A Lipid Metabolism-Related Genes Prognosis Biomarker Associated With The Tumor Immune Microenvironment in Colorectal Carcinoma

Chao YANG
Renmin Hospital of Wuhan University
Shuoyang Huang
Renmin Hospital of Wuhan University
Yongbin ZHENG (yongbinzheng@whu.edu.cn)
Renmin Hospital of Wuhan University
Fengyu CAO
Renmin Hospital of Wuhan University

Research Article

Keywords: Colorectal carcinoma, Lipid metabolism-related genes, Prognostic value, Tumor immune microenvironment

Posted Date: January 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-145179/v1

A lipid metabolism-related genes prognosis biomarker associated with the tumor immune microenvironment in colorectal carcinoma

Chao Yang[1], Shuoyang Huang[1], Fengyu Cao[1], Yongbin Zheng[1*]

[1]Department of Gastrointestinal Surgery, Renmin Hospital of Wuhan University, Wuhan, Hubei 430060, P.R. China

[*]Correspondence: Yongbin Zheng, E-mail: yongbinzheng@whu.edu.cn.

Abstract

Background: Lipid metabolic reprogramming was considered as a new hallmark of malignant tumors. It has been reported to play a crucial biological role in cell proliferation, energy homeostasis and signal-transduction. However, the important value of lipid metabolism-related genes(LMRGs) in prognostic prediction and the tumor immune microenvironment has not been explored by large sample studies in colorectal cancer(CRC).

Methods: In this study, the lipid metabolism status of 1086 CRC samples was analyzed using RNA expression profiles and clinical data from the Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases, of which the former was determined as training set and the latter as validation set. The risk signature was constructed by the univariate Cox regression and Least Absolute Shrinkage and Selection Operator(LASSO) COX regression. The patients were stratified into high- and low-risk groups according to the median value of the risk score.

Immune and mutation landscape between low- and high-risk CRC patients were also explored. Additionally, we established a nomogram integrating the risk signature and clinical factors to improve risk assessment of CRC patients.

Results: A four LMRGs signature, including PROCA1, CCKBR, CPT2 and FDFT1, was constructed to predict the prognosis of CRC. The risk signature as an independent prognostic factor for CRC was associated with a variety of parameters. Survival analysis showed that patients with low risk score had a better prognosis. There were different immune landscapes between low and high-risk CRC patients, especially in monocytes, dendritic cells, M0 and M2-like macrophages. Patients in the low-risk group were more likely to have higher tumor mutation burden, stem cell characteristics and level of PD-L1 expression. In addition, it was found that genes that played crucial biological functions in tumorigenesis (including TP53, PI3K and MUC16) had significant differences in mutation frequency between two groups.

Conclusion: A lipid metabolism-related risk signature for predicting the prognosis of CRC was identified in this study. Furthermore, this prognostic signature may be a potential biomarker for predicting the efficacy of chemotherapy and anti-PD-L1 therapy in CRC.

Keywords: Colorectal carcinoma, Lipid metabolism-related genes, Prognostic value, Tumor immune microenvironment

Introduction

Lipids, a general term for fatty acids, triglycerides and cholesterol, are hydrophobic or amphiphilic small molecules soluble in nonpolar solvents but not in water. Lipids generally receive less attention than other biological molecules that make up cells, such as nucleic acids and proteins[1]. Lipids are one of the important nutrients needed by the human body supplying the energy and essential fatty acids. As an indispensable part of cellular and organelle membranes in mammals, lipids participate in many key biological functions in specific cell areas under normal physiological conditions. Furthermore, lipids can act as a second messenger and involve in intracellular signaling[2].

Dyslipidemia is closely related to the occurrence and development of various metabolic diseases and malignancies. Metabolic reprogramming was considered as a new hallmark of malignant tumors[3]. Although most of the knowledge of cancer metabolism alterations has focused on glucose metabolism (called Warburg effect) at present, the abnormal lipid metabolism in cancer cells has been gradually recognized in the past few years[3]. A rapidly proliferating cancer cell requires more energy than a normal cell and meets its biological needs by activating an endogenous production pathway or increasing the intake[6]. ATP generated by fatty acid oxidation is an important energy source for cancer cells when energy provision is insufficient. Adipocyte and free fatty acid

in hypoxic tumor microenvironment could markedly conduce to cancer proliferation, progression, invasion and metastasis[7, 8]. It was reported that cancer cells relied mostly on the endogenous adipogenesis rather than uptake of exogenous fatty acids which was more common in normal cells[9]. Hence, abnormal lipid metabolism, especially fatty acids synthesis and oxidation, has increasingly been regarded as an important feature of metabolic reprogramming.

Epidemiological studies have shown that serum triglyceride levels were associated with susceptibility to colorectal cancer (CRC)[10, 11]. The research of Wang et al confirmed that alterations in the abundance of individual lipids were observed in CRC using shotgun lipidomics[12]. The study showed that the expression of lipogenic enzyme involved in *de novo* adipogenesis (fatty acid synthesis) was increased in CRC, including fatty-acid synthase, acetyl-CoA carboxylase, carnitine palmitoyltransferase, while the enzyme involved in fatty acid oxidation was decreased. As a mitochondrial serine/threonine phosphatase, PGAM5 regulates a variety of metabolic pathways in vivo. Research by Zhu et al. showed that blocking PGAM5 would reduce lipid metabolism and inhibit the occurrence of CRC in mice[13]. Gong and his colleagues showed that the reprogramming of lipid metabolism in tumor-associated fibroblasts in the tumor microenvironment significantly enhanced the invasion and metastasis of CRC[14]. Besides, drug resistance in antiangiogenic therapy

is frequently appeared in cancer treatment, and underlying molecular mechanism might include lipid metabolism reprogramming[15]. Iwamoto et al noticed that blocking of carnitine palmitoyl transferase 1A, a key enzyme in lipid metabolism, can obviously revert the sensitivity of antiangiogenic therapy. Thus, they brought a promising cancer therapy concept to overcome drug resistance by combining conventional therapy and targeted lipid metabolism[8].

Existing studies have confirmed the closely relationship between altered lipid metabolism and CRC on tumorigenesis, progression and treatment. Previous studies showed that lipid metabolism-related genes(LMRGs) signature had high prognostic value in papillary thyroid cancer or diffuse glioma[16, 17]. However, the prognostic value of LMRGs in CRC has not been verified by large sample studies. The present study aimed to develop a novel risk signature based on the LMRGs for providing additional information on risk assessment, as well as clinical-decision making in CRC.

Materials and Methods

Study population and data collection

The flow chart of this research was presented in Figure 1. Level 3 RNA sequencing data(RNA-seq), mutation data and matched clinical information were obtained from the TCGA CRC cohort (https://portal.gdc.cancer.gov/repository) as a training set. The raw CEL

data and paired clinical information of CRC patients was downloaded from GEO GSE39583 cohort (https://www.ncbi.nlm.nih.gov/geo/) as a validation group.

Both TCGA and GEO database are publicly available, thus the present study was exempted from the approval of local ethics committees.

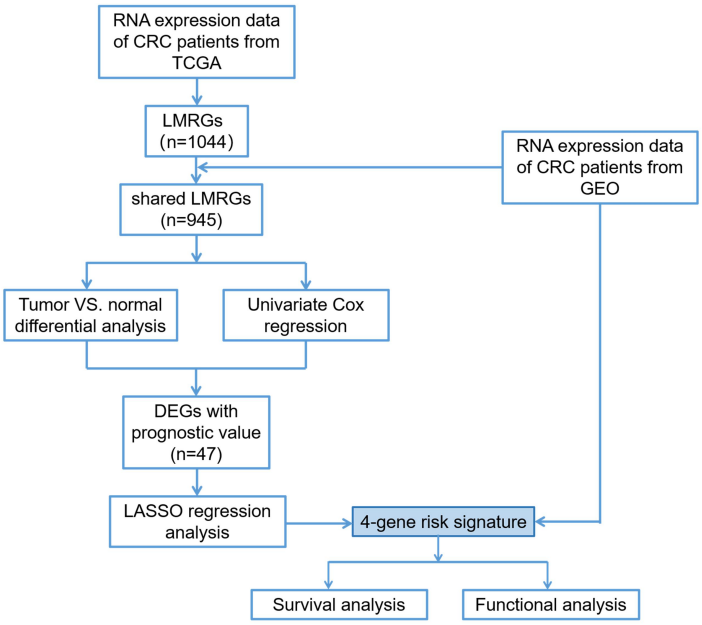

Figure 1 Flowchart of study design. (LMRGs: lipid metabolism-related genes; DEGs: differentially expressed genes; LASSO: the Least Absolute Shrinkage and Selection Operator; TCGA: the Cancer Genome Atlas; GEO: the Gene Expression Omnibus)

Identification of IMRGs

The five IMRGs sets were collected from the Molecular Signature Database[17](Supplementary Table 1). A total of 1044 genes were found to be involved in lipid metabolism process after removing overlapping genes.

