense(2). It should be recognized that the expression of proto-oncogenes depends not only on the genes themselves, but also on epigenetic modifications without altering the gene sequence(2, 3).

Epigenetics is a research hotspot in recent years. It is defined as no change in DNA sequence but heritable change in gene expression(3). Previously, epigenetic researchers focused on DNA and histone modifications. It has even been suggested, for example, mRNA was only useful for messaging. While,with the rapid development of high-throughput sequencing technology, it was found that mRNA also

underwent various modifications during exon splicing, such as -methyladenosine (m6A),N^1-methyladenosine (m^1A) and pseudouridine methylation 5'- hat and 3'- tail(4–6). Among these modifications, m^6A RNA methylation, which was widely found in the mRNA, lncRNA as well as miRNA, was recognized as the most prominent and abundant form of internal modifications in eukaryotic cells, of whose abundance account for 0.1–0.4% total adenosine residues(7, 8). These modifications affected mRNA splicing, nucleation, stabilization, translation, and other mRNA metabolism processes in regulating gene expression. To date, 171 RNA modifications have been identified(9). Research on epigenetic modification of m^6A is increasing. M^6A is methylated to RNA adenine (A) and is one of the most abundant variants of most eukaryotic mRNAs and long chain non-coding RNA (lncRNAs)(10).M^6A methylation is also detected in tRNA, RIbosomal RNA (rRNA), micrornas. Similar to DNA and protein modifications, m^6A methylation is dynamically and reversibly regulated by methyltransferase ("author"), binding protein ("reader") and demethylase ("eraser")(11). RNA was subjected to methylation in the presence of "authors" such as METTL3, METTL14, WTAP, RBM15, KIAA1429, and ZC3H13(12–18). Then "reader" includes YTHDF1, YTHDF2, YTHDC1, YTHDC2, and HNRNPC recognized these m^6A in RNA processing, nuclear output(11, 16, 19). Depending on the "Erasers" (FTO, ALKBH5), the m^6A is restored to adenosine and thus to achieve demethylation modification(20). Once the modification is lost, physiological functions such as cell differentiation and embryonic development are affected and gene expression is abnormally regulated(4, 21) .

In previous studies, most of the researchers only considered the prognostic effect of m^6A-related genes in LUAD. In recent study, they barely focused on the combination of m^6A RNA methylation and m^6A-related genes and their roles for prognostic analysis in LUAD(22–24).

Therefore, based on the GEO and TCGA databases, we conducted a comprehensive analysis to disclose the correlation between mRNA methylation and clinical characteristics in LUAD patients. In this study, we evaluated the relationship between gene expression profile of 21 m^6A-related genes, 883 m^6A RNA methylation, and malignant clinical features. Five genes were identified, which are useful for identifying novel therapeutic targets genes and prognostic stratification and treatment strategy development.

Methods

Expression data

M^6A-related genes, m^6A RNA methylation regulators expression data and clinical data of LUAD patients were obtained from the publicly available GEO database (https://www.ncbi.nlm.nih.gov/geo/). The clinical characters and RNA-seq data of LUAD patients were obtained from the available GEO databases (GSE31210) and the LUAD cases with clinical survival information including survival status and survival time were selected for building the prognostic model as the training set. Another two mRNA expression datasets and corresponding clinical datasets used in this study were obtained from the public GEO (GSE50081) and TCGA (http://cancergenome.nih.gov/). Clinical details of LUAD patients in the training

set(GSE31210) and test set (GSE50081) were shown in Table 1. We selected samples with survival time less than five years in TCGA dataset for prognostic analysis, resulting in 461 samples, which were shown also in Table 1. In the GEO data, samples were mainly in clinical stage and .

Table 1
clinical characteristics of the GEO and TCGA datasets

Characteristic	GSE31210	GSE50081	TCGA
Age (years)			
> 61	122	104	290
≤ 61	104	24	171
Sex			
Female	121	63	253
Male	105	65	208
Vital status			
Alive	191	76	287
Dead	35	52	174
Pathological stage			
Stage	168	92	242
Stage	58	36	113
Stage	0	0	76
Stage	0	0	24
Unknown	0	0	6

Selection of m^6A RNA methylation regulatory factors and m^6A-related genes

The 21 m^6A RNA methylation regulators were collected(25) (Table S1). Fifteen of them were expressed in GSE31210. According to the different roles in the methylation process, they were divided into three types: methyltransferase (writers); binding protein (readers); and demethylase (erasers). A total of 3413 m^6A-related genes related to LUAD were identified in the m6Avar database (http://m6avar. renlab.org/)(26), and 883 of the m^6A-related genes were expressed in the GEO database. All these two data sets (m^6A RNA methylation regulators, and m^6A RNA methylation related genes associated with LUAD) were integrated, including 897 candidate genes.

Establishment of the m ^{6}A-related genes and m6A RNA methylation regulators risk score model

Using univariable cox regression analysis and receiver operating characteristic (ROC) curve analysis, we identified the genes significantly associated with patients' overall survival (OS) in the training group (GSE31210). Then we reduced the number of the genes by random survival forest algorithm (RSFVH). prognostic model was constructed as follows:

$$\textbf{Risk Score} = \sum_{i=1}^{N} \text{Expi} \times Coefi$$

where N is the number of gene, Exp is the gene expression value and Coef is coefficient of the gene expression in cox regression analysis. The final prognostic gene signature was screened out with the largest area under curve (AUC) value in all the constructed models. Each patient was assigned 1023 risk scores, since ten genes form 2^10−1 = 1023 combinations.

According to the median risk score, all of the samples were divided into low and high subtypes. Then the 'survival' package of R software was used to analyze the survival of the low and high subtypes. In additional, the Kaplan-Meier method was employed to examine the survival curves and compare the difference in survival across different scoring subgroups.

Estimation of outcome signature for patients' prognosis and its relationship with clinical characteristics

To investigate the associations between the signature and clinical, other molecular characteristics, the chi-square test was used for categories of variables and the Wilcoxon rank-sum test was usable for continuous variables, respectively. The Kaplan-Meier method and log-rank test were utilized to estimate the association between the signature and OS. The proportional hazards assumption was verified for each variable before fitting Cox models. Multivariate Cox proportional hazards models were utilized to study the association between OS and factors (the signature, age, sex, and pathological stage) in training and test datasets, showing the P value, HR and 95% CI of each variable through 'forestplot' R package.

External validation of the genes in the gene signature

The expression of the genes in the gene signature were further validated at the mRNA level (The Oncomine database, https://www.oncomine.org/resource/main.html; TIMER database, https://cistrome.shinyapps.io/timer/, and GEPIA database, Gene Expression Profiling Interactive Analysis, http://gepia.cancer-pku.cn/index.html), and at the protein level (The Human Protein Atlas database,

http://www.proteinatlas.org). The cBioPortal for Cancer Genomics (http://cbioportal.org) was explored to investigate the genetic alterations of the prognostic genes in the gene signature.

Statistical analysis

Kaplan–Meier analysis was used to assess the three survival risk groups (GSE31210, GSE50081 and TCGA) seperated by the median risk score. Cox regression analysis was performed to explore the independence of the signature. ROC and TimeROC were used to analyse survival prediction performance. Function prediction of prognostic genes was analyzed by clusterProfiler. R program (www.r-project.org) with packages including pROC, TimeROC, clusterProfiler, randomForestSRC and survival were used to perform the above analyses.

Results

Patient population

All 226 ,128 and 461 patients diagnosed with LUAD were collected from the GEO (GSE31210, GSE50081) and TCGA database, respectively. A total of 897 m^6A-related genes out of 9057 expressed genes were identified in GSE31210 dataset. From Table 1, the median age of the enrolled patients was 61 years. The ratio of male vs female was 1.15:1, with 191 live cases and 35 death cases. The longest survival was 10 years. Gene expression data was mainly distributed in stage I-II of LUAD.

Construction of the risk score modle the m ^{6} A-related genes and m ^{6} A RNA methylation regulators risk score

Performing the cox regression and ROC analysis, a total of 129 genes were discovered, which were significantly associated with OS and had a good ability to predict survival ($P < 0.05$, $AUC > 0.6$, Table S2). Further, we screened out ten prognostic genes by RSFVH analysis based on importance scores (Fig. 1a-b). Then we brought the prognostic genes into the risk prediction model and got 2^{10}-1 = 1023 possible signatures in the training dataset. ROC analyses were performed in all the 1023 signatures to find out the signature with the strongest predictive ability (Table S3). The final signature including five genes (DENND1A, KBTBD6, KIF4A, BMPER, YTHDC2) was screened out with the maximum AUC (AUC signature = 0.762; Fig. 1c; Table 2). The selected risk model is as follows: Risk score = (1.664 × expression value of DENND1A) + (-1.249 × expression value of KBTBD6) + (1.736 × expression value of KIF4A)+ (-1.721 × expression value of BMPER) + (-2.015 × expression value of YTHDC2)

Table 2
Prognosis of five genes in the signature

Database ID	Gene symbol	Gene name	Coefficient	P	Expression with Poor prognosis
ENSG00000119522	DENND1A	DENN domain containing 1A	1.66	0.00	high
ENSG00000165572	KBTBD6	kelch repeat and BTB domain containing 6	-1.25	0.00	low
ENSG00000090889	KIF4A	kinesin family member 4A	1.74	0.00	high
ENSG00000164619	BMPER	BMP binding endothelial regulator	-1.72	0.00	low
ENSG00000047188	YTHDC2	YTH domain containing 2	-2.02	0.00	low

The validation of performance in predicting overall survival

We used the risk model to calculate the risk score for each patient. The median risk score divided patients of the training dataset into either the high-risk (n = 113) or low-risk group (n = 113). The Kaplan–Meier analysis results showed patients in the low-risk group lived longer than patients in the high-risk group ($P < 0.001$; Fig. 2a). Then we tested the prognostic value of the gene signature in test dataset (n = 128). Kaplan–Meier analysis found the survival of patients with high risk scores was lower than that of patients with low risk scores in test group ($P = 0.009$; Fig. 2b). In independent data (TCGA, n = 461), the survival of patients with high risk scores was lower than that of patients with low risk scores in TCGA-test group ($P < 0.001$; Fig. 2c). The relationship of gene expression, risk score and survival information was showed in Fig. 3. With the increase of risk scores, death toll raised both in the training set (Fig. 3a) and test set (Fig. 3b-c).

We found that the 5-year survival rate in the training group (GSE31210) was 57.52% in the low-risk group and 33.63% in the high-risk group (Fig. 2a). The overall survival rate was 45.58%, indicating that this model feature could basically differentiate the data better. Survival was also significantly improved in two independent data validation groups (GSE50081 and TCGA). In GSE50081, the low risk group was 50% and the high risk group was 35.94% (Fig. 2b). The overall survival rate was 42.97% and the grouping label was also obvious. In TCGA, we selected sample data of survival time less than 5 years (a total of 461 cases), so we saw the 3 year survival rate, respectively. In low-risk group was 22.08% and in high-risk survival rate was 14.35%(Fig. 2c). The overall survival rate was 18.22%. Through the analysis, we found that the queue sample, the effect of the model had subsided, but the original caused the change of survival rate was not high.

By using timeROC in five-year survival circumstances, we found that the label had a very good prediction effect (Fig. 2d-f). In the training group GSE31210, the AUC value is 0.799. In the validation group

GSE50081, AUC value is 0.619. In the TCGA, we chose timeROC analysis of 4.9 years (because of the selection of data was not more than 5 years of patient samples), and the AUC is 0.716.

The relationship between the signature and clinical characteristics

The correlations were further explored between the prognostic model and various clinical features. In the training and test datasets, by chi-square test, and the gene signature was related with Pathological_stage (Table 3). Then we further performed univariate and multivariable Cox regression analysis to test the predictive independence of the gene signature (Table 4). Multivariable Cox regression results verified that the gene signature was an independent predictive factor and could independently predict patients' clinical outcome in training and test datasets (High- vs. Low-risk, GSE31210, HR = 17.48, 95% CI 4.16–73.53, P < 0.001, n = 226; GSE50081, HR = 1.86, 95% CI 1.05–3.30, P = 0.03, n = 128; TCGA, HR = 1.65, 95% CI 1.20–2.23, P = 0.0019, n = 461,Table 4).

Table 3
Clinical information and signature Chi-square table

GSE31210 dataset (n = 226)				
Variables	Status	low	high	*P*
Age				1.00
	≤ 61	61	61	
	> 61	52	52	
Gender				0.06
	Female	68	53	
	Male	45	60	
Pathological_stage				0.00
	1	95	73	
	2	18	40	
Signature				0.00
	Low expression	113	0	
	High expression	0	113	
GSE50081 dataset (n = 128)				
Age				1.00
	≤ 61	12	12	
	> 61	52	52	
Gender				0.29
	Female	35	28	
	Male	29	36	
Pathological_stage				0.03
	1	52	40	
	2	12	24	
Signature				0.00
	Low expression	64	0	
	High expression	0	64	
TCGA dataset (n = 461)				

GSE31210 dataset (n = 226)				
Age				0.01
	≤ 61	72	99	
	> 61	159	131	
Gender				0.00
	Female	144	109	
	Male	87	121	
Pathological_stage				0.01
	0	4	2	
	1	140	102	
	2	49	64	
	3	29	47	
	4	9	15	
Signature				0.00
	Low expression	231	0	
	High expression	0	230	

Table 4
Univariable and multivariable Cox regression analysis of the signature with LUAD survival

		Univariable analysis				Multivariable analysis			
Variables		HR	95% CI of HR		*P*	HR	95% CI of HR		*P*
			lower	upper			lower	upper	
GSE31210 dataset(*n* = 226)									
Age	> 61 vs.≤61	1.43	0.73	2.78	0.29	1.70	0.86	3.36	0.12
Sex	Male vs.Female	1.52	0.78	2.96	0.22	0.94	0.47	1.87	0.86
Pathological stage	II vs I,	4.23	2.17	8.24	0.00	3.37	1.68	6.76	0.00
Signature	High risk vs. low risk	20.33	4.87	84.79	0.00	17.49	4.16	73.53	0.00
GSE50081 set (n = 128)									
Age	> 61 vs.≤61	2.09	0.89	4.89	0.09	1.88	0.80	4.41	0.15
Sex	Male vs.Female	1.35	0.78	2.34	0.29	1.44	0.82	2.51	0.20
Pathological stage	II vs I,	2.53	1.45	4.44	0.00	2.35	1.32	4.17	0.00
Signature	High risk vs. low risk	2.09	1.19	3.68	0.01	1.86	1.05	3.30	0.03
TCGA set (n = 461)									
Age	> 61 vs.≤61	0.89	0.66	1.21	0.46	1.06	0.77	1.44	0.74
Sex	Male vs.Female	1.26	0.93	1.70	0.13	1.21	0.89	1.64	0.22
Pathological stage	I II vs III IV,	1.57	1.36	1.81	0.00	1.53	1.33	1.77	0.00
Signature	High risk vs. low risk	1.81	1.33	2.47	0.00	1.65	1.20	2.26	0.00

Functional annotation of the survival-related m^{6}A RNA methylation-related gene set

Univariate analysis was used in GSE31210 to explore the prognostic potential of the candidate m^{6}A RNA methylation-related gene set. The results showed that 129 candidate genes, including m^{6}A RNA

methylation regulatory factor ELAVL1, METTL14, ZC3H13 and YTHDC2, were significantly associated with OS of LUAD.

Four m^6A RNA methylation regulatory factors were indicated to predict favorable overall survival (ZC3H13: HR = 0.44 ; 95% CI, 0.21 to 0.90. METTL14: HR = 0.38; 95% CI, 0.19 to 0.78. YTHDC2: HR = 0.13; 95% CI 0.05 to 0.34, ELAVL1: HR = 0.26; 95% CI, 0.12 to 0.57). We then used the bioinformatic tool STRING to analyze functional protein association networks between these 129 candidate genes. The results indicated that lysine methyltransferase 2A (KMT2A), notch receptor 1(NOTCH1), collagen type III alpha 1 chain collagen type III alpha 1 chain (COL3A1) were the hub genes (Fig. 4).

All statistically enriched terms (Gene Ontology (GO)) by Metascape, and accumulative hypergeometric p-values and enrichment factors were calculated and used for filtering. The significant terms were then hierarchically clustered into a tree, based on Kappa statistical similarities among their gene memberships.The results indicated that basement membrane, inner ear receptor cell differentiation, and inner ear receptor cell stereocilium organization were the most significantly enriched (Fig. 4).

External validation using online database about genes in the signature

Consistent with our results, DENND1A and KIF4A were found to be significantly overexpressed, while BMPER were significantly underexpressed in LUAD in the Oncomine (Fig. 5a), which was almost the same in both TIMER database (Fig. 6) and GEPIA database (Figure S1). Interestingly, the aberrant expression of these five genes was frequently observed in various cancer and showed some tissue-dependent pattern. For example, DENND1A was overexpressed in esophageal cancer, head and neck cancer. However, DENND1A was underexpressed in brain and CNS cancer.