Data processing

The RNA-seq transcriptome data(FPKM) from TCGA cohort were converted into $\log_2$(TPM+1) for normalized counts. The robust multichip

average (RMA) was used to normalize the raw data from GEO cohort by the R package affy.

Criteria for study exclusion were: (1) Patients with unknown survival status, follow-up information, and disease stage. (2) Patients who died within a follow-up period of 30 days. Consequently, 544 cases (500 tumor and 44 normal samples) meeting the criteria were included in the training set and 542 cases (523 tumor and 19 normal samples) in validation set (Supplementary Table 2). The TCGA cohort was used to establish a risk signature, and the GEO set were used for validation.

Construction a risk signature based on the LMRGs

The shared LMRGs in the GEO and TCGA sets were selected for subsequent analysis. Differentially expressed genes(DEGs) between tumor and normal tissue samples were screened in TCGA cohort by the "limma" R package with a false discovery rate(FDR) <0.05. Meanwhile, the LMRGs with prognostic value were identified using univariate Cox analysis. Nextly, the overlapping genes between genes with prognostic value and DEGs were identified by venn diagram for the follow analysis. Subsequently, the Least Absolute Shrinkage and Selection Operator(LASSO) COX regression was used to select the best predicting model based on these overlapping genes in TCGA CRC patients. The LASSO analysis was performed using R package "glmnet" and the optimal value of penalty parameter was determined by 10-fold

cross-validation. The risk signatures were generated from the TCGA and GEO cohort according to the expression of LMRGs and corresponding coefficients simultaneously. The risk score of each sample was calculated as the following formula. (expGene: the expression level of LMRGs in TCGA or GEO cohort; Coef: the coefficient of LMRGs in LASSO regression model in training set)

$$\text{risk } score = \sum_{i=1}^{n} (\exp Genei \times Coefi)$$

Prognostic value of the risk signature in training and validation group

The patients were stratified into high- and low-risk groups according to the median value of the risk score. Kaplan Meier(K-M) survival curve with the log-rank test were performed to show the prognostic ability of the risk signature. Additionally, the area under curves(AUCs) of receiver operating characteristic(ROC) at 1-, 3-, and 5-year were calculated simultaneously to evaluate the performance of those two signatures.

Gene set enrichment analysis (GSEA)

To explore potential molecular mechanism between the two groups, the GSEA was carried out between the high- and low-risk groups. The annotated gene set list, h.all.v7.2.symbols.gmt (Hallmarks), was selected as the reference gene set from the Molecular Signature Database[18].

Independence of the risk signature from other clinicopathological parameters

To determine the independence of the risk signature from other clinical parameters, univariate and multivariate analysis of risk score with age, gender and tumor stage were performed. The forest plots were used to show the independence of the risk score.

Correlation between the risk signature and other clinicopathological parameters

The association was further explored between the risk signature and clinicopathological parameters including age, gender, tumor stage, pathological T stage, N stage and M stage. Patients were stratified into subgroups of age≤65 years and age≥65 years, female and male, pathological tumor stage I + II and stage III + IV, T1 + T2 and T3 + 4, N0 and N1 + 2, M0 and M1. K-M survival analysis of the aforementioned paired subgroups was performed.

Cancer stem cells are highly dependent on lipid metabolism to maintain their stem cell characteristics. One study showed that cancer stem cells in CRC had higher lipid metabolism level than tumor cells or normal colonic epithelial cells[19]. Malta et al had developed a novel analysis tool to assess the stemness features based on the gene expression[20]. In this study, the mRNA expression-based stemness index (mRNAsi) of each CRC patient was downloaded from the research of Malta et al. Besides, CD133 is a marker gene of many tumor stem cells. Hence, the relationship of risk score with mRNAsi score and CD133

mRNA also were analyzed.

Nomogram construction and validation

A nomogram integrating the risk signature and other clinicopathological factors was established for prognostic evaluation using the "rms" package. The AUCs of ROC were demonstrated to assess the predictive capability of the nomogram. Calibration curves were simultaneously established to assess the predictive accuracy of the nomogram.

Estimation of relative abundance of immune cell types in different risk groups

The CIBERSORT algorithm (https://cibersort.stanford.edu), an approach to quantify the relative abundance of immune cell types based on specific gene expression profiles, was used to assess the distribution of 22 immune cell types in CRC samples[21]. Moreover, the *P*-value, correlation coefficient and root mean squared error were also presented to evaluate the accuracy of the results in each patient. Samples with *P*-value < 0.05 were retained to compare immune cell abundance in different risk groups.

Mutation analysis

The top 20 genes with the highest mutation frequency in CRC were analyzed in both high and low risk groups using the GenVisR R package.

The tumor mutational burden(TMB) were associated with the

efficacy of curative resection combined with followed adjuvant chemotherapy (fluoropyrimidine plus oxaliplatin regimen) in CRC[22, 23]. TMB was also an independent predictor of response to treatment with immunocheckpoint inhibitors (ICPI) in a variety of tumors[24, 25]. Therefore, the relationship between risk score and TMB was also explored in this study. The TMB score was generated by the total number of mutations divided by the number of exons in each sample. The exon size is often approximately estimated at 38 megabase.

Statistical analysis

All of statistical analyses and drawing in this study were conducted by the R (version 4.0.2) software or Graphpad Prism (version 8.3.0). T-test was applied to analyze differences of continuous variables. The Fisher's exact or Chi square test were employed for the comparison of categorical variables. Log-Rank test was used to estimate the difference among K-M survival curves. *P*-value<0.05(two-tailed) was considered significant.

RESULT

Identification of differentially expressed and prognosis-related genes in LMRGs

The expression data of LMRGs was extracted from TCGA and GEO cohort, respectively. A total of 945 shared LMRGs were matched. There were 729 differentially expressed genes(DEGs) between normal and

tumor tissues, including 365 up-regulated and 364 down-regulated have been identified in training cohort when cut-off was set as FDR< 0.05 (Fig. 2A). At the same time, univariate Cox analysis was employed to filtrate genes with prognostic value in the 945 intersected genes. Finally, 57 LMRGs shown to be prognostically related in the training set. Conclusively, there were 47 shared LMRGs that were both differentially expressed and prognostic genes (Fig. 2B, C).

Functional enrichment analysis showed that the enriched terms were related to lipid metabolism. The top five terms in biological process included. steroid metabolic process, glycerolipid metabolic process, lipid localization, phospholipid metabolic process and lipid transport (Fig. 2D). Kyoto Encyclopedia of Genes and Genomes(KEGG) analysis found that these 47 LMRGs were mainly engaged in fat digestion and absorption, PPAR signaling pathway, glycerolipid metabolism, cholesterol metabolism and adipocytokine signaling pathway (Fig. 2E).

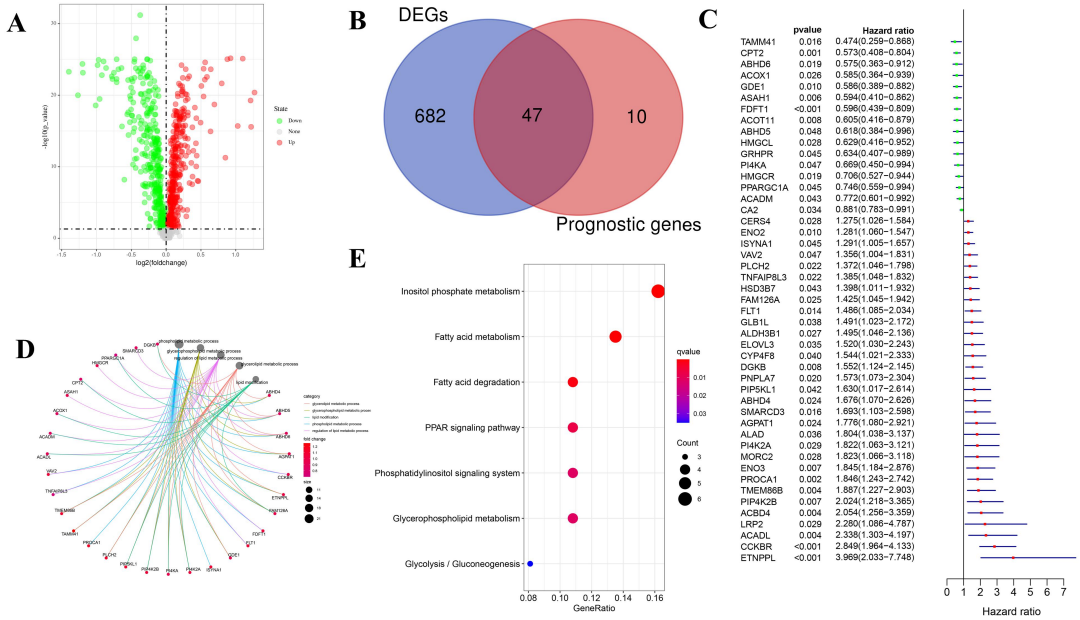

Figure 2 Identification of the differentially expressed and prognosis-related genes in the TCGA cohort. (A) Volcano map showed the differentially expressed genes between normal and tumor tissues. Red meant up-regulated genes and green represented down-regulated genes. (B) Venn diagram showed the 47 differentially expressed genes which also had prognostic value. (C) Forest plot for the univariate Cox regression analysis of the 47 overlapping genes. (D) The main GO enrichment terms of the 47 overlapping genes. (E) The most significant KEGG pathways of the 47 overlapping genes. (GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes)