Survival analyses for each gene in the signature (DENND1A, KBTBD6, KIF4A, and BMPER) were performed in the cohort of GSE31210, GSE50081 and TCGA datasets (Fig. 7). DENND1A high expression patients group displayed remarkable shorter OS than DENND1A low expression patients group in GSE31210. While, KIF4A high expression patients group displayed remarkable shorter OS than KIF4A low expression patients group not only in GSE31210, but also in GSE50081 and TCGA dataset. Compared to the KBTBD6 low expression groups, KBTBD6 high expression patient group had more OS in GSE31210 (Fig. 7a, $P < 0.05$). Compared to the BMPER low groups,BMPER high patient group had more OS in GSE31210 (Fig. 7a, $P < 0.05$). Compared to YTHDC2 low expression groups, YTHDC2 high expression patients group had more OS in GSE31210 and TCGA data (Fig. 7a and c, $P < 0.05$).

We then reviewed the proteomic data and found YTHDC2 protein was reported significantly down-expressed in non-small cell lung cancer(27). KIF4A protein was reported significantly up-expressed in non-small cell lung cancer(28). The representative protein expression of KBTBD6, KIF4A, and YTHDC2 was explored in the Human Protein Profiles and shown in Fig. 5b. However, DENND1A and BMPER was not found on the website. BMPER possessed the most frequent genetic alterations (6%) among the five genes. Meanwhile, amplification mutation and deep deletion were the most common alterations among the five genes (Fig. 5c).

Taking together, aberrant expression of the five genes were further validated in LUAD, and genetic alteration might help explain the aberrant expression of these genes to some extent.

Discussion

At the post-transcriptional level, more than 160 kinds of chemical modifications had been found in a variety of RNAs(9, 29). Among these modifications, more and more evidence showed that m^6A modification makes a vital difference in hypertension and cardiovascular diseases, as well as in tumorigenesis and metastasis. Therefore, the identification of m6A-related genes and m^6A RNA methylation regulators in fatal LUAD may offers valuable therapeutic targets to us.

Doctors usuauslly diagnose LUAD as advanced, and there's a high death rate in it. A lots of studies have illuminated that m^6A process is related to lung cancer, which makes m^6A-related gene as potential biomarker for clinical practice. According to our research, the classification of m^6A-related genes in LUAD patients is in association with prognosis. We identify a signature that consists of one m^6A RNA methylation regulators (YTHDC2) and four m^6A-related genes (DENND1A, KBTBD6, KIF4A, and BMPER) by using different statistical and machine learning methods.

Up to now, little is known about the role of YTHDC2 in tumorigenesis. As one of the YTH domain families, YTHDC2 can influence gene expression by binding m^6As(30). Additional RNA binding and protein-protein interaction domains guide the rapid expression and degradation of mRNA, thereby influencing the stability of their mRNA interaction partners(31). Recent reports show that YTHDC2 may have carcinogenic in colon cancer cells and hepatocellular carcinoma cells(32). However, TCGA database show a positive correlation between YTHDC2 expression and the prognosis of head and neck squamous cell carcinoma, suggesting that YTHDC2 may also be a tumor suppressor gene.

DENND1A can regulate the migration and invasion of gastric cancer cells through the interaction of EGF-GRB2-dennd1A-RAB35 as the regulatory center(33). DENND1A also plays roles in hyperandrogenemia and in polycystic ovary syndrome(34).

KIF4A is a member of the KIF family. KIF proteins get involved in lots of crucial cellular biological functions, which are made up of mitosis,intracellular vesicles and organelle transport(35). More and more evidence shows that KIF members are involved in the development and progression of human cancers(36–38). KIF4A is reported to be abnormally expressed and plays an important role in the progression of various solid cancers(28, 39, 40).

The function of BMPER is rarely reported in cancer(41). In B-cell lymphoma, 58% of the BMPER gene promoter was found to be methylated. Some reports showed that BMPER can promote the invasion and migration of cervical cancer cells and other biological behaviors(42). BMPER expression is closely related to the OS of patients, and high BMPER expression is an independent risk factor for poor prognosis. Inhibition of BMPER expression can decrease the proliferation, migration and invasion of ovarian cancer

cell lines CAOV3 and OVCAR3. What's more, it is reported that BMPER can promote the invasion and migration of fibroblasts(43).

Because of the reversible effect of m^6A on mRNA expression, we believe that m^6A-related genes may have different functional patterns and networks when participating in malignant tumors. Thus, m^6A-related genes may have different expression patterns in LUAD. In previous research, little was known about the interaction of m^6A-related genes. Moreover, it's worth nothing that whether the TP53 mutant affects the expression of m^6A RNA methylation regulators and m^6A-related genes is still unclear, and more evidence is needed to clarify their mechanism.

Conclusion

In conclusion, our study systematically analyzed the expression, prognostic value,protein-protein interaction, and potential function of m^6A RNA methylation regulators and m^6A-related genes.We found that the expression of m^6A RNA methylation regulators and m^6A-related genes was closely related to the clinicopathologial characteristics of LUAD. Five-gene signature have been identified that might help effectively identify new therapeutic targets or strategies for LUAD. In summary, our study provides important clues for further studies on the role of RNA m^6A methylation and related genes in LUAD.

Abbreviations

m^6A: N6-methyladenosine; LUAD: Lung adenocarcinoma; ROC: Receiver operating characteristic; GEO:Gene Expression Omnibus; TCGA:The Cancer Genome Atlas; m1A: N1-methyladenosine; RSFVH:random survival forest algorithm; lncRNAs:long chain non-coding RNA; OS: overall survival; GO: Gene Ontology

Declarations

Acknowledgements

We thank all individuals who take part in this research.

Authors' contributions

CY, WZ: study conception and design; BG, HZ, CY: manuscript writing; BG, RW,CY: literature review; all authors: data interpretation and discussion; all authors: final editing and approval of the manuscript in its present form.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and materials

The datasets supporting the conclusions of this article are available in public data (TCGA database and the GEO database).

Ethics approval and consent to participate

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Consent for publication

All authors agree on publication of the results of the present manuscript.

Competing interests

The authors declare that they have no competing interest.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2018;68(6):394-424.
2. Verma M. The Role of Epigenomics in the Study of Cancer Biomarkers and in the Development of Diagnostic Tools. Advances in experimental medicine and biology. 2015;867:59-80.
3. Wang KC, Chang HY. Epigenomics: Technologies and Applications. Circulation research. 2018;122(9):1191-9.
4. Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature. 2012;485(7397):201-6.
5. Spenkuch F, Motorin Y, Helm M. Pseudouridine: still mysterious, but never a fake (uridine)! RNA biology. 2014;11(12):1540-54.
6. Dominissini D, Nachtergaele S, Moshitch-Moshkovitz S, Peer E, Kol N, Ben-Haim MS, et al. The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. Nature. 2016;530(7591):441-6.
7. Alarcón CR, Lee H, Goodarzi H, Halberg N, Tavazoie SF. N6-methyladenosine marks primary microRNAs for processing. Nature. 2015;519(7544):482-5.
8. Zhao BS, Roundtree IA, He C. Post-transcriptional gene regulation by mRNA modifications. Nat Rev Mol Cell Biol. 2017;18(1):31-42.

9. Boccaletto P, Machnicka MA, Purta E, Piatkowski P, Baginski B, Wirecki TK, et al. MODOMICS: a database of RNA modification pathways. 2017 update. Nucleic Acids Res. 2018;46(D1):D303-d7.
10. Wei CM, Gershowitz A, Moss B. Methylated nucleotides block 5' terminus of HeLa cell messenger RNA. Cell. 1975;4(4):379-86.
11. Meyer KD, Jaffrey SR. Rethinking m(6)A Readers, Writers, and Erasers. Annual review of cell and developmental biology. 2017;33:319-42.
12. Bokar JA, Shambaugh ME, Polayes D, Matera AG, Rottman FM. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. Rna-a Publication of the Rna Society. 1997;3(11):1233-47.
13. Wang Y, Li Y, Toth JI, Petroski MD, Zhang Z, Zhao JC. N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. Nature cell biology. 2014;16(2):191-8.
14. Ping XL, Sun BF, Wang L, Xiao W, Yang X, Wang WJ, et al. Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. Cell research. 2014;24(2):177-89.
15. Horiuchi K, Kawamura T, Iwanari H, Ohashi R, Naito M, Kodama T, et al. Identification of Wilms' tumor 1-associating protein complex and its role in alternative splicing and the cell cycle. The Journal of biological chemistry. 2013;288(46):33292-302.
16. Patil DP, Chen CK, Pickering BF, Chow A, Jackson C, Guttman M, et al. m(6)A RNA methylation promotes XIST-mediated transcriptional repression. Nature. 2016;537(7620):369-73.
17. Schwartz S, Mumbach MR, Jovanovic M, Wang T, Maciag K, Bushkin GG, et al. Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. Cell Rep. 2014;8(1):284-96.
18. Wen J, Lv R, Ma H, Shen H, He C, Wang J, et al. Zc3h13 Regulates Nuclear RNA m(6)A Methylation and Mouse Embryonic Stem Cell Self-Renewal. Molecular cell. 2018;69(6):1028-38.e6.
19. Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D, et al. N6-methyladenosine-dependent regulation of messenger RNA stability. Nature. 2014;505(7481):117-20.
20. Huang Y, Yan J, Li Q, Li J, Gong S, Zhou H, et al. Meclofenamic acid selectively inhibits FTO demethylation of m6A over ALKBH5. Nucleic Acids Res. 2015;43(1):373-84.
21. Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. Cell. 2012;149(7):1635-46.
22. Zhu J, Wang M, Hu D. Deciphering N(6)-Methyladenosine-Related Genes Signature to Predict Survival in Lung Adenocarcinoma. BioMed research international. 2020;2020:2514230.
23. Zhuang Z, Chen L, Mao Y, Zheng Q, Li H, Huang Y, et al. Diagnostic, progressive and prognostic performance of m(6)A methylation RNA regulators in lung adenocarcinoma. International journal of biological sciences. 2020;16(11):1785-97.
24. Zhang Y, Liu X, Liu L, Li J, Hu Q, Sun R. Expression and Prognostic Significance of m6A-Related Genes in Lung Adenocarcinoma. Medical science monitor : international medical journal of experimental and clinical research. 2020;26:e919644.

25. Zhang B, Wu Q, Li B, Wang D, Wang L, Zhou YL. m(6)A regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in gastric cancer. Molecular cancer. 2020;19(1):53.
26. Zheng Y, Nie P, Peng D, He Z, Liu M, Xie Y, et al. m6AVar: a database of functional variants involved in m6A modification. Nucleic Acids Res. 2018;46(D1):D139-d45.
27. Sun S, Han Q, Liang M, Zhang Q, Zhang J, Cao J. Downregulation of m(6) A reader YTHDC2 promotes tumor progression and predicts poor prognosis in non-small cell lung cancer. Thoracic cancer. 2020;11(11):3269-79.
28. Taniwaki M, Takano A, Ishikawa N, Yasui W, Inai K, Nishimura H, et al. Activation of KIF4A as a prognostic biomarker and therapeutic target for lung cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2007;13(22 Pt 1):6624-31.
29. Roundtree IA, Evans ME, Pan T, He C. Dynamic RNA Modifications in Gene Expression Regulation. Cell. 2017;169(7):1187-200.
30. Kretschmer J, Rao H, Hackert P, Sloan KE, Höbartner C, Bohnsack MT. The m(6)A reader protein YTHDC2 interacts with the small ribosomal subunit and the 5'-3' exoribonuclease XRN1. RNA (New York, NY). 2018;24(10):1339-50.
31. Bailey AS, Batista PJ, Gold RS, Chen YG, de Rooij DG, Chang HY, et al. The conserved RNA helicase YTHDC2 regulates the transition from proliferation to differentiation in the germline. eLife. 2017;6.
32. Tanabe A, Tanikawa K, Tsunetomi M, Takai K, Ikeda H, Konno J, et al. RNA helicase YTHDC2 promotes cancer metastasis via the enhancement of the efficiency by which HIF-1α mRNA is translated. Cancer letters. 2016;376(1):34-42.
33. Ye B, Duan B, Deng W, Wang Y, Chen Y, Cui J, et al. EGF Stimulates Rab35 Activation and Gastric Cancer Cell Migration by Regulating DENND1A-Grb2 Complex Formation. Frontiers in pharmacology. 2018;9:1343.
34. McAllister JM, Modi B, Miller BA, Biegler J, Bruggeman R, Legro RS, et al. Overexpression of a DENND1A isoform produces a polycystic ovary syndrome theca phenotype. Proc Natl Acad Sci U S A. 2014;111(15):E1519-27.
35. Goldstein LS, Philp AV. The road less traveled: emerging principles of kinesin motor utilization. Annual review of cell and developmental biology. 1999;15:141-83.
36. Corson TW, Zhu CQ, Lau SK, Shepherd FA, Tsao MS, Gallie BL. KIF14 messenger RNA expression is independently prognostic for outcome in lung cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2007;13(11):3229-34.
37. Bie L, Zhao G, Wang YP, Zhang B. Kinesin family member 2C (KIF2C/MCAK) is a novel marker for prognosis in human gliomas. Clinical neurology and neurosurgery. 2012;114(4):356-60.
38. Shimo A, Tanikawa C, Nishidate T, Lin ML, Matsuda K, Park JH, et al. Involvement of kinesin family member 2C/mitotic centromere-associated kinesin overexpression in mammary carcinogenesis. Cancer ence 2010;99(1):62-70.

39. Cao Q, Song Z, Ruan H, Wang C, Yang X, Bao L, et al. Targeting the KIF4A/AR Axis to Reverse Endocrine Therapy Resistance in Castration-resistant Prostate Cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2020;26(6):1516-28.
40. Zou JX, Duan Z, Wang J, Sokolov A, Xu J, Chen CZ, et al. Kinesin family deregulation coordinated by bromodomain protein ANCCA and histone methyltransferase MLL for breast cancer cell growth, survival, and tamoxifen resistance. Molecular cancer research : MCR. 2014;12(4):539-49.
41. Bethge N, Honne H, Andresen K, Hilden V, Trøen G, Liestøl K, et al. A gene panel, including LRP12, is frequently hypermethylated in major types of B-cell lymphoma. PLoS One. 2014;9(9):e104249.
42. Huan C, Yang T, Liang J, Xie T, Cheng L, Liu N, et al. Methylation-mediated BMPER expression in fibroblast activation in vitro and lung fibrosis in mice in vivo. Sci Rep. 2015;5:14910.
43. Heinke J, Kerber M, Rahner S, Mnich L, Lassmann S, Helbing T, et al. Bone morphogenetic protein modulator BMPER is highly expressed in malignant tumors and controls invasive cell behavior. Oncogene. 2012;31(24):2919-30.
44. Zhang C, Zhang M, Ge S, Huang W, Lin X, Gao J, et al. Reduced m6A modification predicts malignant phenotypes and augmented Wnt/PI3K-Akt signaling in gastric cancer. Cancer medicine. 2019;8(10):4766-81.

Figures

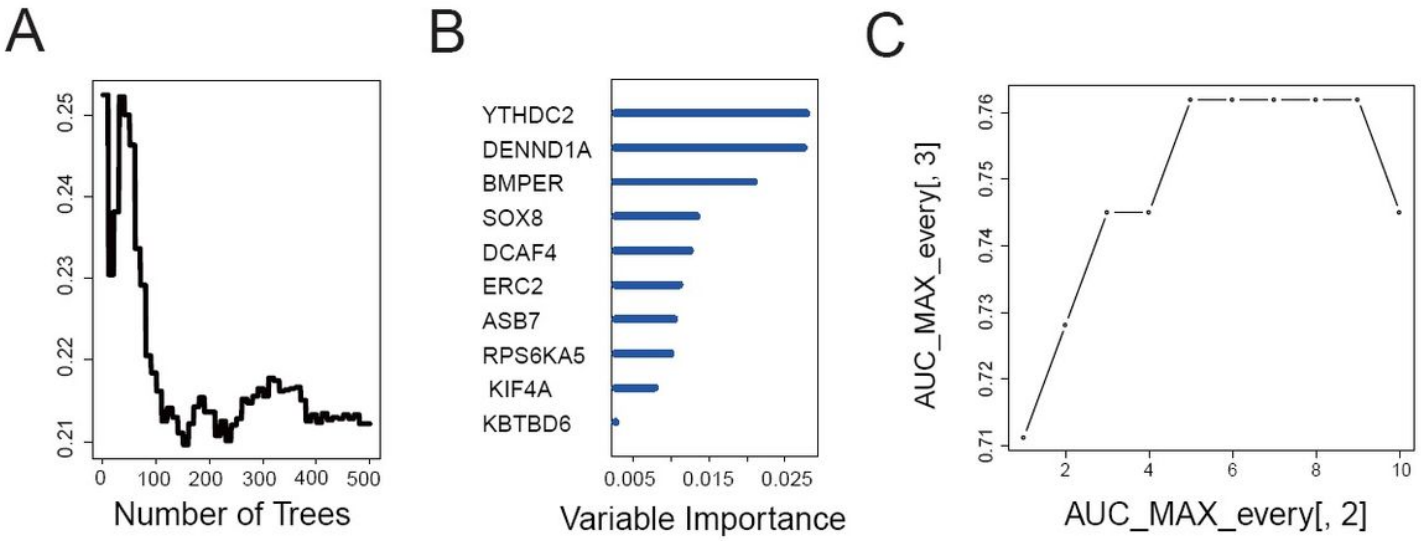

Figure 1

a,b random survival forests-variable hunting analysis reveals the error rate for the data as a function of trees and uses the associated score to filter m6A-related genes and m6A RNA methylation regulators. c ROC for selected prognostic signature from all 1023 signatures

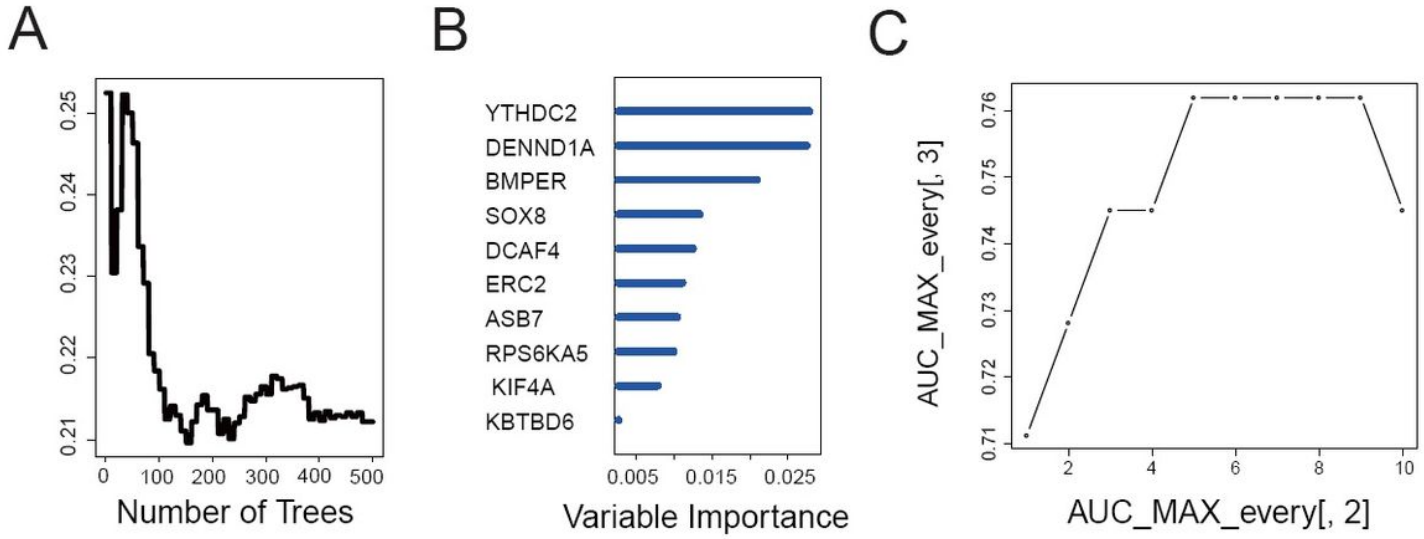

Figure 1

a,b random survival forests-variable hunting analysis reveals the error rate for the data as a function of trees and uses the associated score to filter m6A-related genes and m6A RNA methylation regulators. c ROC for selected prognostic signature from all 1023 signatures

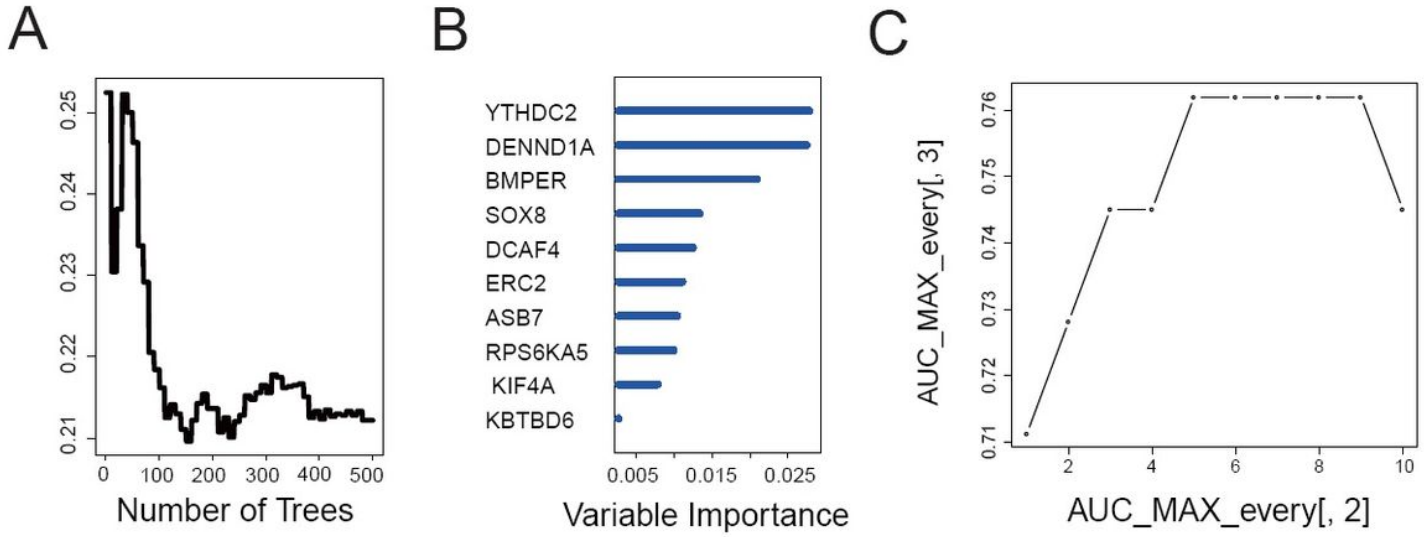

Figure 1

a,b random survival forests-variable hunting analysis reveals the error rate for the data as a function of trees and uses the associated score to filter m6A-related genes and m6A RNA methylation regulators. c ROC for selected prognostic signature from all 1023 signatures

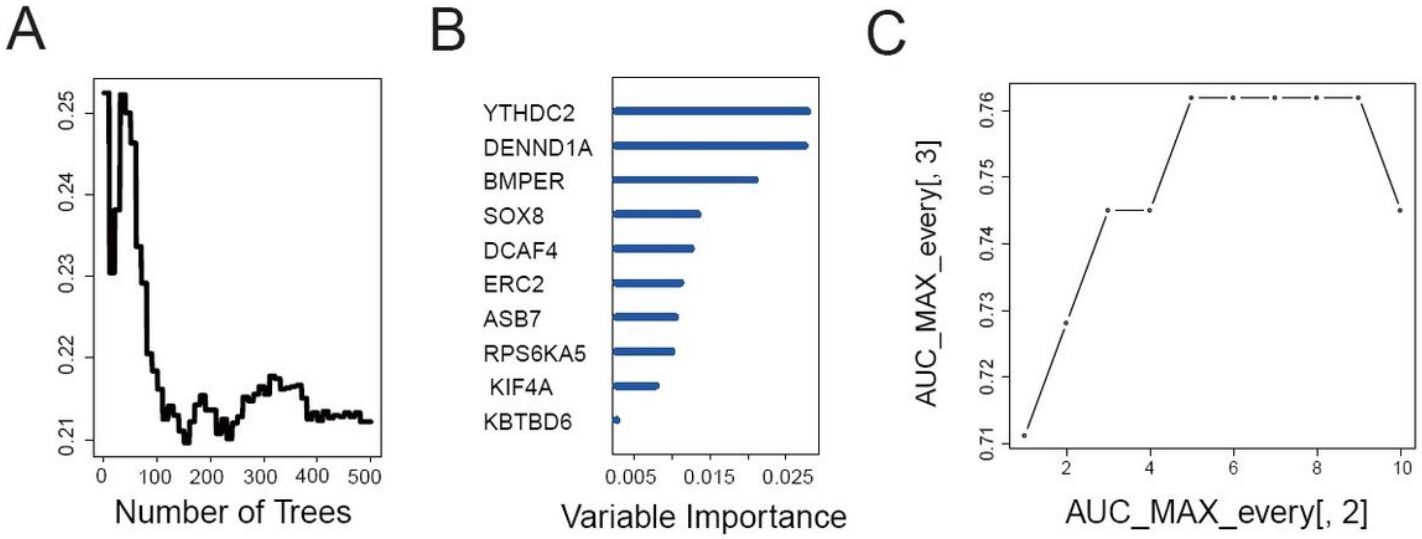

Figure 1

a,b random survival forests-variable hunting analysis reveals the error rate for the data as a function of trees and uses the associated score to filter m6A-related genes and m6A RNA methylation regulators. c ROC for selected prognostic signature from all 1023 signatures

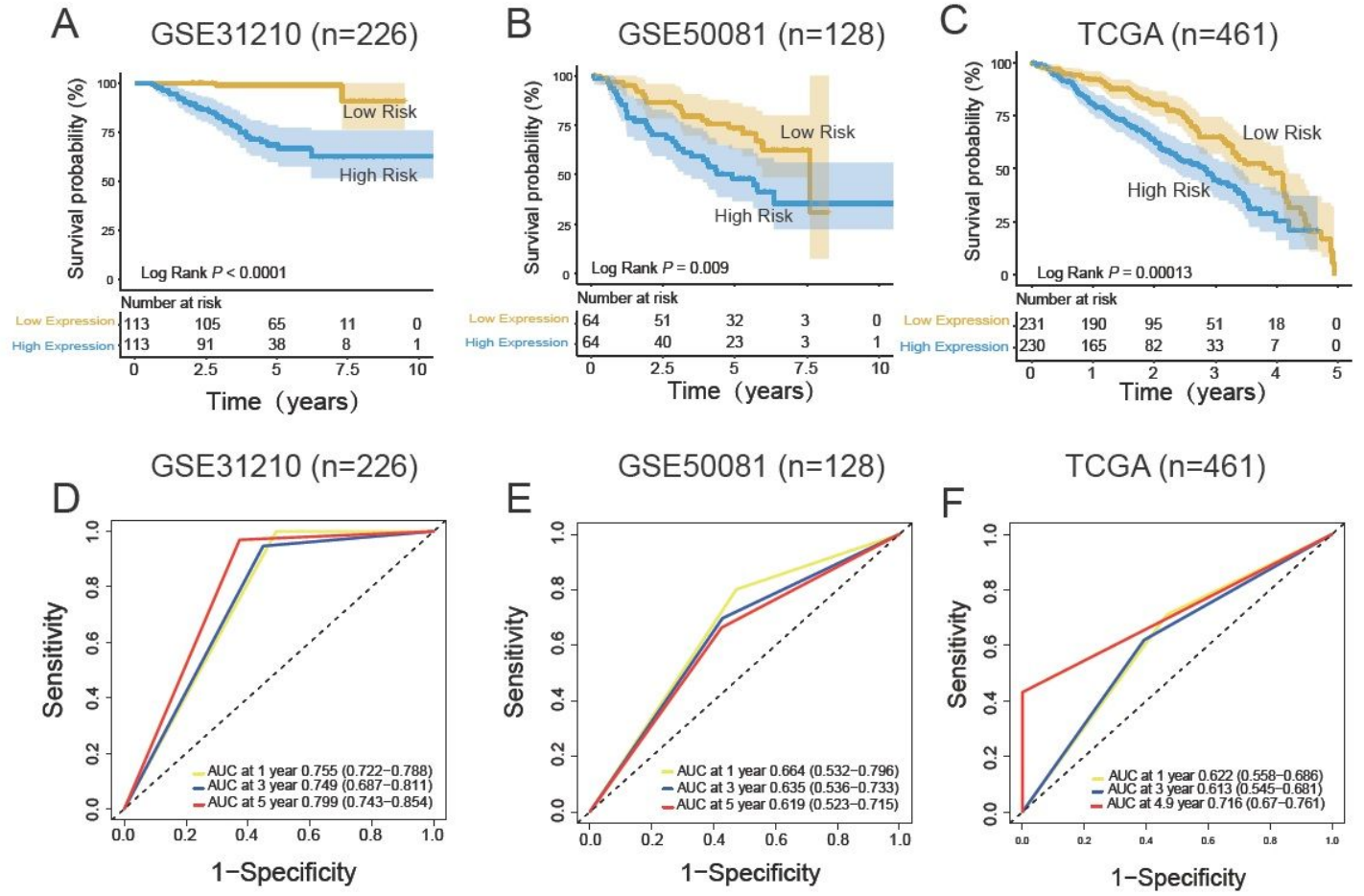

Figure 2

m6A-related genes and m6A RNA methylation regulators signature predicts overall survival of patients of LUAD. a,b,c Kaplan–Meier survival curves classify patients into high- and low-risk groups by the m6A-related genes and m6A RNA methylation regulators signature in the training dataset (GSE31210), and test dataset (GSE50081 and TCGA). P values were calculated by log-rank test. d,e,f m6A-related genes and m6A RNA methylation regulators signatures were used for predicting survival among 1, 3 and 5 years by TimeROC analysis

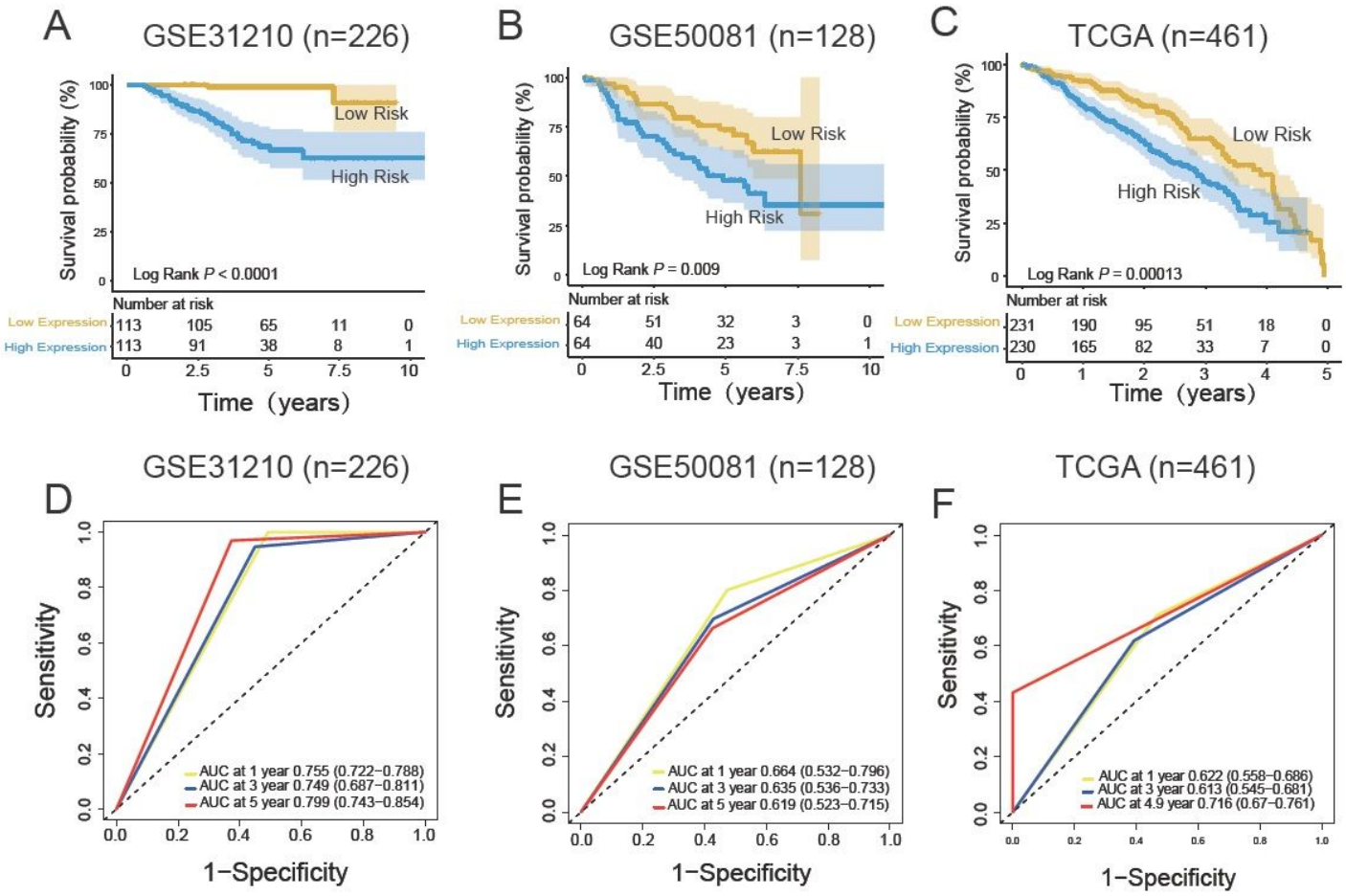

Figure 2

m6A-related genes and m6A RNA methylation regulators signature predicts overall survival of patients of LUAD. a,b,c Kaplan–Meier survival curves classify patients into high- and low-risk groups by the m6A-related genes and m6A RNA methylation regulators signature in the training dataset (GSE31210), and test dataset (GSE50081 and TCGA). P values were calculated by log-rank test. d,e,f m6A-related genes and m6A RNA methylation regulators signatures were used for predicting survival among 1, 3 and 5 years by TimeROC analysis