Construction and validation of gene signature

Nextly, the expression profile of the 47 LMRGs was used to establish a risk signature using the LASSO COX regression analysis (Fig. 3A, 3B). Finally, 4 LMRGs, namely PROCA1, CCKBR, CPT2 and FDFT1, were identified to establish the optimal lipid metabolism-related risk signature. The risk score for each patient was calculated by following same formula: risk score=(PROCA1*0.03071) + (CCKBR*0.58956) +(CPT2*0.00972) +(FDFT1*0.01381).

The predictive value of this risk signature was evaluate by ROC. The AUCs of this signature were 0.6901 at 1 year, 0.6776 at 3 years and

0.5945 at 5 years in training set (Fig. 3C). The patients were dichotomized to two risk group according to the median risk score. Kaplan-Meier survival curves showed that high-risk patients had a significantly shorter overall survival than low-risk cases in training set (Fig. 3D, $P<0.001$). The Fig.3E showed that more patients died in the high-risk group, while the majority survived in the low-risk group. In training set, the principal components analysis(PCA) was used to get the expression pattern in low and high-risk groups. The groups did not present a significant distinction on the risk status when PCA was performed with the all genes (Fig.3F). However, the risk groups were observably distinguished regarding to the risk signature(Fig.3G).

In validation set, the high-risk patients still had a significantly shorter overall survival than low-risk group (Fig. 3H, $P<0.05$). And, the AUCs were 0.6152 at 1 year, 0.6332 at 3 years and 0.6358 at 5 years (Fig. 3I). The risk score distribution curve and survival status were showed simultaneously in Fig. 3J.

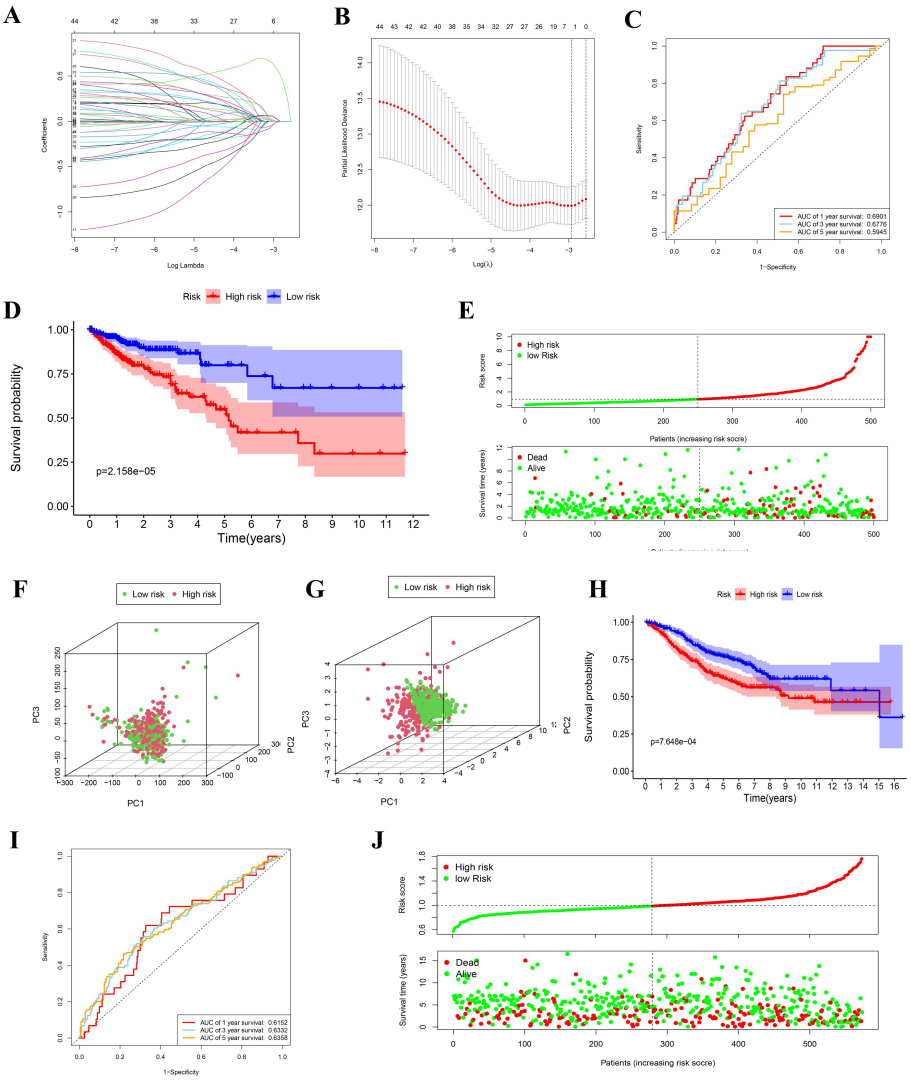

Figure 3 Development and validation of the risk signature. (A) The LASSO COX regression coefficient profiles of the 47 LMRGs. (B) 10-time cross-validation for tuning parameter selection in the LASSO mode. (C) ROC at 1, 3 and 5 years (AUCs = 0.6901, 0.6776 and 0.5945). (D) The survival curve in the high and low risk groups showed significant differences (P< 0.001). (E) Risk score distribution(above) and survival status(bottom) for patients in different risk groups. (F) The PCA showed that the distribution in different directions were general on the basis of the whole gene sets. (G) The PCA showed that the risk signature could distinguish patients in different risk groups markedly. (H) The survival curves showed significant differences in validation sets (P< 0.001). (I) ROC at 1, 3 and 5 years in validation set (AUCs = 0.6152, 0.6332 and 0.6358). (E) Risk score distribution(above) and survival status(bottom) for patients in different risk groups of the validation set. (LASSO: Least Absolute Shrinkage and Selection Operator; the PCA: the principal components analysis; AUCs: the area under curves; ROC: receiver operating characteristic)

Independent prognostic value of the risk signature

The result showed that the risk score based on the four-gene

signature was an independent prognostic factor in training set with hazard ratio (HR)= 6.146 (95% confidence interval (CI)= 3.376-11.190; *P*<0.001, Fig. 4A) by univariate Cox analysis and HR=4.315 (95% CI= 2.321-8.022; *P*<0.001, Fig. 4B) by multivariate Cox analysis. Additionally, similar result was acquired in the validation set showing the independence of risk signature with HR= 5.822 (95% CI= 2.567-13.205; *P*<0.001, Fig. 4C) and 5.395 (95% CI= 2.272-12.809; *P*<0.001, Fig. 4D) by univariate and multivariate Cox analysis, respectively.

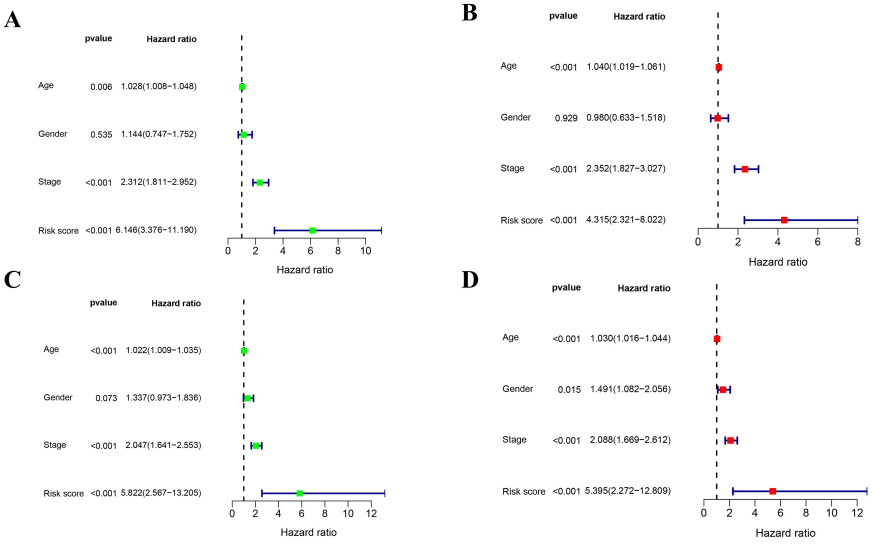

Figure 4 Univariate and multivariate Cox regression analysis showed risk score was an independent prognostic factor both in training and validation set. (A)The result of univariate Cox regression analysis in training set. (B) The result of multivariate Cox regression analysis in training set. (C) The result of univariate Cox regression analysis in validation set. (D) The result of multivariate Cox regression analysis in validation set.