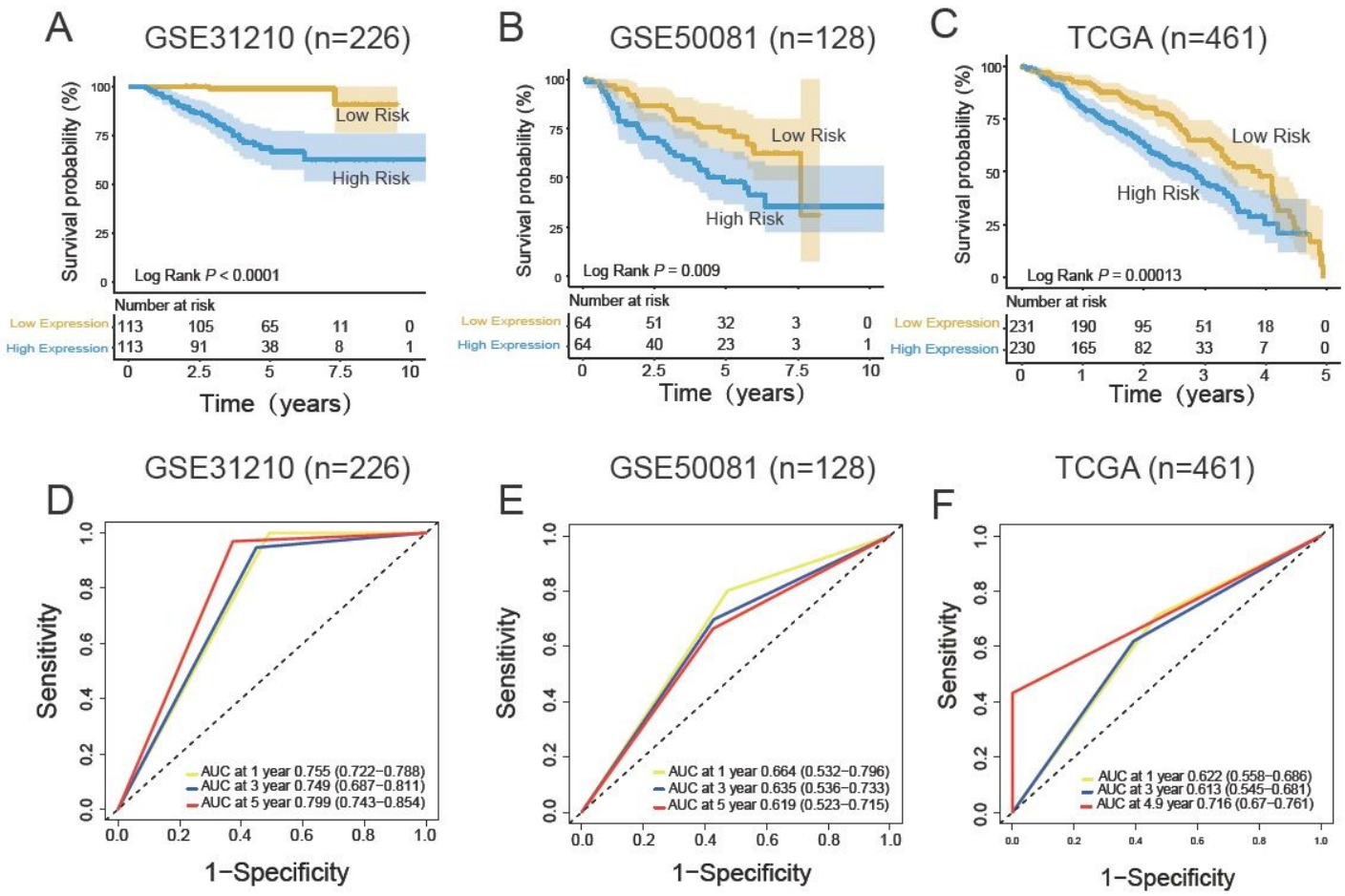

Figure 2

m6A-related genes and m6A RNA methylation regulators signature predicts overall survival of patients of LUAD. a,b,c Kaplan–Meier survival curves classify patients into high- and low-risk groups by the m6A-related genes and m6A RNA methylation regulators signature in the training dataset (GSE31210), and test dataset (GSE50081 and TCGA). P values were calculated by log-rank test. d,e,f m6A-related genes and m6A RNA methylation regulators signatures were used for predicting survival among 1, 3 and 5 years by TimeROC analysis

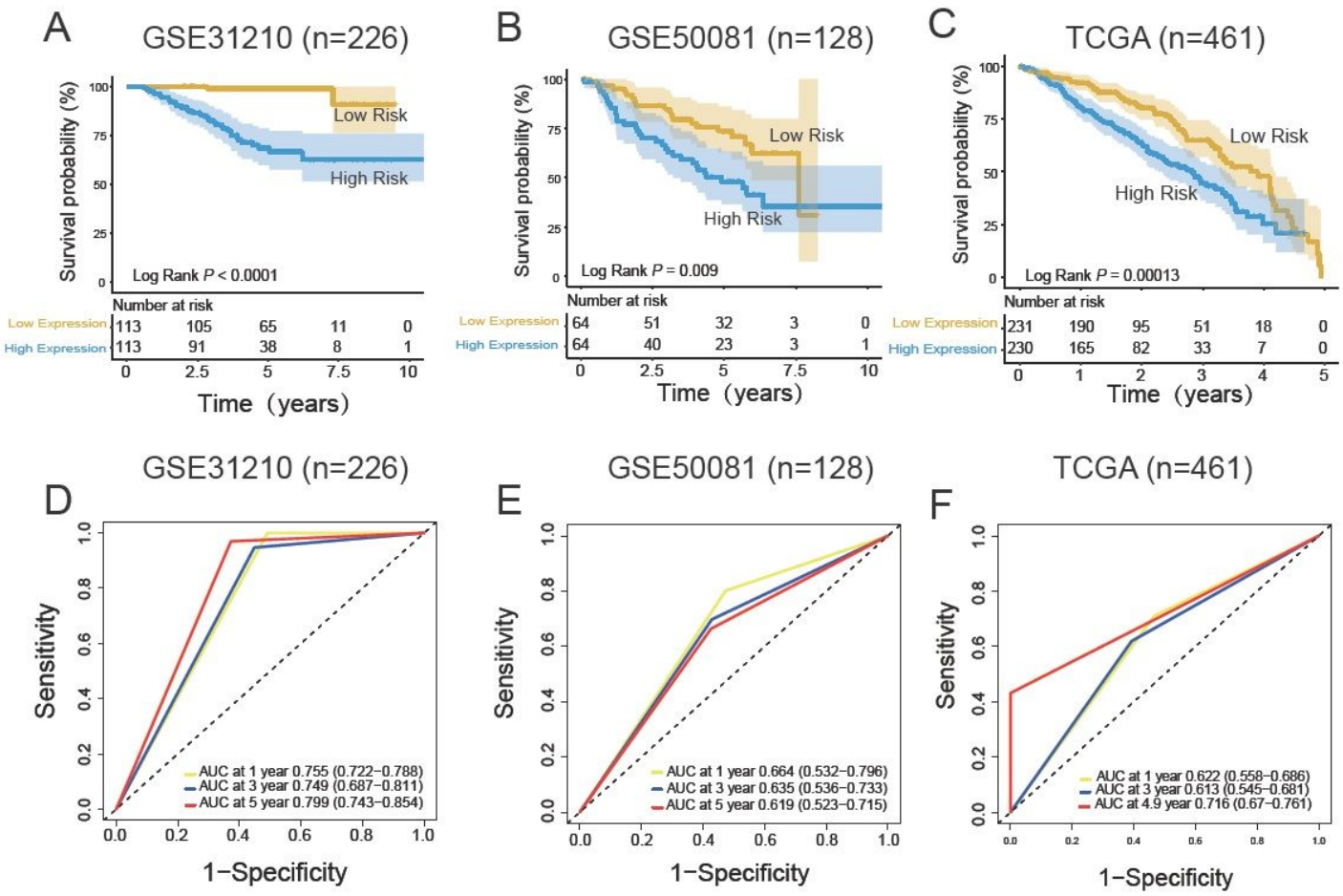

Figure 2

m6A-related genes and m6A RNA methylation regulators signature predicts overall survival of patients of LUAD. a,b,c Kaplan–Meier survival curves classify patients into high- and low-risk groups by the m6A-related genes and m6A RNA methylation regulators signature in the training dataset (GSE31210), and test dataset (GSE50081 and TCGA). P values were calculated by log-rank test. d,e,f m6A-related genes and m6A RNA methylation regulators signatures were used for predicting survival among 1, 3 and 5 years by TimeROC analysis

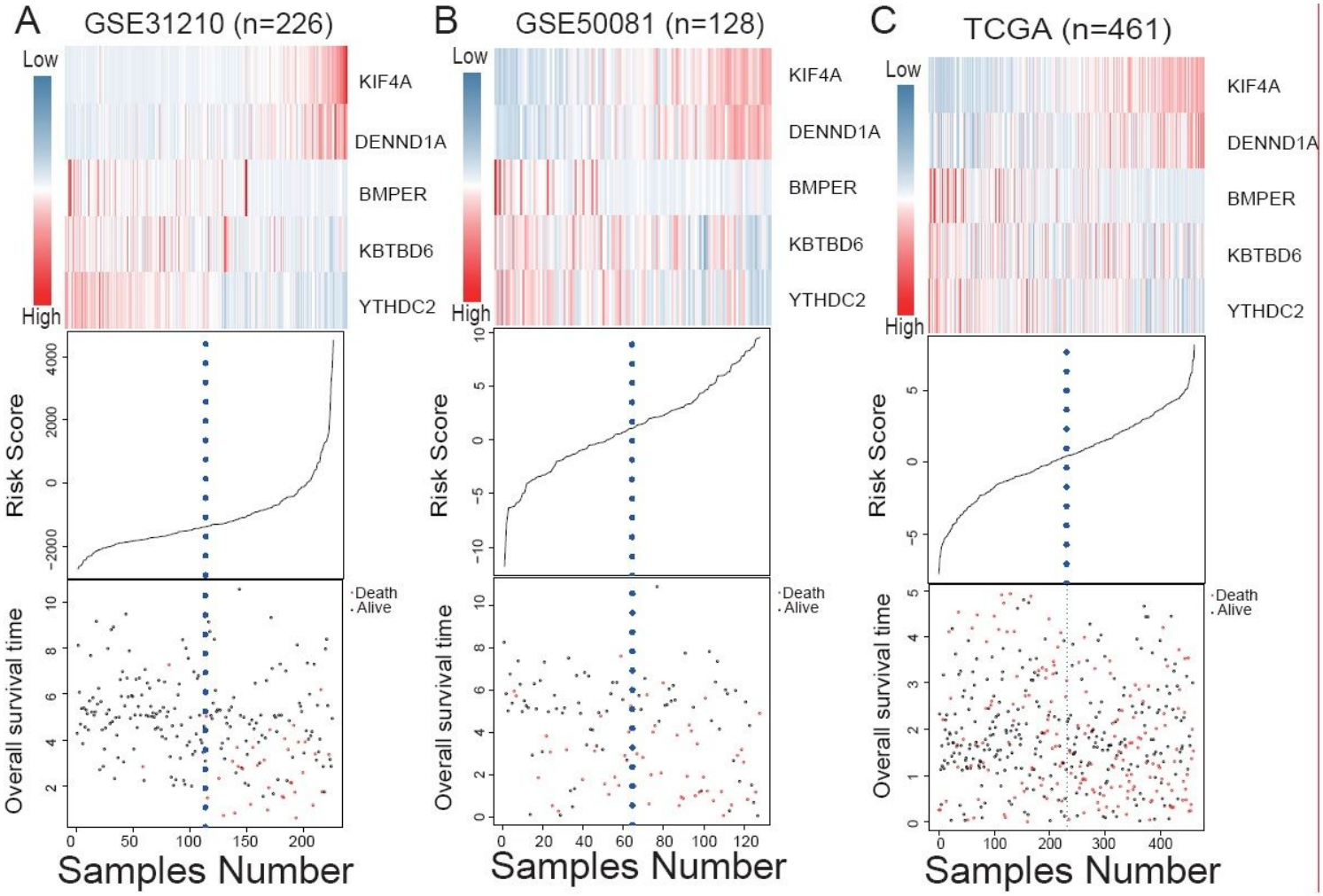

Figure 3

Evaluation of the risk predictive model in training set and test set. a,b,c The distribution of m6A RNA methylation regulators and m6A-related genes expression level, patients' survival status and risk score between high- and low-risk group

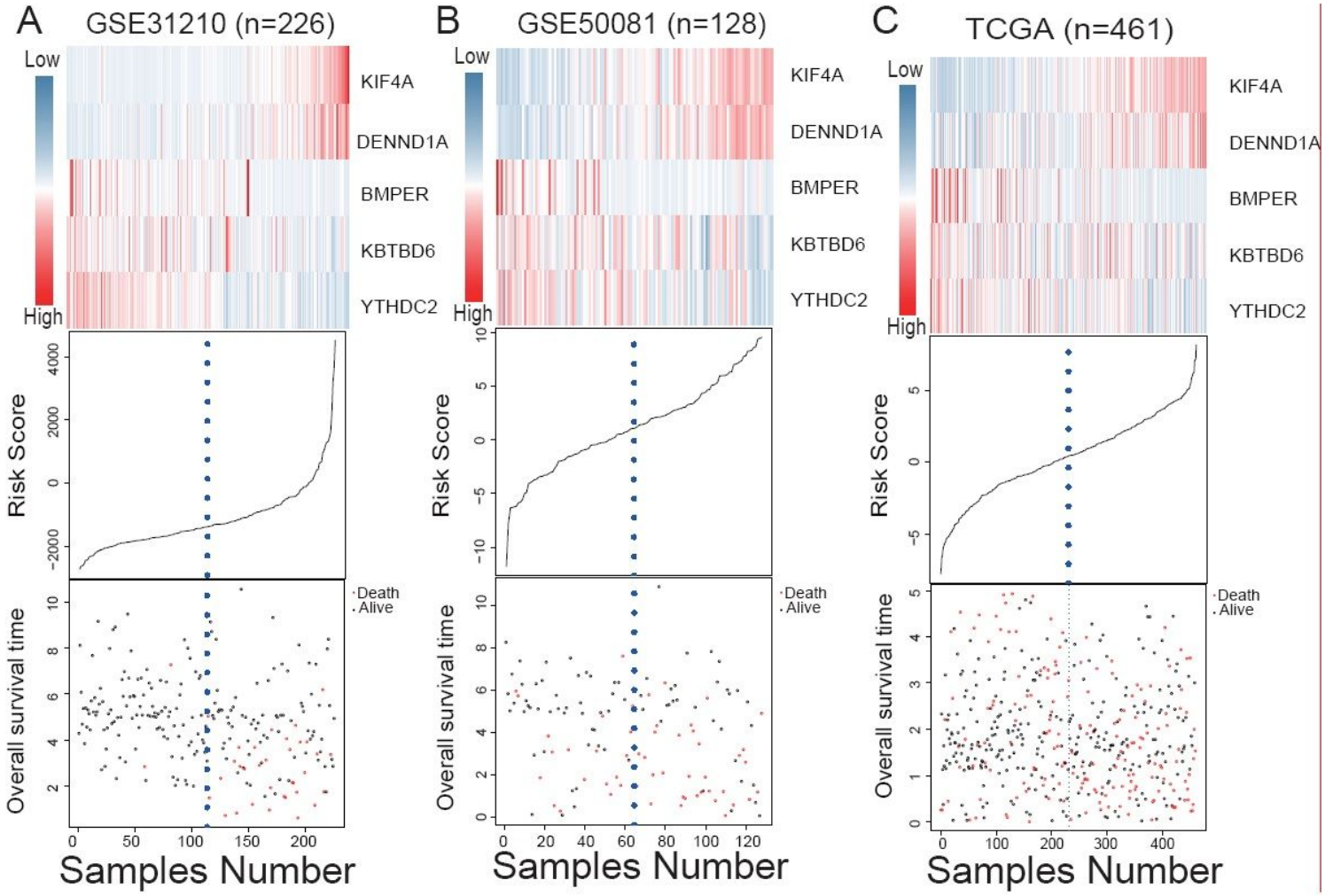

Figure 3

Evaluation of the risk predictive model in training set and test set. a,b,c The distribution of m6A RNA methylation regulators and m6A-related genes expression level, patients' survival status and risk score between high- and low-risk group