Relationships between the risk score and clinicopathological features

Risk score was significantly associated with age (*P*<0.0001, Fig. 5A) but not with gender (*P*=0.4168, Fig. 5B). In addition, the relationship between obesity and risk score was also explored. The patient was

defined as obesity when body mass index (BMI) was equal or greater than 30[26]. Interestingly, no significant difference was found in lipid metabolism-related risk score between the normal weight and obesity groups (P=0.4168, Fig. 5C).

The result showed that the risk score was positively correlated with the degree of tumor progression. For instance, the risk score in TNM stage III + IV patients were markedly higher than those in tumor stage I + II (P<0.0001, Fig. 5D). Similar results have been achieved in other subgroups, including pathological T stage (I + II vs. III + IV, P<0.0001, Fig. 5E), N stage (N0 vs. N1 + 2, P<0.0001, Fig. 5F), M stage (M0 vs. M1, P = 0.0024, Fig. 5G) and tumor location (left vs. right, P<0.0001, Fig. 5H). Another finding of this study was that there were higher stem cell characteristics in the low-risk group, including significantly higher level of CD133 mRNA (P = 0.0479, Fig. 5I) and mRNAsi (P<0.0001, Fig.5J).

Subgroup analysis was performed to further assess whether the risk signature still had prognostic value within specific clinical parameters. The results showed that this risk signature remained a powerful prognostic prediction to each subgroup in age (age ≥ 65 or < 65, Fig. 5K and 5L), gender (male or female, Fig. 5M and 5N), tumor stage (III + IV, Fig. 5O), T stage (III + IV, Fig. 5P), N stage (N0 and N1 + 2, Fig. 5Q and 5R), M stage (M0 and M1, Supplementary Fig.1). Obviously, the overall

survival of high-risk patients was significantly shorter than that of low-risk patients in subgroups of different clinicopathological characteristics.

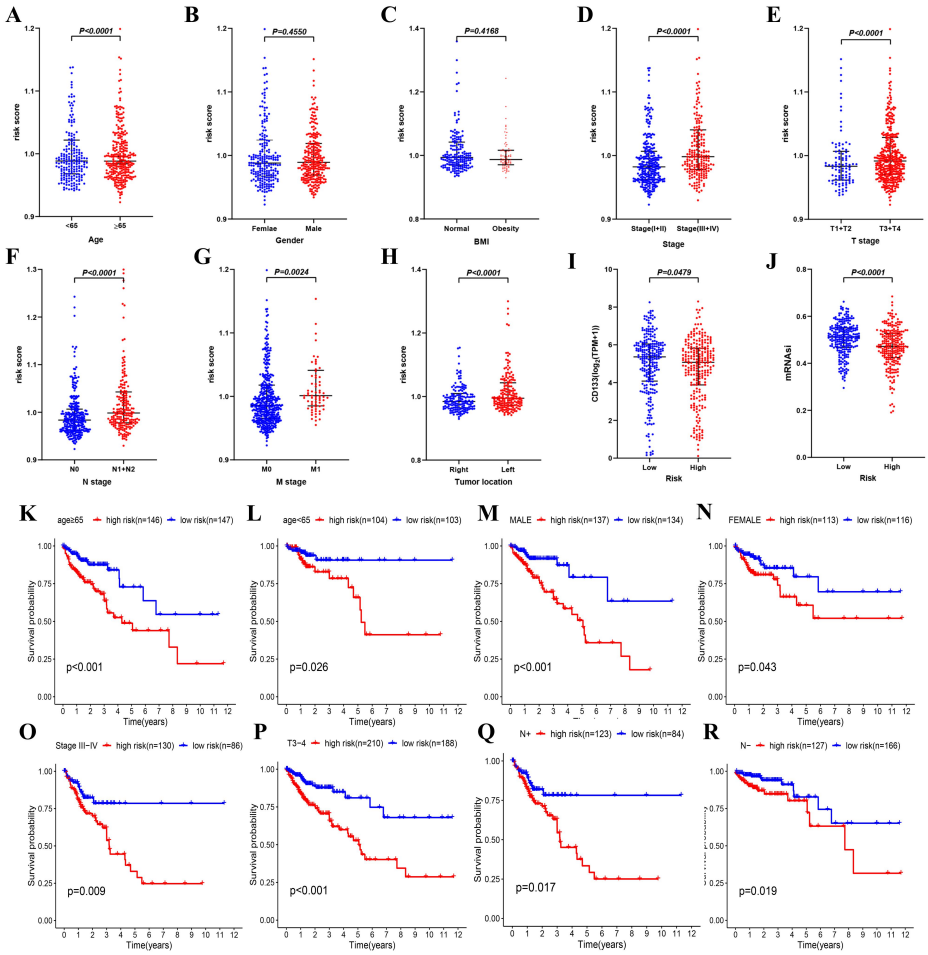

Figure 5 Stratified analysis of the risk signature in the training set. The relationships between the risk signature and age (A), gender (B), BMI (C), tumor stage (D), T stage (E), N stage (F), M stage (G), tumor location (H), CD133 mRNA (I) and mRNAsi (J). A poor prognosis was observed in high risk group in majority of clinical stratification, including age (K-L), gender (M-N), tumor stage (O), T stage (P) and N stage (Q-R). (BMI: Body Mass Index; mRNAsi: the mRNA expression-based stemness index)

Construction and validation of a nomogram combining clinicopathological features and risk signature

A nomogram was constructed based on several factors, namely the age, gender, disease stage and risk score, to provide a visualization

method for clinicians to predict the probability of 1-, 3- and 5-year overall survival in CRC patients. Each patient would get a total point from this nomogram, and higher total point indicated worse outcome for the patient (Fig. 6A). This nomogram had a high potential for clinical utility with the AUC of 0.7652 at 1 year, 0.8058 at 3 years and 0.7972 at 5 years by ROC in training set (Fig. 6B). Moreover, the calibration plots indicated that the actual observation probability was very close to the predictive probability of the nomogram (Fig. 6C).

In validation set, the nomogram was showed in Fig. 6D. The AUCs of ROC were 0.7325 at 1 year, 0.7280 at 3 years and 0.7173 at 5 years (Fig. 6E). The predictive probability of nomogram and the actual observed values also showed good consistency (Fig. 6F).

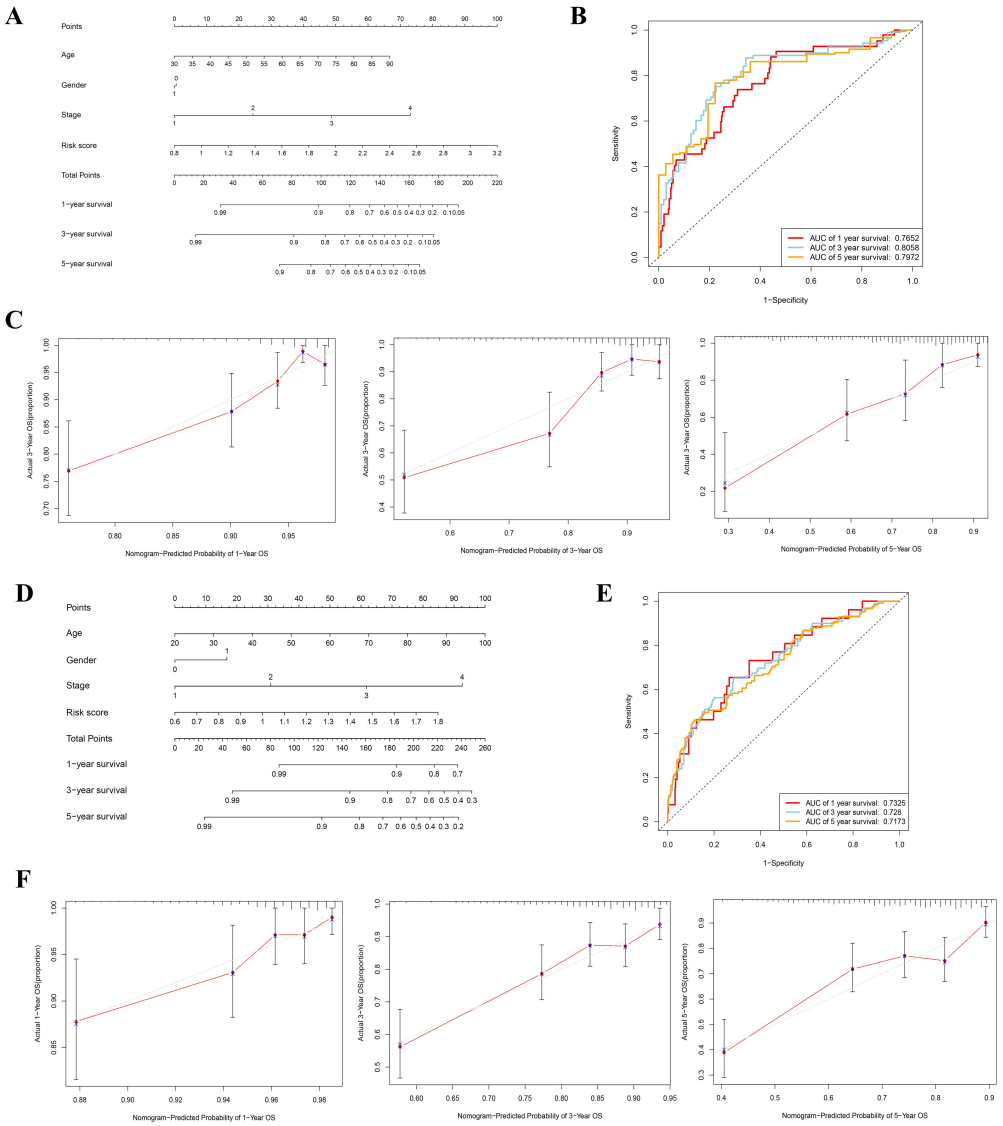

Figure 6 Construction and validation of the nomogram combining clinicopathological features and risk signature. (A) The nomogram in training set. (B) The prognostic value of nomogram for predicting 1-, 3-, and 5-year overall survival rate in training set. (C) Calibration plots of the nomogram in training set. (D) The nomogram in validation set. (E) The prognostic value of nomogram for predicting 1-, 3-, and 5-year overall survival rate in validation set. (F) Calibration plots of the nomogram in validation set.

Immune landscapes differ between low- and high-risk CRC patients

The abundance differences of immune cells between the high- and low-risk groups were further investigated. The result showed the proportions of 22 immune cells in each sample varied markedly (Fig. 7A and 7B). 487 samples of the total set were included in the subsequent

analysis when the cut-off value was set to *P* less than 0.05. The difference analysis showed that monocytes (*P*=0.011), M0 (*P*=0.001) and M2-like macrophages (*P*<0.001) were enriched in the high-risk group dramatically, while the patients in low-risk group had higher level of plasma cells (*P*=0.012), T cells CD4 memory resting (*P*=0.008) and dendritic cells activated(*P*=0.004) (Fig. 7C).

PD-L1, as an immune checkpoint molecule, was often closely related to immunotherapy response[22, 27]. This study indicated the risk score was negatively correlated with PD-L1 mRNA in training set (R=-0.125, *P*=0.0052, Fig. 7D).

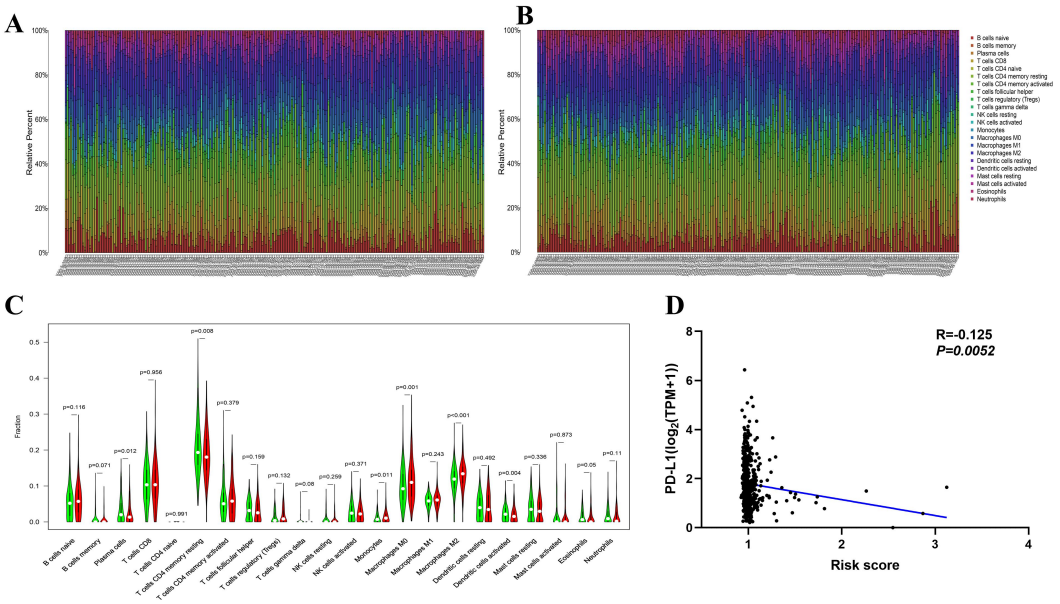

Figure 7 The risk signature was correlated with tumor-infiltrating immune cells. (A) The distribution pattern of immune cells in high risk group. (B) The distribution pattern of immune cells in low risk group. (C) There were significant differences in the abundance of many immune cells in two groups, especially monocytes (*P*=0.011), M0 (*P*=0.001) and M2-like macrophages (*P*<0.001). (Green represented the low risk group, and red meant the high risk group.) (D) Risk score was negatively correlated with PD-L1 mRNA (R=-0.125, *P*=0.0052).

Mutation landscape of the risk signature

The top 20 genes with the highest frequency mutation in CRC were shown in two waterfall plots, respectively (Fig. 8A and 8B). There was different genetic mutation landscape in high and low risk groups. Genes that played crucial biological functions in tumorigenesis, including TP53, PI3K and MUC16, had significant differences in mutation frequency between high and low risk groups (Fig. 8C).

A total of 429 samples of TMB data were obtained for this analysis. The analysis result displayed that the TMB score was higher in low-risk group (n=212) than in high-risk (n=217) (*P*=0.0002, Fig. 8D). This could be one of the reasons for the better prognosis of patients in the low-risk group.

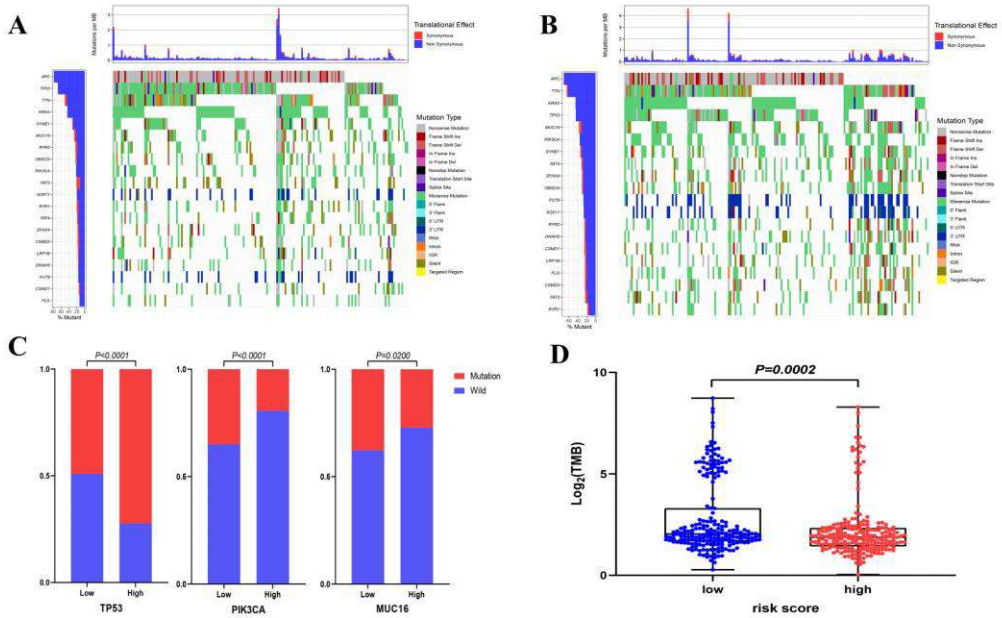

Figure 8 Mutation landscape between the low and high risk. (A) The top 20 genes with the highest frequency of mutation in high group were shown by waterfall plot. (B) The top 20 genes with the highest frequency of mutations in low group were shown by waterfall plot. (C) The key genes, including TP53, PI3K and MUC16, had significant differences in mutation frequency between high and low risk groups. (D) The low risk group had a higher TMB level(*P*=0.0002). (TMB: tumor mutational burden).