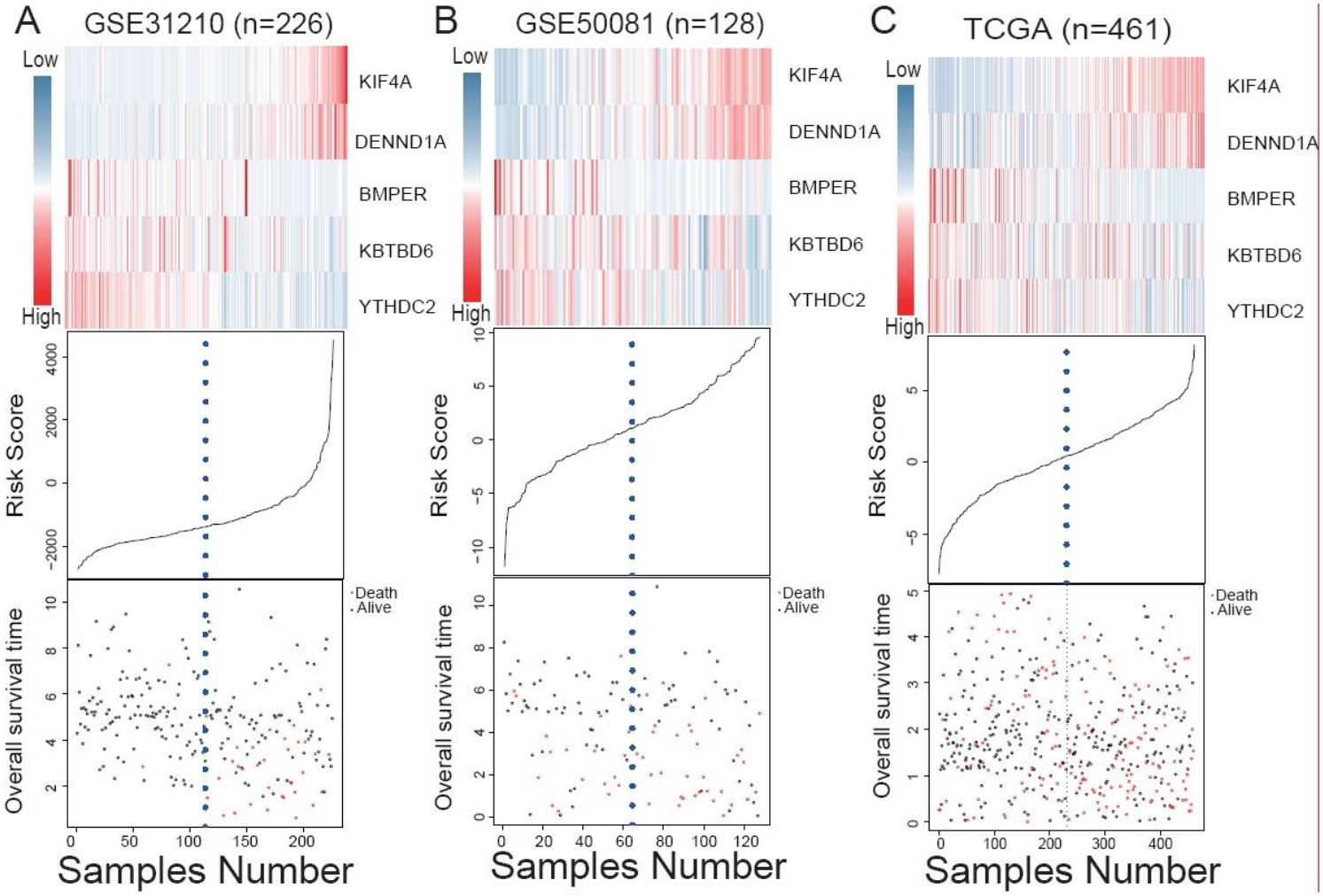

Figure 3

Evaluation of the risk predictive model in training set and test set. a,b,c The distribution of m6A RNA methylation regulators and m6A-related genes expression level, patients' survival status and risk score between high- and low-risk group

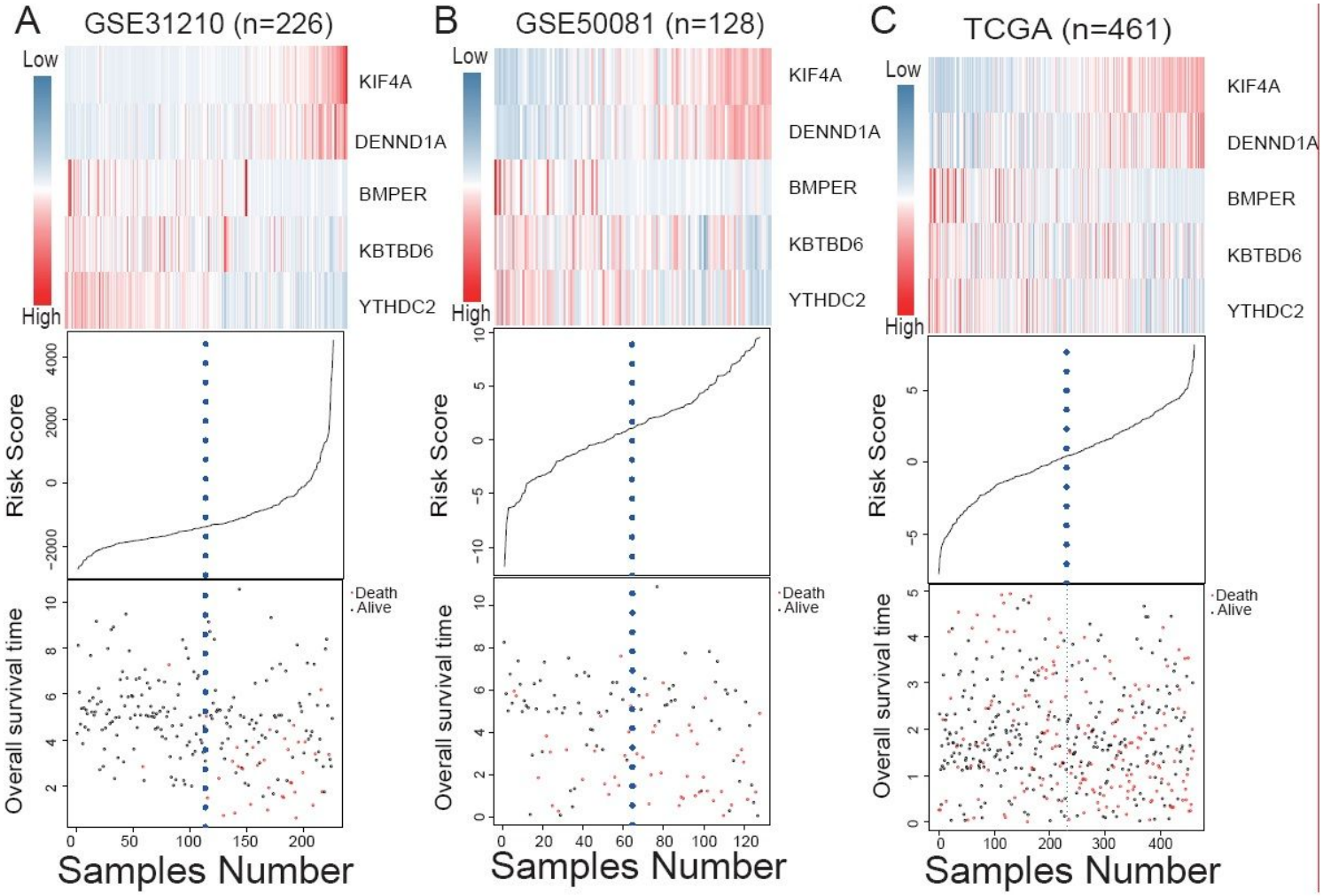

Figure 3

Evaluation of the risk predictive model in training set and test set. a,b,c The distribution of m6A RNA methylation regulators and m6A-related genes expression level, patients' survival status and risk score between high- and low-risk group

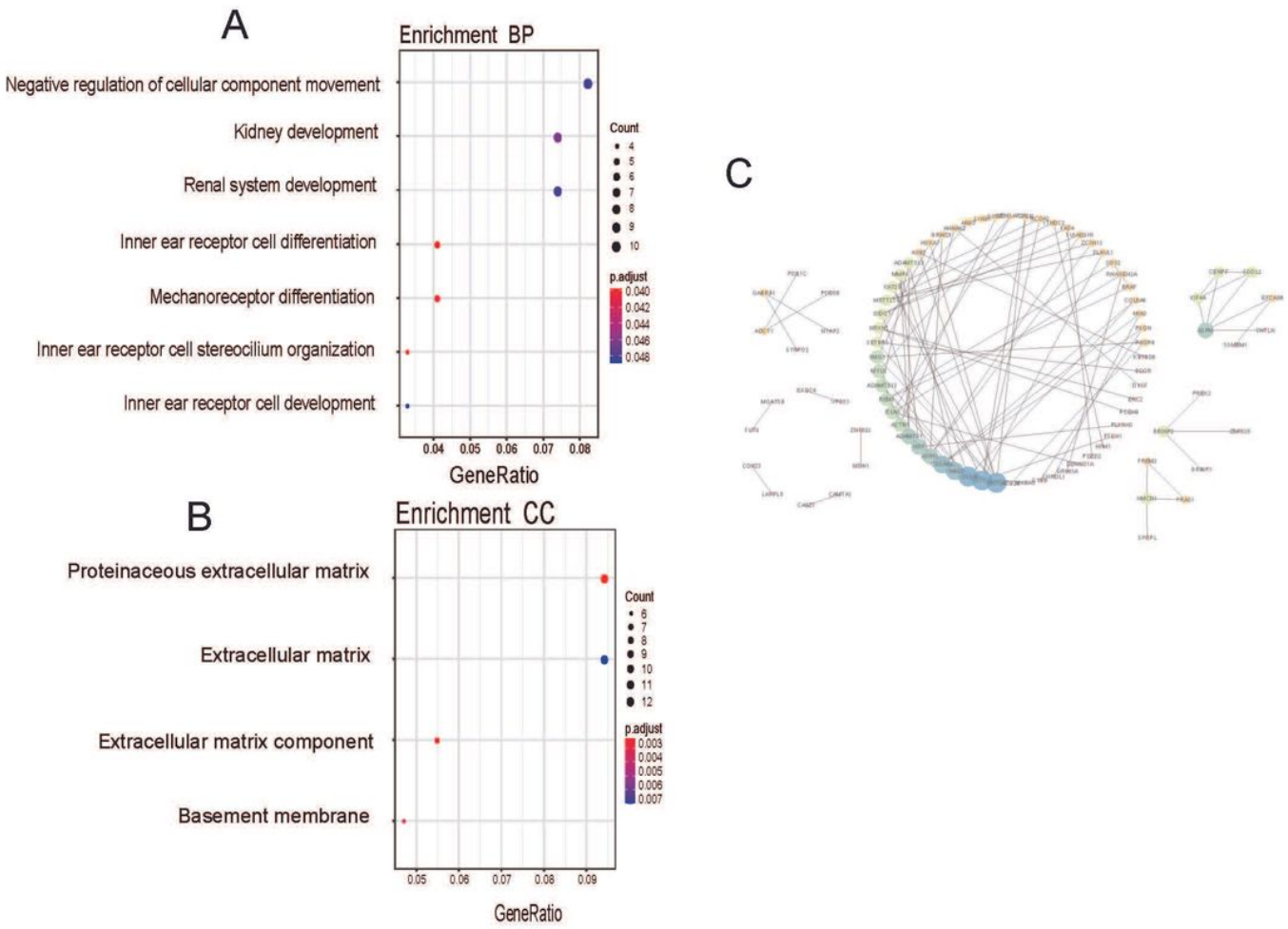

Figure 4

Function prediction and Protein-Protein interaction for the Genes with a significant prognosis

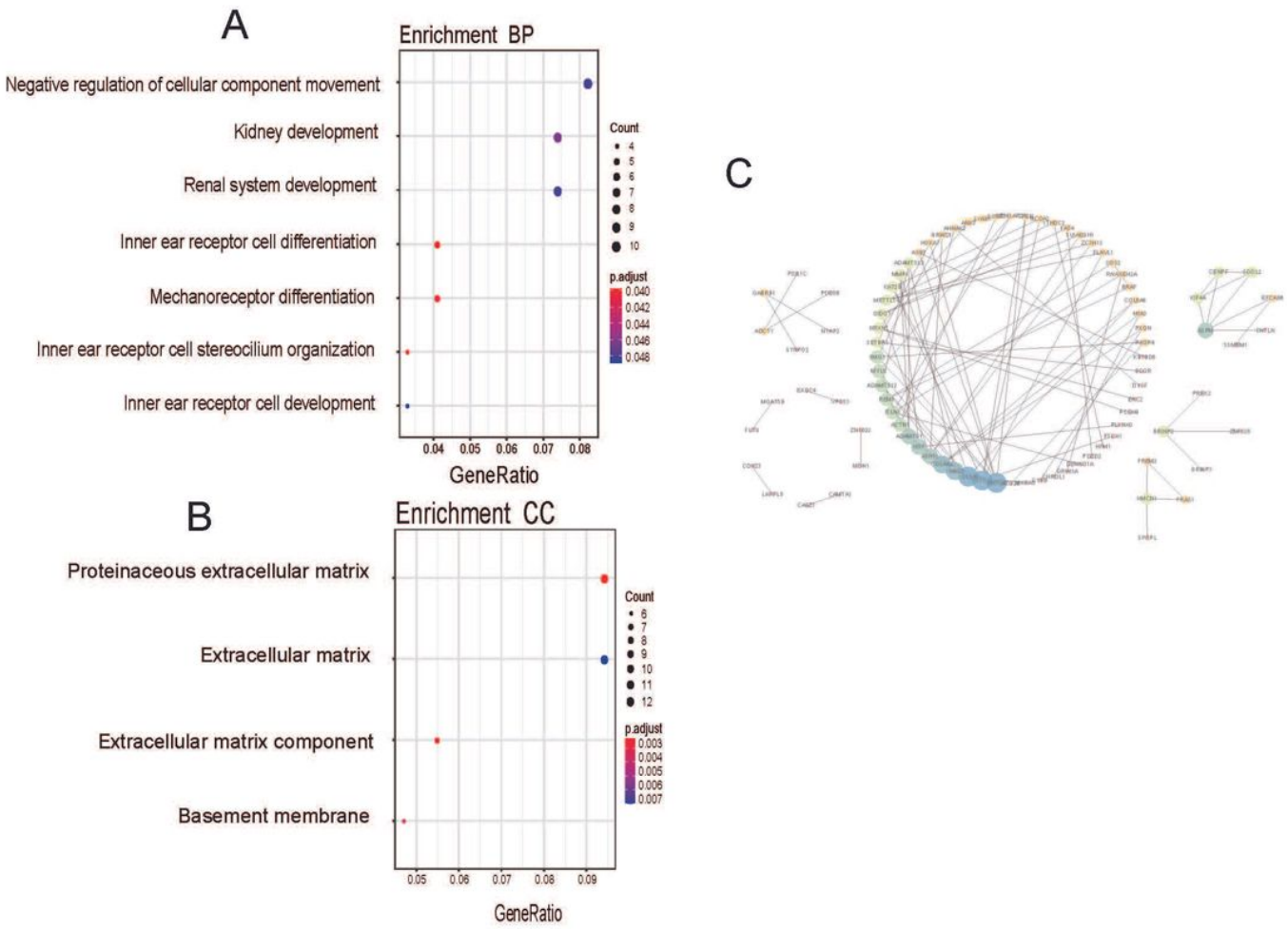

Figure 4

Function prediction and Protein-Protein interaction for the Genes with a significant prognosis