Distinct biological function pathways characterize high- and low-risk CRC patients

The GSEA was also performed to explore whether relevant signaling pathways differ between two risk groups. The result showed that the high-risk group was associated with Hedgehog signaling, KRAS signaling, Wnt/β catenin signaling, apical junction, epithelial mesenchymal transition and angiogenesis. While functional enrichment in the low-risk group focused on energy-metabolism-related functions, including fatty acid metabolism and oxidative phosphorylation (Fig. 9).

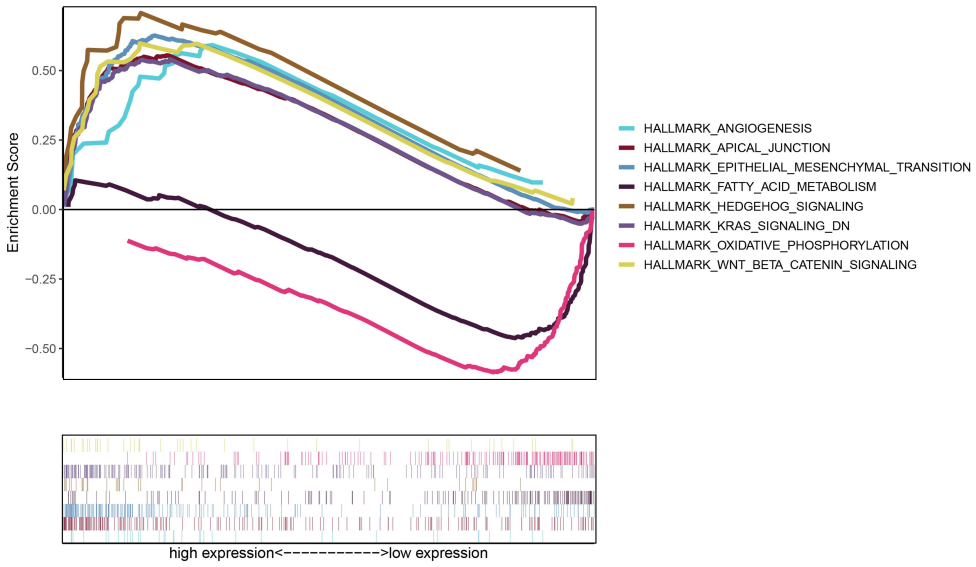

Figure 9 Distinct biological function pathways characterize in two groups. The results showed that the high risk group was associated with Hedgehog signaling, KRAS signaling, Wnt/β catenin signaling, apical junction, epithelial mesenchymal transition and angiogenesis, while functional enrichment in the low risk group focused on energy-metabolism-related functions, including fatty acid metabolism and oxidative phosphorylation.

Discussion

According to the latest statistics, CRC ranked the third most common cancer in male while the second in female, accounting for about

10 percent of cancer-related deaths worldwide each year[28]. Abnormal metabolism was often observed in CRC, which was conducive to the continuous proliferation and extensive metastasis of tumor cells[29]. Metabolic reprogramming, including abnormal aerobic glycolysis and glutamine metabolism, was one of the characteristics of malignant tumors. Lipid metabolic reprogramming has been reported to play a crucial biological role in cell proliferation, energy homeostasis and signal-transduction[30, 31]. In-depth study of lipids metabolomics characteristics can result in better understanding the progression of tumors and provide potential metabolic targets for the development of new treatment methods[31]. In this study, the lipid metabolism status of 1086 CRC samples was analyzed using gene expression data from the public database, and a risk signature was constructed. This was the first CRC prognostic risk signature constructed from LMRGs so far. The risk signature could well distinguish the high risk patients among CRC. Pathological parameters were often used for risk stratification management and tumor-node-metastasis (TNM) staging system was currently the most widely used risk evaluation tool. The risk signature based on four LMRGs might be an effective supplement to TNM staging to provide more accurate prognostic information. Nomogram was a commonly used visualization tool for prognosis assessment in clinical practice. The nomogram, combining the risk signature and clinical

characteristics, have revealed better potential to contribute to risk stratification than clinical parameters alone. In general, the risk signature was based on the expression level of specific genes which was an economy-friendly and more practical method compared to next generation sequencing.

The correlation between the risk signature and 22 immune cells was evaluated to explore the role of lipid metabolism in the tumor immune microenvironment. According to our investigation, risk signature based on lipid metabolism was related to the level of monocyte infiltration, especially the M2-like macrophages. In high score group, there were superior level of M2-like macrophages, which suggested that high-risk patients might have an unfavorable tumor immune microenvironment. Tumor-associated macrophages (TAMs) were the most important immune cell component in the tumor microenvironment. TAMs, often characterized by M2-like macrophages, play a variety of tumor-promoting effects in the tumor microenvironment. Studies have shown that there were different metabolic patterns between pro-inflammatory and anti-inflammatory macrophages[32, 33]. Activated pro-inflammatory macrophages often depend on glycolytic pathway for energy, while immunosuppressed macrophages are more inclined to use fatty acid oxidation to provide energy[34]. Research by Wu et al. found that fatty acids in the tumor microenvironment, especially unsaturated

fatty acids, might promote the polarization of monocytes to the M2-like macrophages, showing a strong immunosuppressive phenotype[35]. Therefore, the lipid metabolism risk signature somehow represents the alteration in the immune microenvironment of CRC. In addition, the risk signature was also associated with the immune checkpoint marker PD-L1, suggesting that the signature has the potential as metabolic markers for immunotherapy in CRC.

Some metabolic regulators have been considered as oncogenes or tumor suppressors. Carnitine palmitoyl transferase II(CPT2) is a rate-limiting enzymes for mitochondrial fatty acid transportation which plays a critical role in regulating fatty acid oxidation. Patients with lower CPT2 expression level had better disease control rate than patients with higher CPT2 in gastric cancer and CRC[36]. Blocking fatty acid oxidation by knocking out CPT1A/CPT2 via CRISPR-mediated would inhibit the invasive phenotype of radiotherapy-resistant breast cancer, suggesting that CPT2 was a potential metabolic target for breast cancer radiotherapy[37]. Besides, the CPT2 level in tumor tissue was associated with oxaliplatin-based chemotherapy sensitivity. Fatty acid catabolism can be inhibited by knocking out CPT2 or using the CPT2 inhibitor perhexiline. This inhibitory effect can promote oxaliplatin and other classic chemotherapy drugs to induce tumor cell apoptosis[36]. In this study, the patients had lower CPT2 levels in the high risk group, which

might explain why the overall survival of the low risk group was longer than that of the high risk group.

Conclusion

The study revealed that LMRGs had prognostic value in CRC and contribution to the formation of suppressive immune microenvironment. LMRGs, especially PROCA1, CCKBR, CPT2 and FDFT1, might be prognostic markers and potential therapeutic targets for CRC. This lipid metabolism-related prognostic signature may be a potential biomarker for predicting the efficacy of chemotherapy and anti-PD-L1 therapy in CRC.

Abbreviations

CRC: colorectal cancer; TCGA: the Cancer Genome Atlas; GEO: the Gene Expression Omnibus; LMRGs: lipid metabolism-related genes; PCA: the principal components analysis; LASSO: least absolute shrinkage and selection operator; FPKM: fragments per kilobase million; TPM: transcripts per million; ROC: receiver operating characteristic; AUC: area under the curves; TMB: tumor mutation burden; mRNAsi: the mRNA expression-based stemness index; DEGs: differentially expressed genes; OS: overall survival; FDR: false discovery rate; DEGs: differentially expressed genes; HR: hazard ratio; CI: confidence interval

Acknowledgements

None.

Authors' contributions

YBZ and CY designed the study; CY, SYH and FYC analyzed the data and drafted the manuscript. YBZ and CY carried out data management and revised the manuscript. All authors approved the final version of the manuscript.

Funding

The study was supported by the grants of the Wu Jieping Medical Foundation (320.2710.1855).

Availability of data and materials

All data can be obtained from TCGA CRC cohort (https://portal.gdc.cancer.gov/repository) and GEO GSE39583 cohort (https://www.ncbi.nlm.nih.gov/geo/).

Ethics approval and consent to participate

Both TCGA and GEO database are publicly available, thus ethical approval was not required for present study.

Consent for publication

Not Applicable.

Competing interests

All authors declare that there were no potential conflicts of interest.

Author details

[1]Department of Gastrointestinal Surgery, Renmin Hospital of Wuhan University, No. 99 ZhangZhiDong Street Wuchang District, Wuhan,

Hubei, P. R. China.