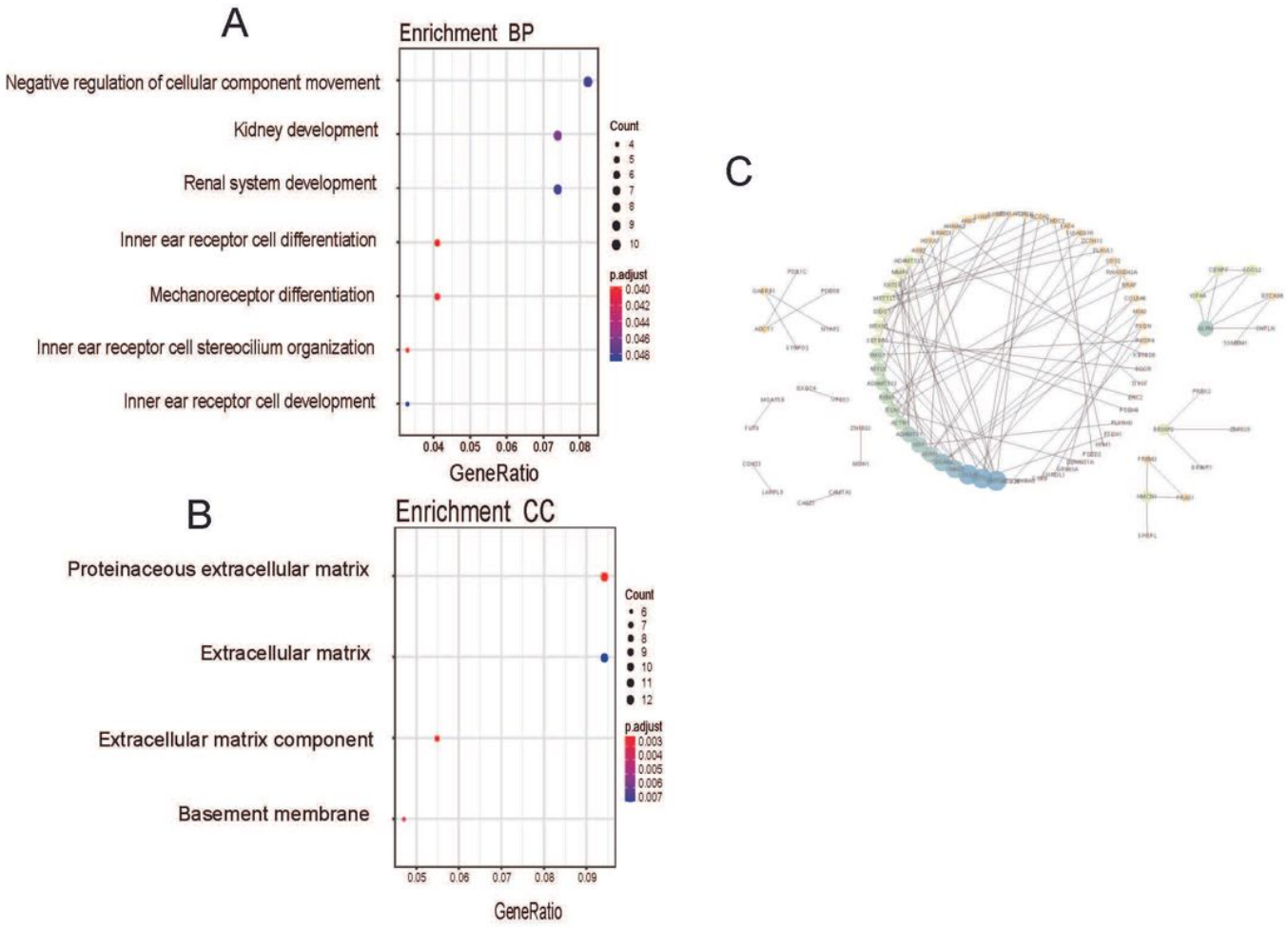

Figure 4

Function prediction and Protein-Protein interaction for the Genes with a significant prognosis

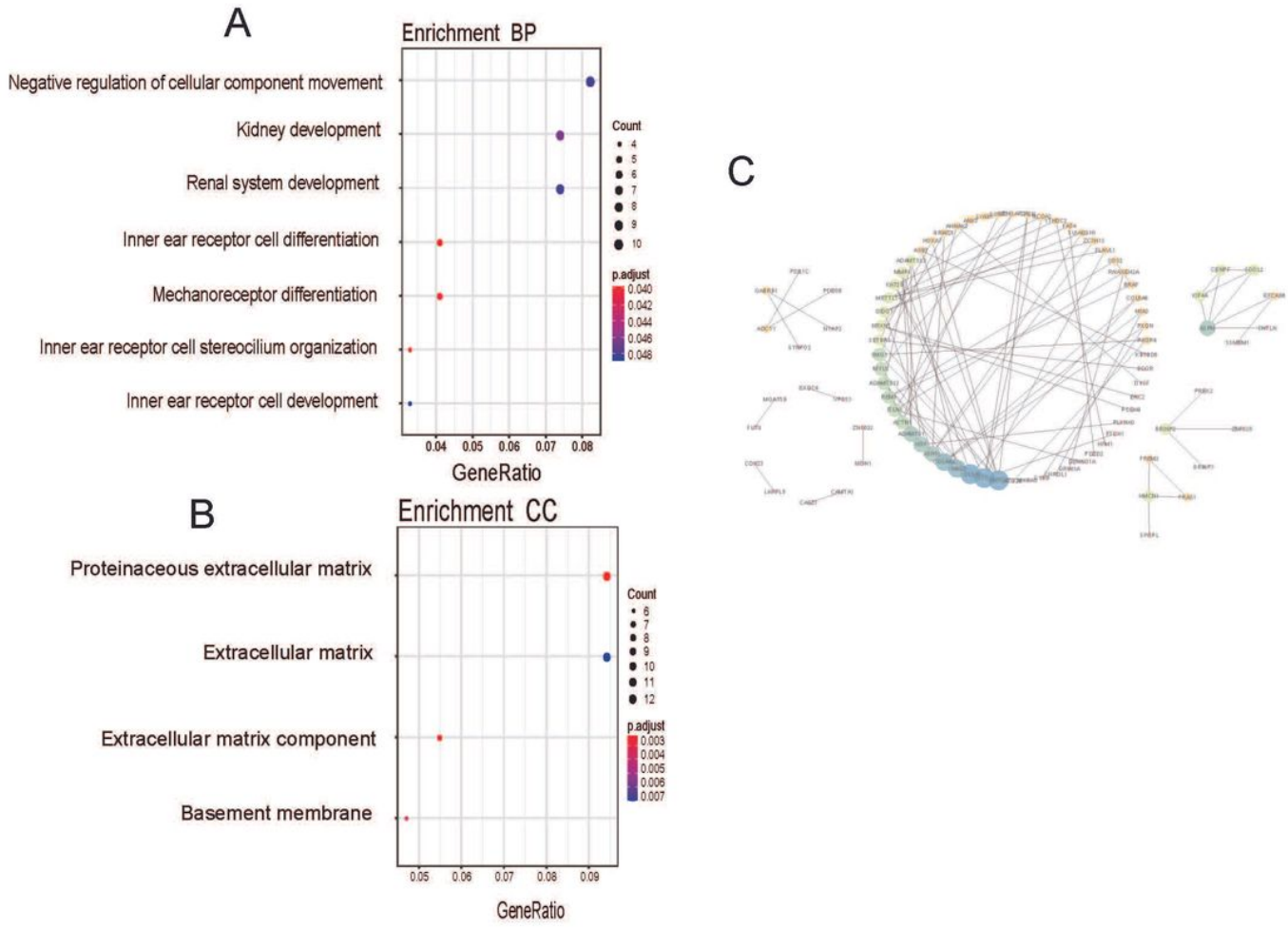

Figure 4

Function prediction and Protein-Protein interaction for the Genes with a significant prognosis

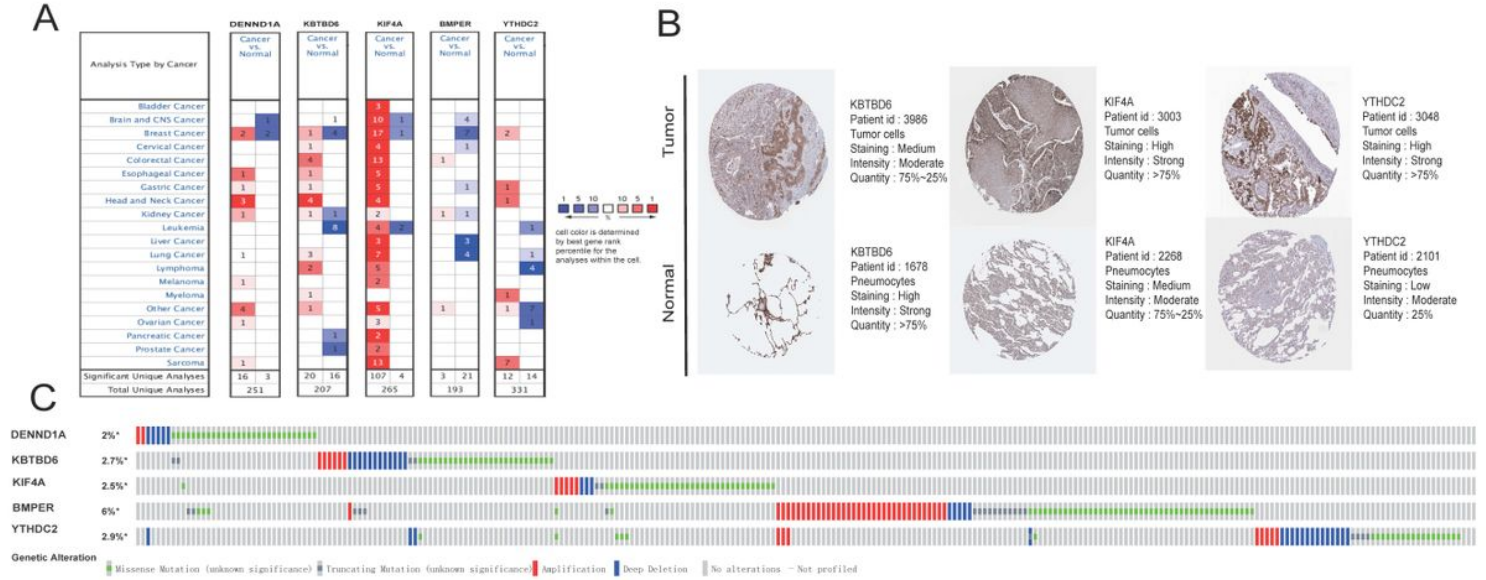

Figure 5

Expression and genetic alterations of the five predictive genes. a The expression profiles of the five genes in the Oncomine database. b The representative protein expression of the three genes in LUAD and normal lung tissue in the Human Protein Atlas database. Data of DENND1A and BMPER were not found in the database. c Genetic alterations of the five genes in LUAD in the cBioportal for Cancer Genomics

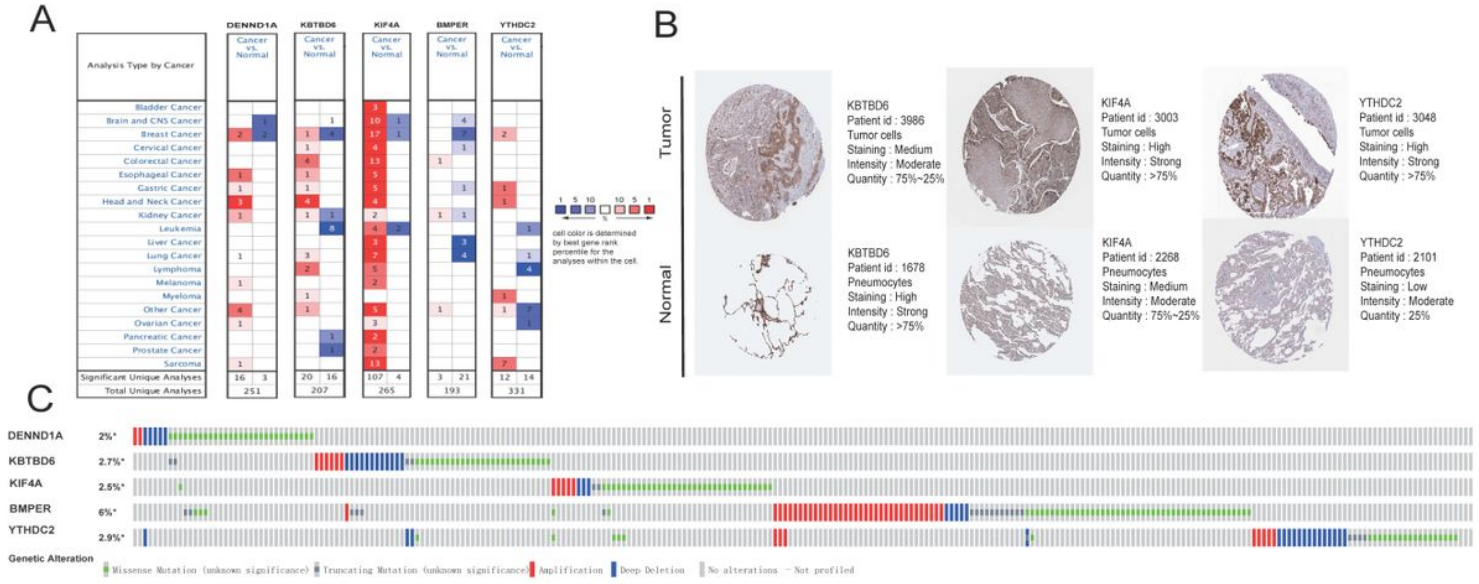

Figure 5

Expression and genetic alterations of the five predictive genes. a The expression profiles of the five genes in the Oncomine database. b The representative protein expression of the three genes in LUAD and normal lung tissue in the Human Protein Atlas database. Data of DENND1A and BMPER were not found in the database. c Genetic alterations of the five genes in LUAD in the cBioportal for Cancer Genomics

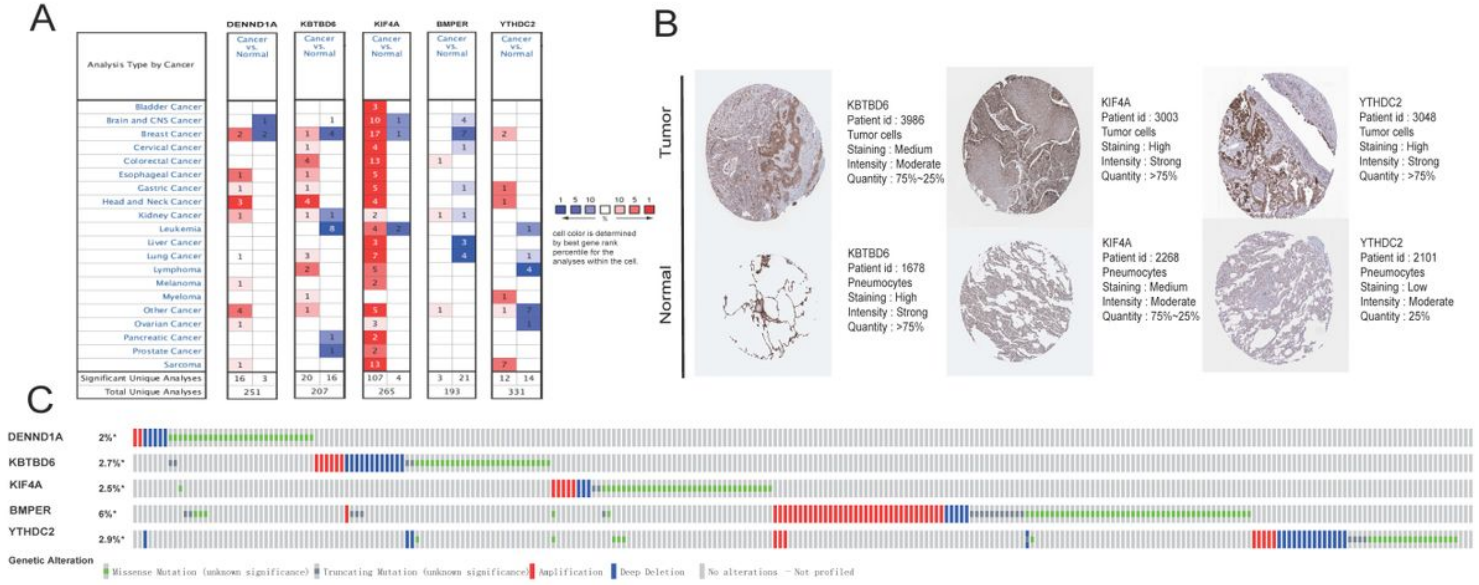

Figure 5

Expression and genetic alterations of the five predictive genes. a The expression profiles of the five genes in the Oncomine database. b The representative protein expression of the three genes in LUAD and

normal lung tissue in the Human Protein Atlas database. Data of DENND1A and BMPER were not found in the database. c Genetic alterations of the five genes in LUAD in the cBioportal for Cancer Genomics