References:

1. Berndt N, Eckstein J, Heucke N, Gajowski R, Stockmann M, Meierhofer D, Holzhutter HG: Characterization of Lipid and Lipid Droplet Metabolism in Human HCC. CELLS-BASEL. 2019, 8(5)-11-17.
2. Lodhi IJ, Semenkovich CF: Peroxisomes: a nexus for lipid metabolism and cellular signaling. CELL METAB. 2014, 19(3):380-392.
3. Thakur C, Chen F: Connections between metabolism and epigenetics in cancers. SEMIN CANCER BIOL. 2019, 57:52-58.
4. Luo X, Cheng C, Tan Z, Li N, Tang M, Yang L, Cao Y: Emerging roles of lipid metabolism in cancer metastasis. MOL CANCER. 2017, 16(1):76.
5. Huang C, Freter C: Lipid metabolism, apoptosis and cancer therapy. INT J MOL SCI. 2015, 16(1):924-949.
6. Beloribi-Djefaflia S, Vasseur S, Guillaumond F: Lipid metabolic reprogramming in cancer cells. ONCOGENESIS. 2016, 5:e189.
7. Lazar I, Clement E, Dauvillier S, Milhas D, Ducoux-Petit M, LeGonidec S, Moro C, Soldan V, Dalle S, Balor S et al: Adipocyte Exosomes Promote Melanoma Aggressiveness through Fatty Acid Oxidation: A Novel Mechanism Linking Obesity and Cancer. CANCER RES. 2016, 76(14):4051-4057.
8. Iwamoto H, Abe M, Yang Y, Cui D, Seki T, Nakamura M, Hosaka K, Lim S, Wu J, He X et al: Cancer Lipid Metabolism Confers Antiangiogenic Drug Resistance. CELL METAB. 2018, 28(1):104-117.
9. Merino SM, Gomez DCM, Moreno RJ, Falagan MS, Sanchez MR, Casado E, Ramirez DMA, Sereno M: Lipid metabolism and lung cancer. Crit Rev Oncol Hematol. 2017, 112:31-40.
10. Rodriguez-Broadbent H, Law PJ, Sud A, Palin K, Tuupanen S, Gylfe A, Hanninen UA, Cajuso T, Tanskanen T, Kondelin J et al: Mendelian randomisation implicates hyperlipidaemia as a risk factor for colorectal cancer. INT J CANCER. 2017,

140(12):2701-2708.

11.Liu T, Peng F, Yu J, Tan Z, Rao T, Chen Y, Wang Y, Liu Z, Zhou H, Peng J: LC-MS-based lipid profile in colorectal cancer patients: TAGs are the main disturbed lipid markers of colorectal cancer progression. ANAL BIOANAL CHEM. 2019, 411(20):5079-5088.

12.Wang Y, Hinz S, Uckermann O, Honscheid P, von Schonfels W, Burmeister G, Hendricks A, Ackerman JM, Baretton GB, Hampe J et al: Shotgun lipidomics-based characterization of the landscape of lipid metabolism in colorectal cancer. Biochim Biophys Acta Mol Cell Biol Lipids. 2020, 1865(3):158579.

13.Zhu Y, Gu L, Lin X, Liu C, Lu B, Cui K, Zhou F, Zhao Q, Prochownik EV, Fan C et al: Dynamic Regulation of ME1 Phosphorylation and Acetylation Affects Lipid Metabolism and Colorectal Tumorigenesis. MOL CELL. 2020, 77(1):138-149.

14.Gong J, Lin Y, Zhang H, Liu C, Cheng Z, Yang X, Zhang J, Xiao Y, Sang N, Qian X et al: Reprogramming of lipid metabolism in cancer-associated fibroblasts potentiates migration of colorectal cancer cells. CELL DEATH DIS. 2020, 11(4):1145-1154.

15.Cao Y: Adipocyte and lipid metabolism in cancer drug resistance. J CLIN INVEST. 2019, 129(8):3006-3017.

16.Ma B, Jiang H, Wen D, Hu J, Han L, Liu W, Xu W, Shi X, Wei W, Liao T et al: Transcriptome Analyses Identify a Metabolic Gene Signature Indicative of Dedifferentiation of Papillary Thyroid Cancer. J Clin Endocrinol Metab. 2019, 104(9):3713-3725.

17.Wu F, Zhao Z, Chai RC, Liu YQ, Li GZ, Jiang HY, Jiang T: Prognostic power of a lipid metabolism gene panel for diffuse gliomas. J CELL MOL MED. 2019, 23(11):7741-7748.

18.Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES et al: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005, 102(43):15545-15550.

19. Tirinato L, Liberale C, Di Franco S, Candeloro P, Benfante A, La Rocca

R,PotzeL, Marotta R, Ruffilli R, Rajamanickam VP et al: Lipid droplets: a new player in colorectal cancer stem cells unveiled by spectroscopic imaging. STEM CELLS. 2015, 33(1):35-44.

20.Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kaminska B, Huelsken J, Omberg L, Gevaert O et al: Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. CELL. 2018, 173(2):338-354.

21.Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA: Robust enumeration of cell subsets from tissue expression profiles. NAT METHODS. 2015, 12(5):453-457.

22.Llosa NJ, Luber B, Siegel N, Awan AH, Oke T, Zhu Q, Bartlett BR, Aulakh LK, Thompson ED, Jaffee EM et al: Immunopathologic Stratification of Colorectal Cancer for Checkpoint Blockade Immunotherapy. CANCER IMMUNOL RES. 2019, 7(10):1574-1579.

23.Lee DW, Han SW, Bae JM, Jang H, Han H, Kim H, Bang D, Jeong SY, Park KJ, Kang GH et al: Tumor Mutation Burden and Prognosis in Patients with Colorectal Cancer Treated with Adjuvant Fluoropyrimidine and Oxaliplatin. CLIN CANCER RES. 2019, 25(20):6141-6147.

24.Klempner SJ, Fabrizio D, Bane S, Reinhart M, Peoples T, Ali SM, Sokol ES, Frampton G, Schrock AB, Anhorn R et al: Tumor Mutational Burden as a Predictive Biomarker for Response to Immune Checkpoint Inhibitors: A Review of Current Evidence. ONCOLOGIST. 2020, 25(1):e147-e159.

25.Arora S, Velichinskii R, Lesh RW, Ali U, Kubiak M, Bansal P, Borghaei H, Edelman MJ, Boumber Y: Existing and Emerging Biomarkers for Immune Checkpoint Immunotherapy in Solid Tumors. ADV THER. 2019, 36(10):2638-2678.

26.Flegal KM, Ogden CL, Fryar C, Afful J, Klein R, Huang DT: Comparisons of Self-Reported and Measured Height and Weight, BMI, and Obesity Prevalence from National Surveys: 1999-2016. Obesity (Silver Spring). 2019, 27(10):1711-1719.

27.Ganesh K, Stadler ZK, Cercek A, Mendelsohn RB, Shia J, Segal NH, Diaz LJ: Immunotherapy in colorectal cancer: rationale, challenges and potential. Nat Rev Gastroenterol Hepatol .2019, 16(6):361-375.

28.Dekker E, Tanis PJ, Vleugels J, Kasi PM, Wallace MB: Colorectal cancer. LANCET. 2019, 394(10207):1467-1480.

29.La Vecchia S, Sebastian C: Metabolic pathways regulating colorectal cancer initiation and progression. SEMIN CELL DEV BIOL .2020, 98(SI):63-70.

30.Santos CR, Schulze A: Lipid metabolism in cancer. FEBS J .2012, 279(15):2610-2623.

31.Munir R, Lisec J, Swinnen JV, Zaidi N: Lipid metabolism in cancer cells under metabolic stress. Br J Cancer .2019, 120(12):1090-1098.

32.Phan AT, Goldrath AW, Glass CK: Metabolic and Epigenetic Coordination of T Cell and Macrophage Immunity. IMMUNITY .2017, 46(5):714-729.

33.Stienstra R, Netea-Maier RT, Riksen NP, Joosten L, Netea MG: Specific and Complex Reprogramming of Cellular Metabolism in Myeloid Cells during Innate Immune Responses. CELL METAB .2017, 26(1):142-156.

34.Huang SC, Everts B, Ivanova Y, O'Sullivan D, Nascimento M, Smith AM, Beatty W, Love-Gregory L, Lam WY, O'Neill CM et al: Cell-intrinsic lysosomal lipolysis is essential for alternative activation of macrophages. NAT IMMUNOL .2014, 15(9):846-855.

35.Wu H, Han Y, Rodriguez SY, Deng H, Siddiqui S, Treese C, Schmidt F, Friedrich M, Keye J, Wan J et al: Lipid droplet-dependent fatty acid metabolism controls the immune suppressive phenotype of tumor-associated macrophages. EMBO MOL MED .2019, 11(11):e10698.