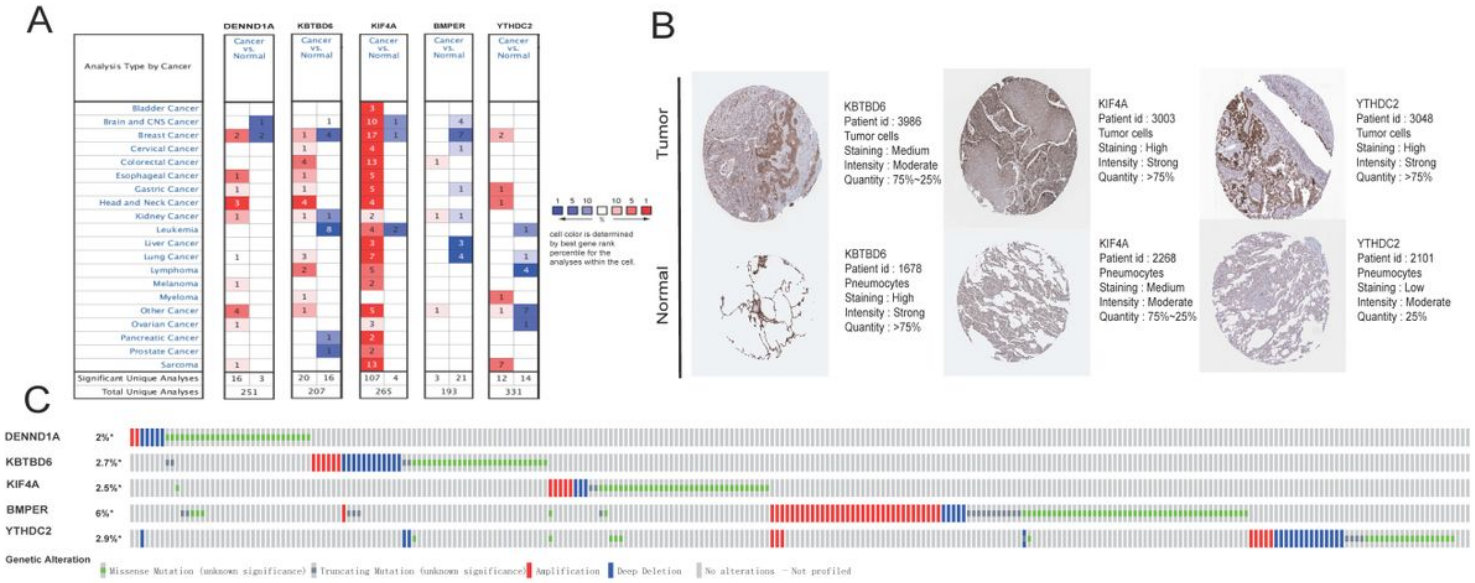

Figure 5

Expression and genetic alterations of the five predictive genes. a The expression profiles of the five genes in the Oncomine database. b The representative protein expression of the three genes in LUAD and normal lung tissue in the Human Protein Atlas database. Data of DENND1A and BMPER were not found in the database. c Genetic alterations of the five genes in LUAD in the cBioportal for Cancer Genomics

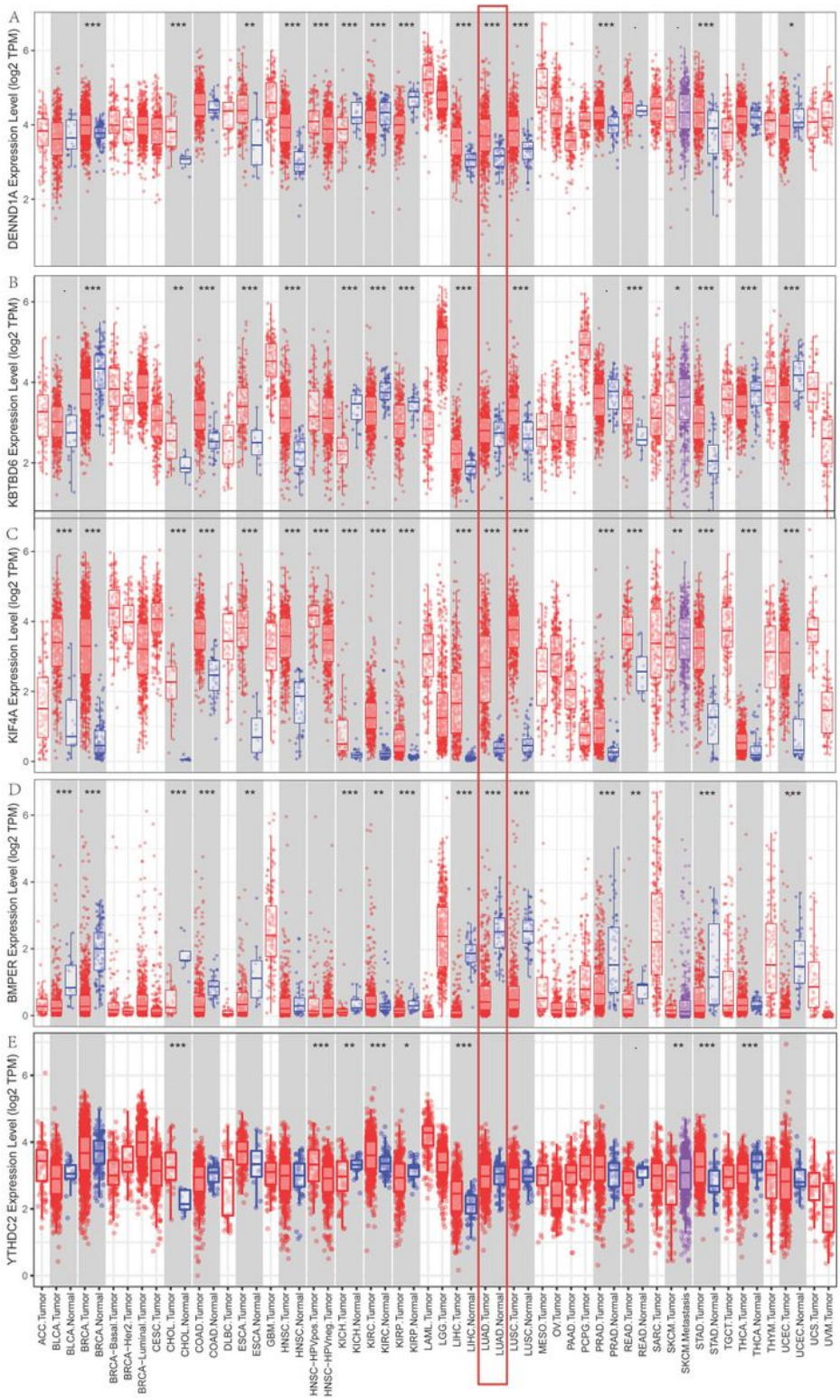

Figure 6

The expression of the five predictive genes in cancers in TIMER

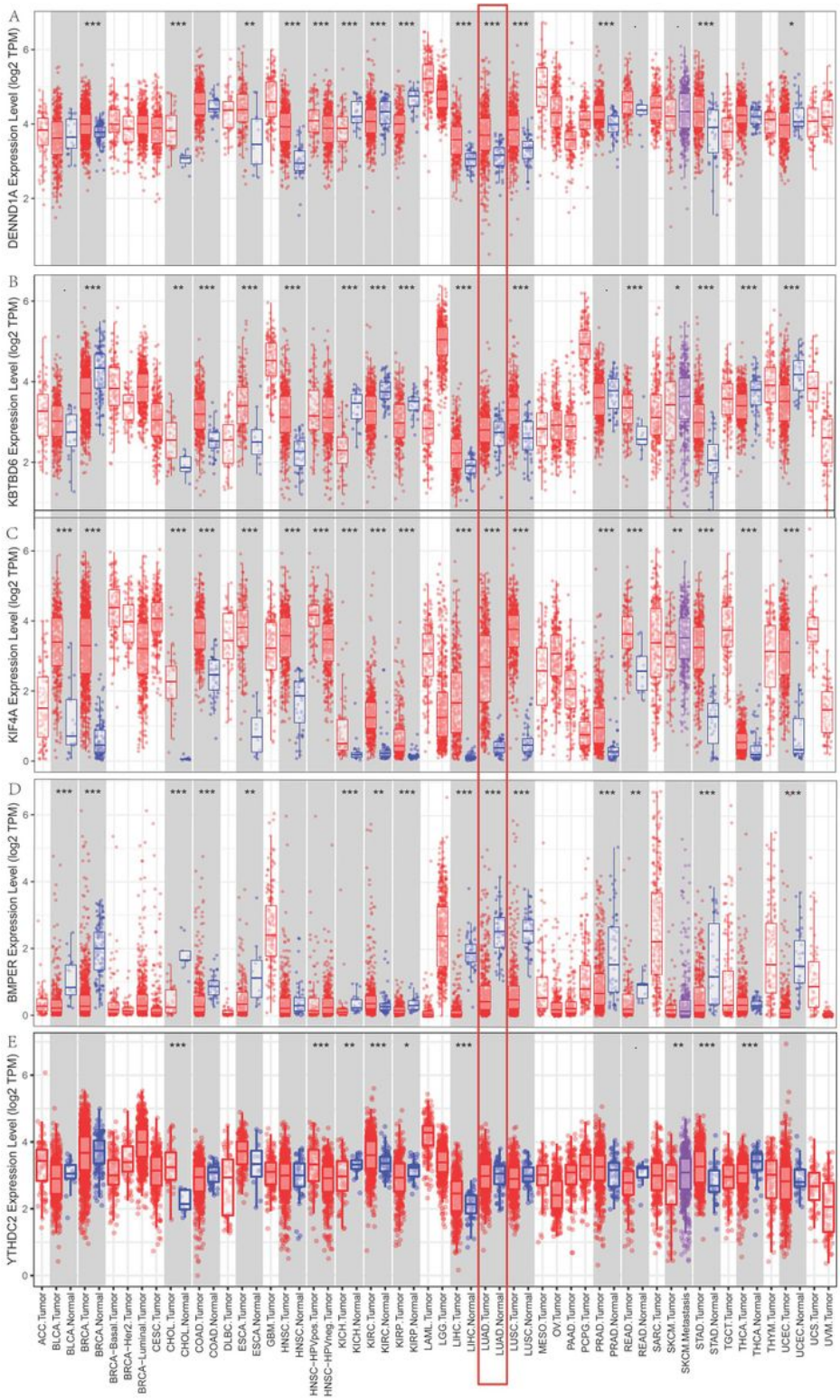

Figure 6

The expression of the five predictive genes in cancers in TIMER

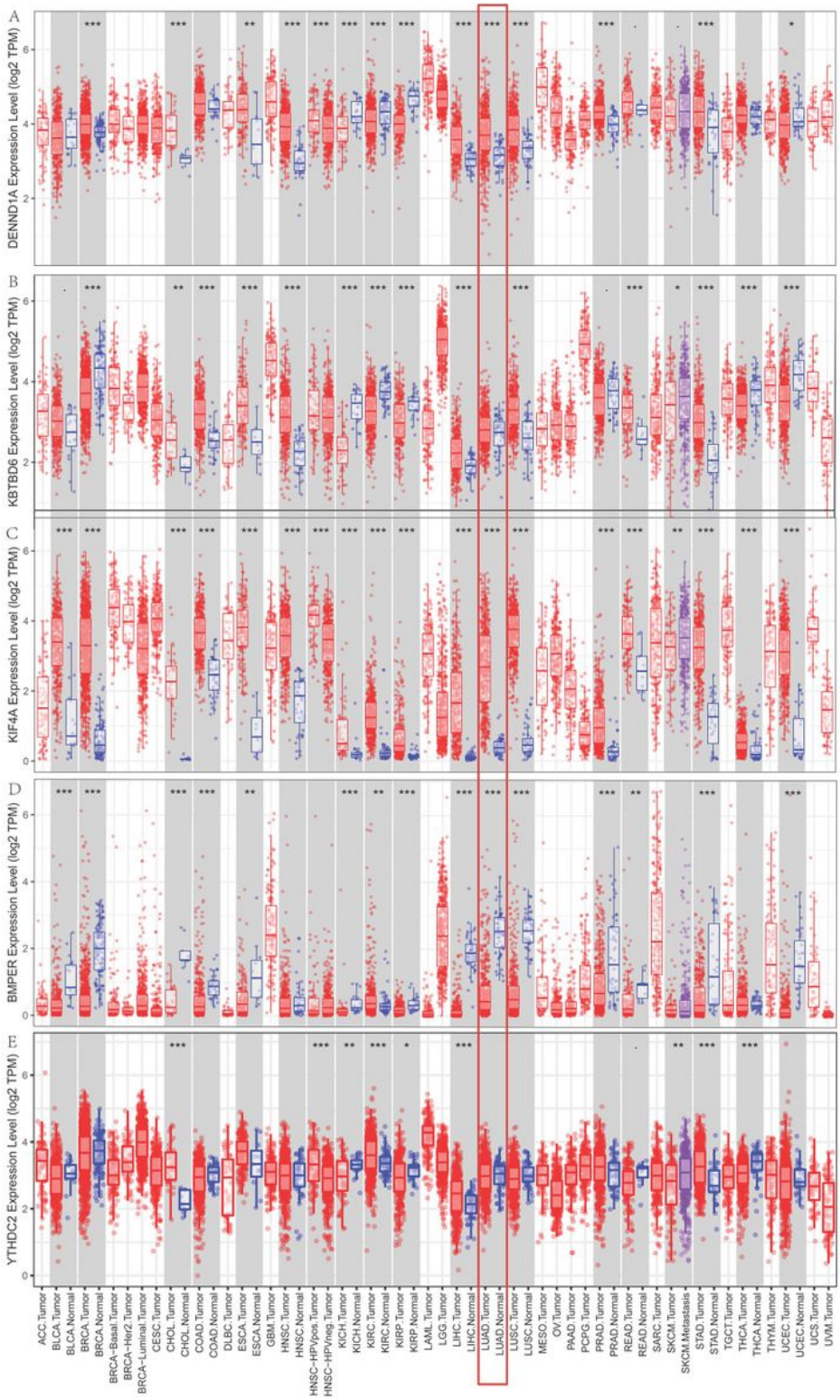

Figure 6

The expression of the five predictive genes in cancers in TIMER

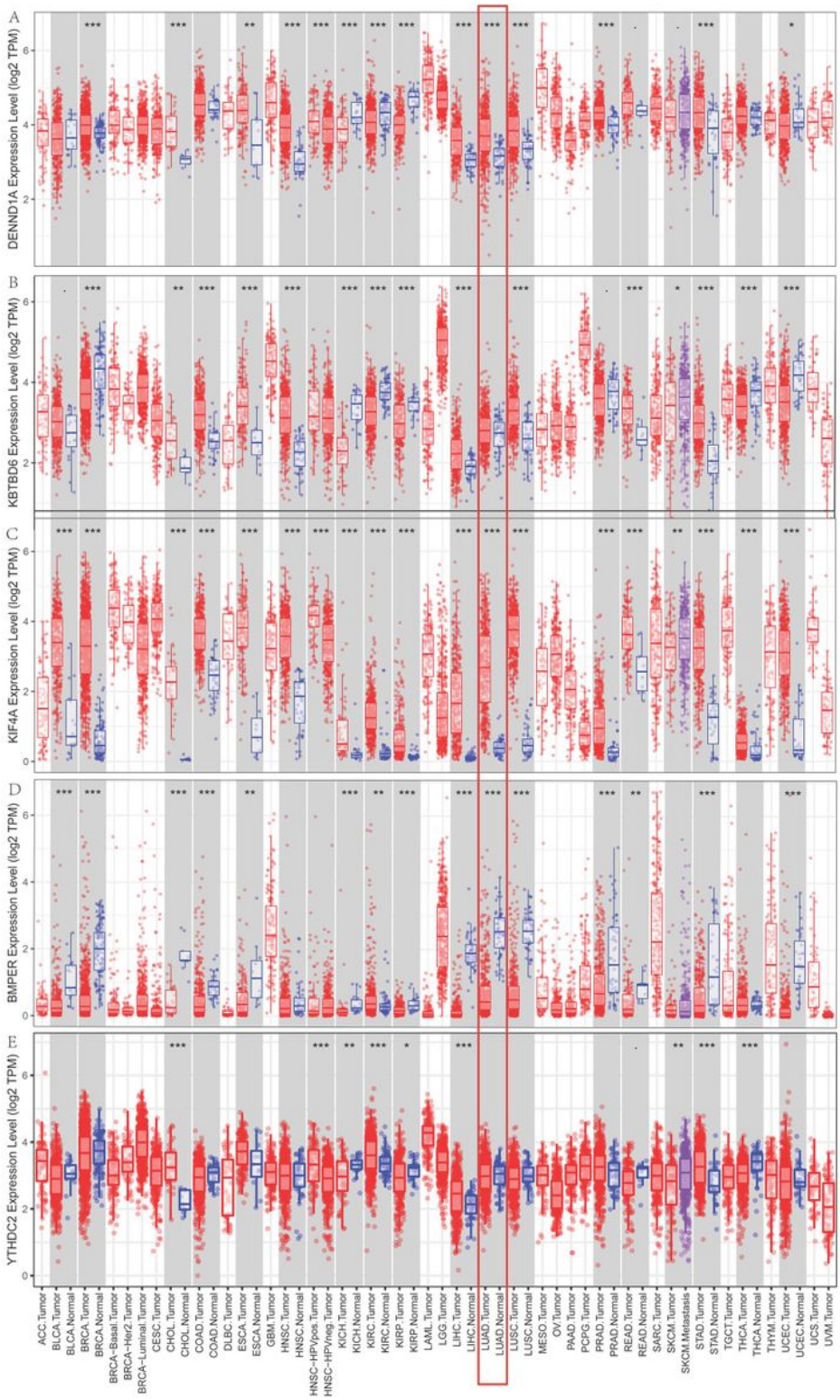

Figure 6

The expression of the five predictive genes in cancers in TIMER

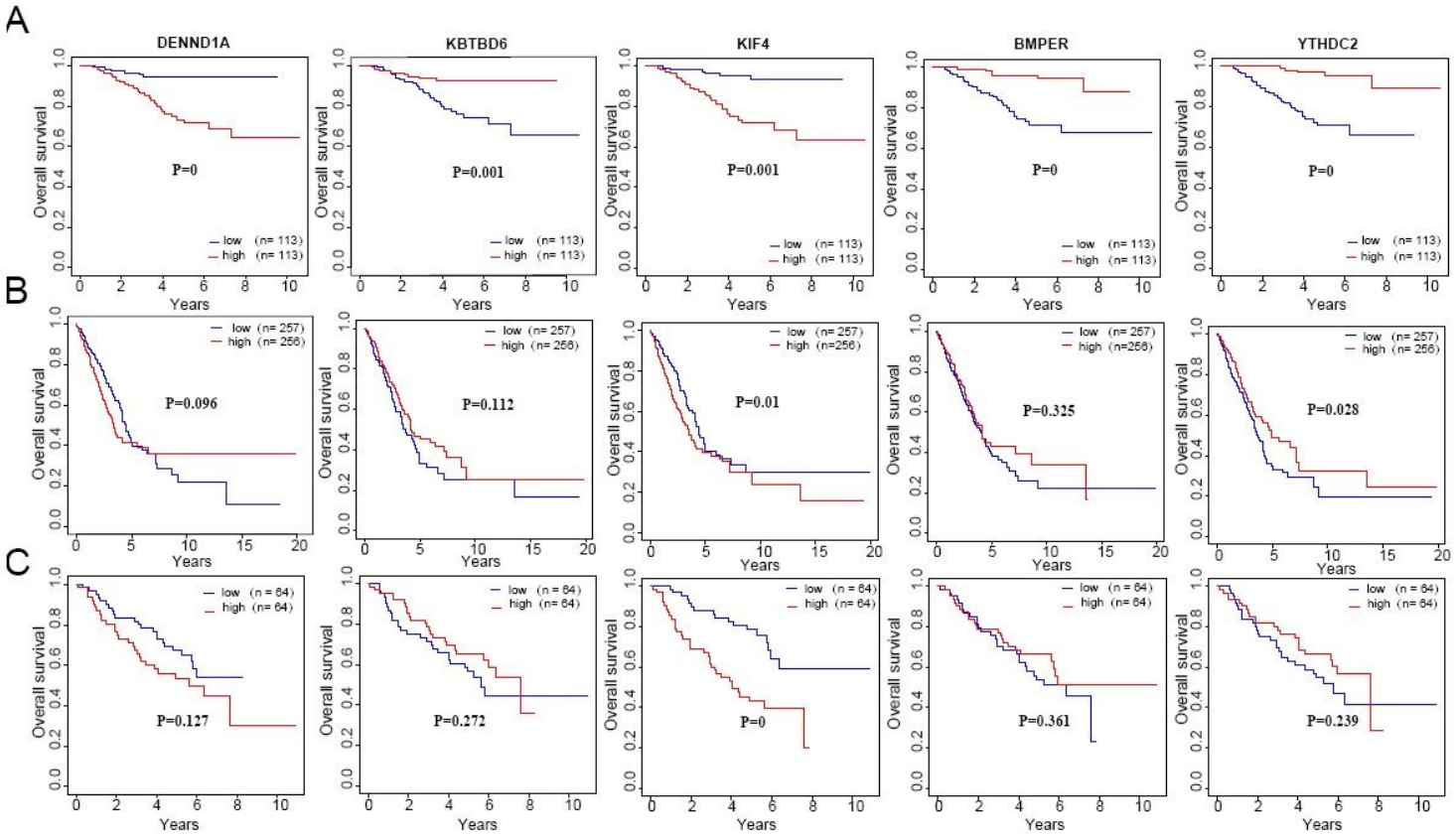

Figure 7

Comparing the low and high expression of five predictive genes in Overall survival in a (GSE31210), b (GSE50081) and c (TCGA)

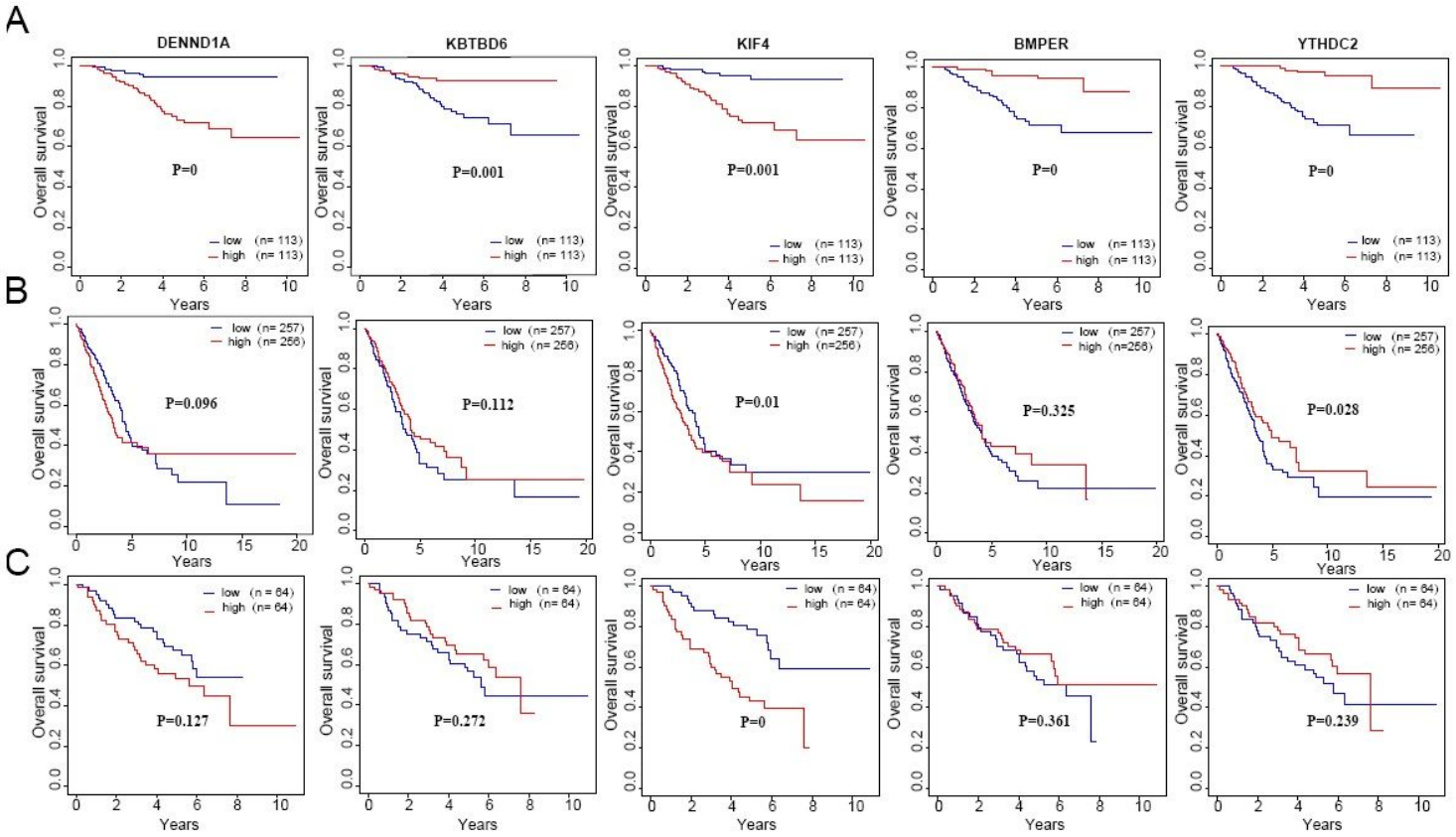

Figure 7

Comparing the low and high expression of five predictive genes in Overall survival in a (GSE31210), b (GSE50081) and c (TCGA)

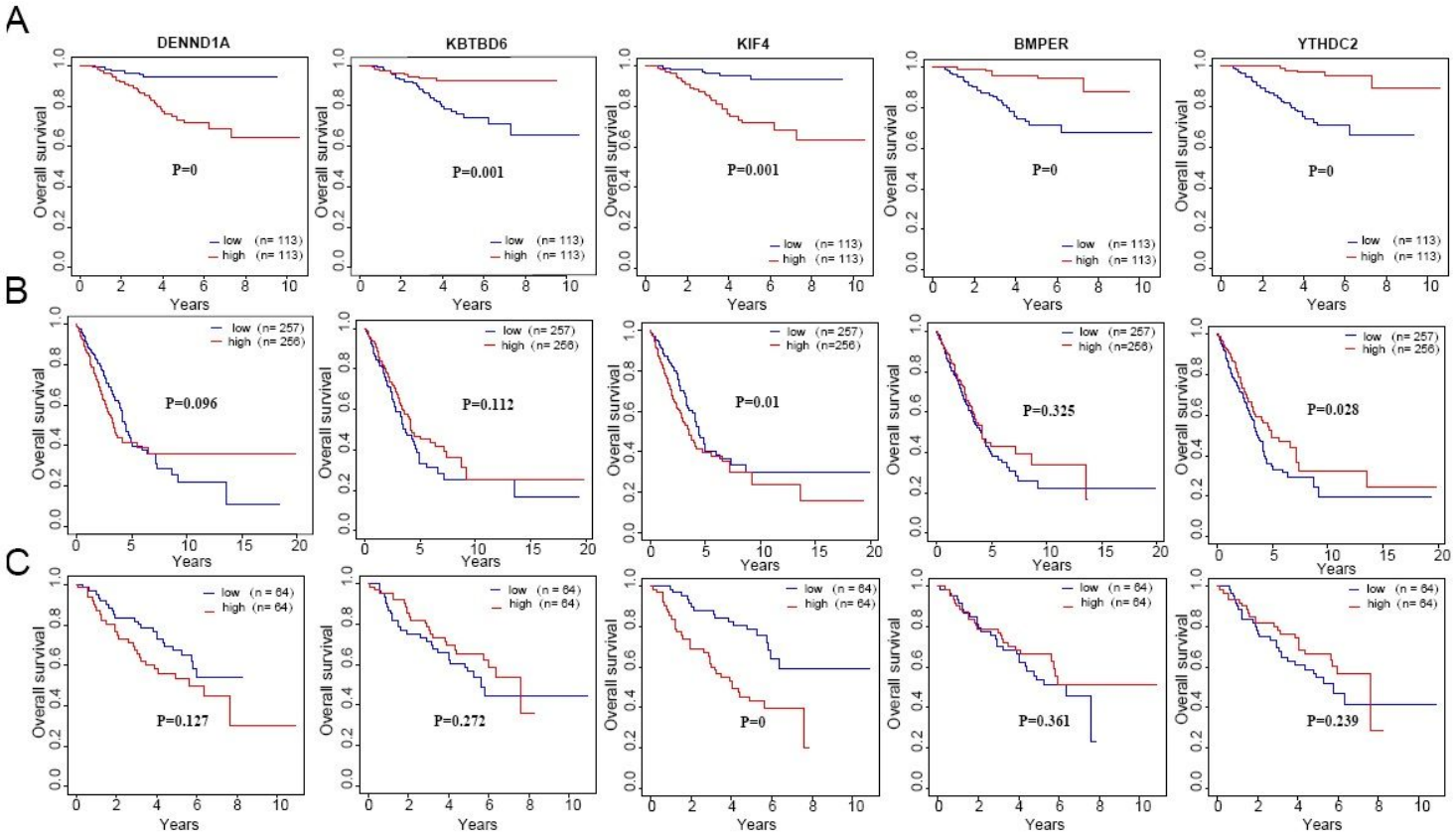

Figure 7

Comparing the low and high expression of five predictive genes in Overall survival in a (GSE31210), b (GSE50081) and c (TCGA)

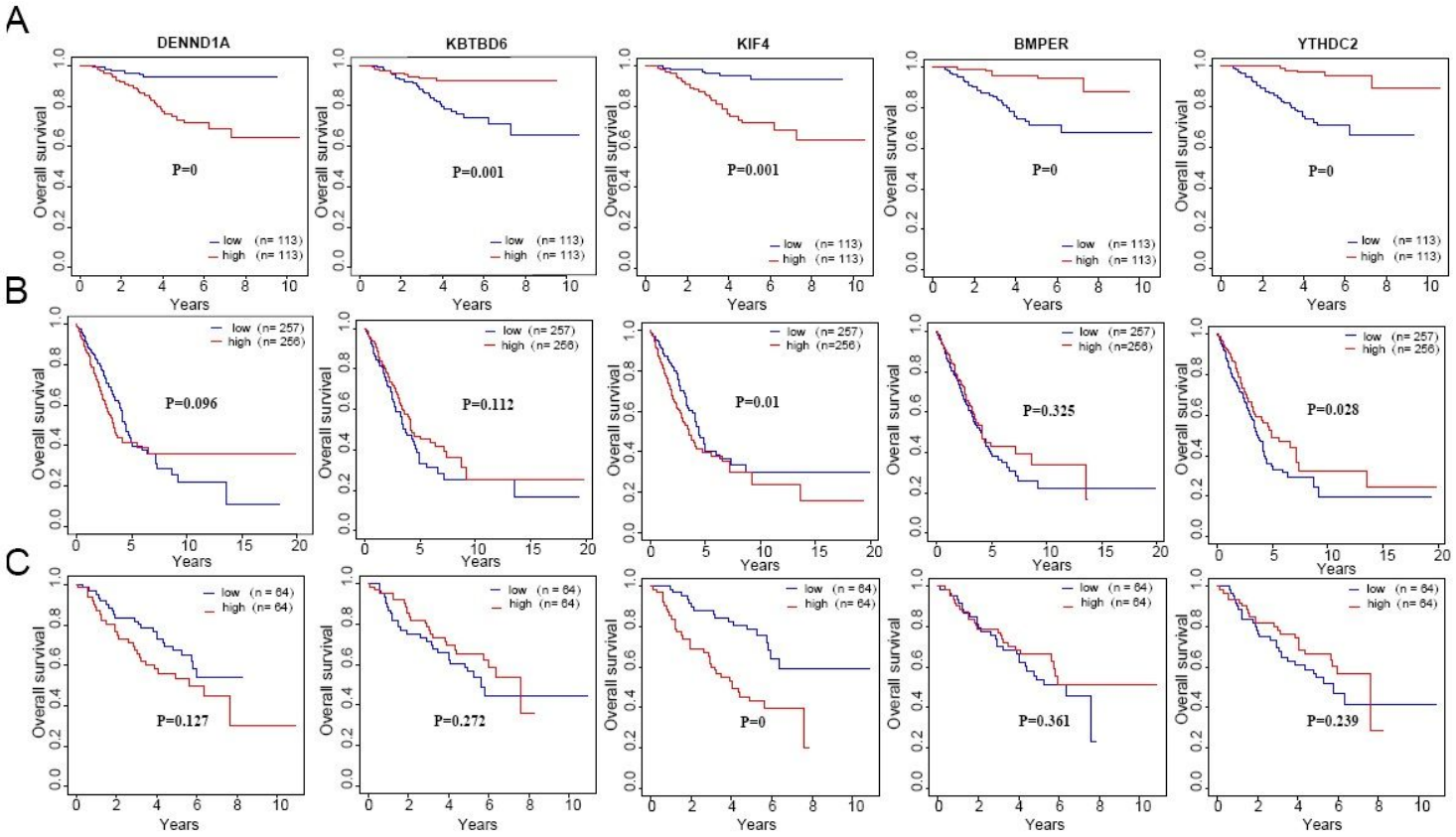

Figure 7

Comparing the low and high expression of five predictive genes in Overall survival in a (GSE31210), b (GSE50081) and c (TCGA)

Supplementary Files

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

- SupplementaryFig.S1.pdf
- SupplementaryFig.S1.pdf
- SupplementaryFig.S1.pdf
- SupplementaryFig.S1.pdf
- SupplementarytableS1.doc
- SupplementarytableS1.doc
- SupplementarytableS1.doc
- SupplementarytableS1.doc
- SupplementarytableS2.xls
- SupplementarytableS2.xls

- SupplementarytableS2.xls
- SupplementarytableS2.xls
- SupplementarytableS3.xls
- SupplementarytableS3.xls
- SupplementarytableS3.xls
- SupplementarytableS3.xls