36.Wang Y, Lu JH, Wang F, Wang YN, He MM, Wu QN, Lu YX, Yu HE, Chen ZH, Zhao Q et al: Inhibition of fatty acid catabolism augments the efficacy of oxaliplatin-based chemotherapy in gastrointestinal cancers. CANCER LETT .2020, 473:74-89.

37.Han S, Wei R, Zhang X, Jiang N, Fan M, Huang JH, Xie B, Zhang L, Miao W, Butler AC et al: CPT1A/2-Mediated FAO Enhancement-A Metabolic Target in

Radioresistant Breast Cancer. FRONT ONCOL. 2019, 9(1201)-112-118.

Figures

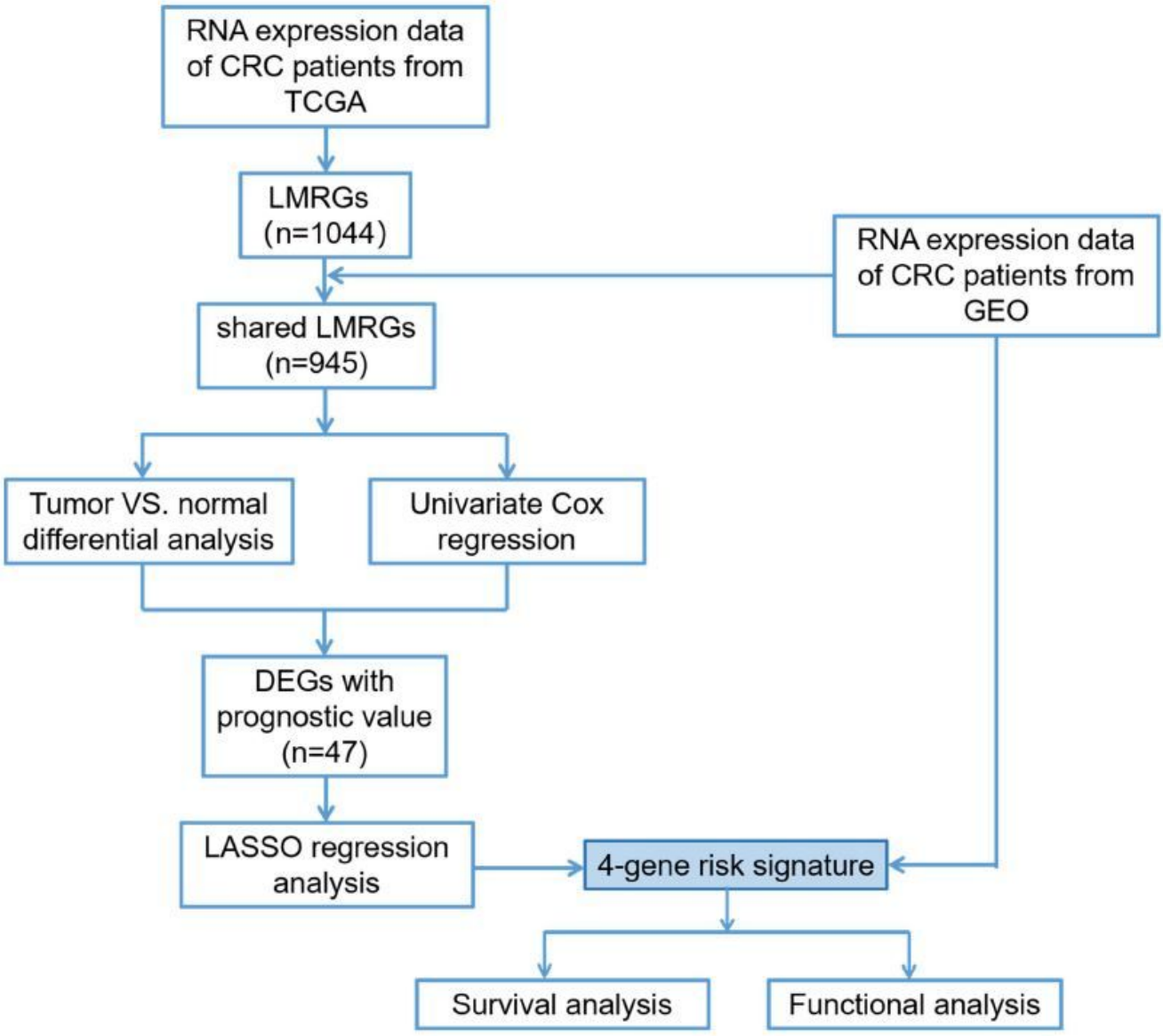

Figure 1

Flowchart of study design. (LMRGs: lipid metabolism-related genes; DEGs: differentially expressed genes; LASSO: the Least Absolute Shrinkage and Selection Operator; TCGA: the Cancer Genome Atlas; GEO: the Gene Expression Omnibus)

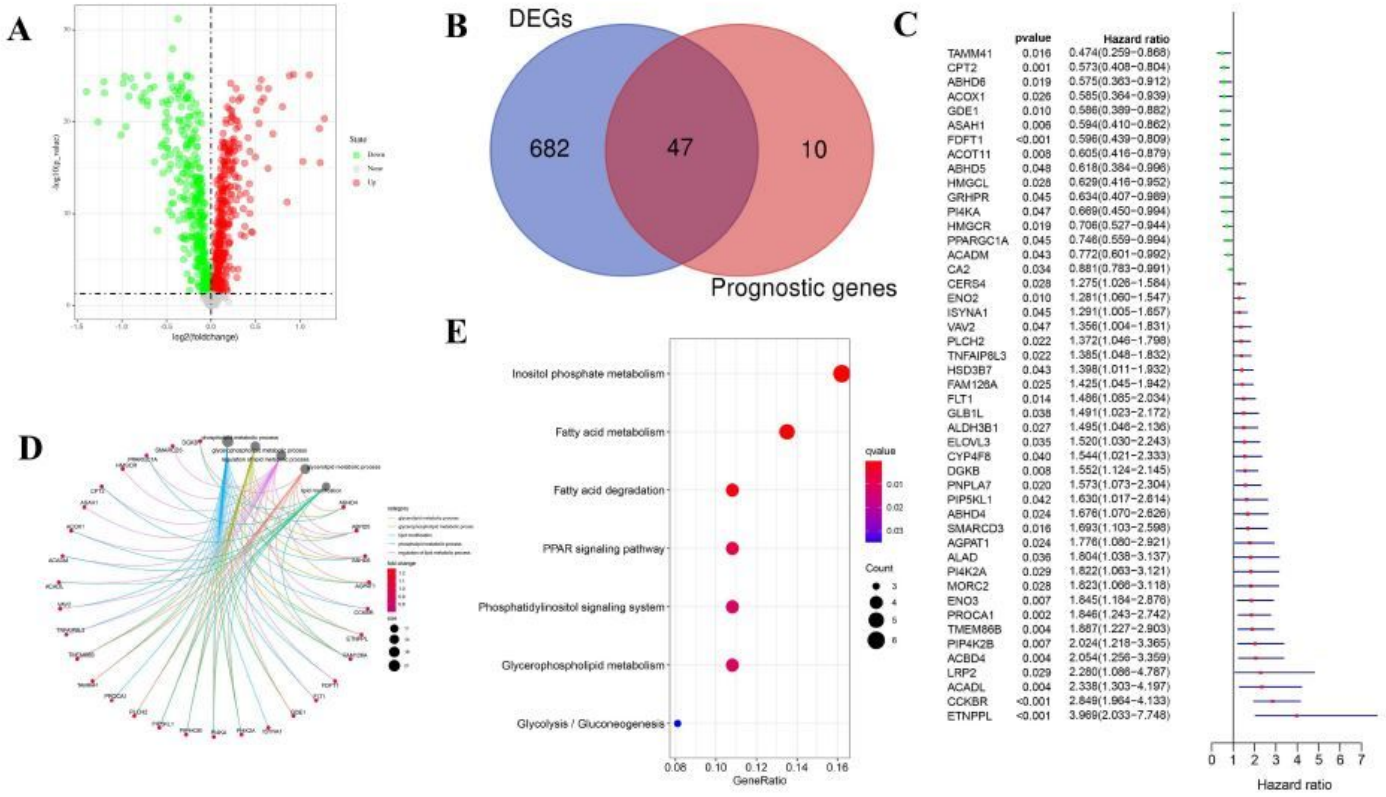

Figure 2

Identification of the differentially expressed and prognosis-related genes in the TCGA cohort. (A) Volcano map showed the differentially expressed genes between normal and tumor tissues. Red meant up-regulated genes and green represented down-regulated genes. (B) Venn diagram showed the 47 differentially expressed genes which also had prognostic value. (C) Forest plot for the univariate Cox regression analysis of the 47 overlapping genes. (D) The main GO enrichment terms of the 47 overlapping genes. (E) The most significant KEGG pathways of the 47 overlapping genes. (GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes)

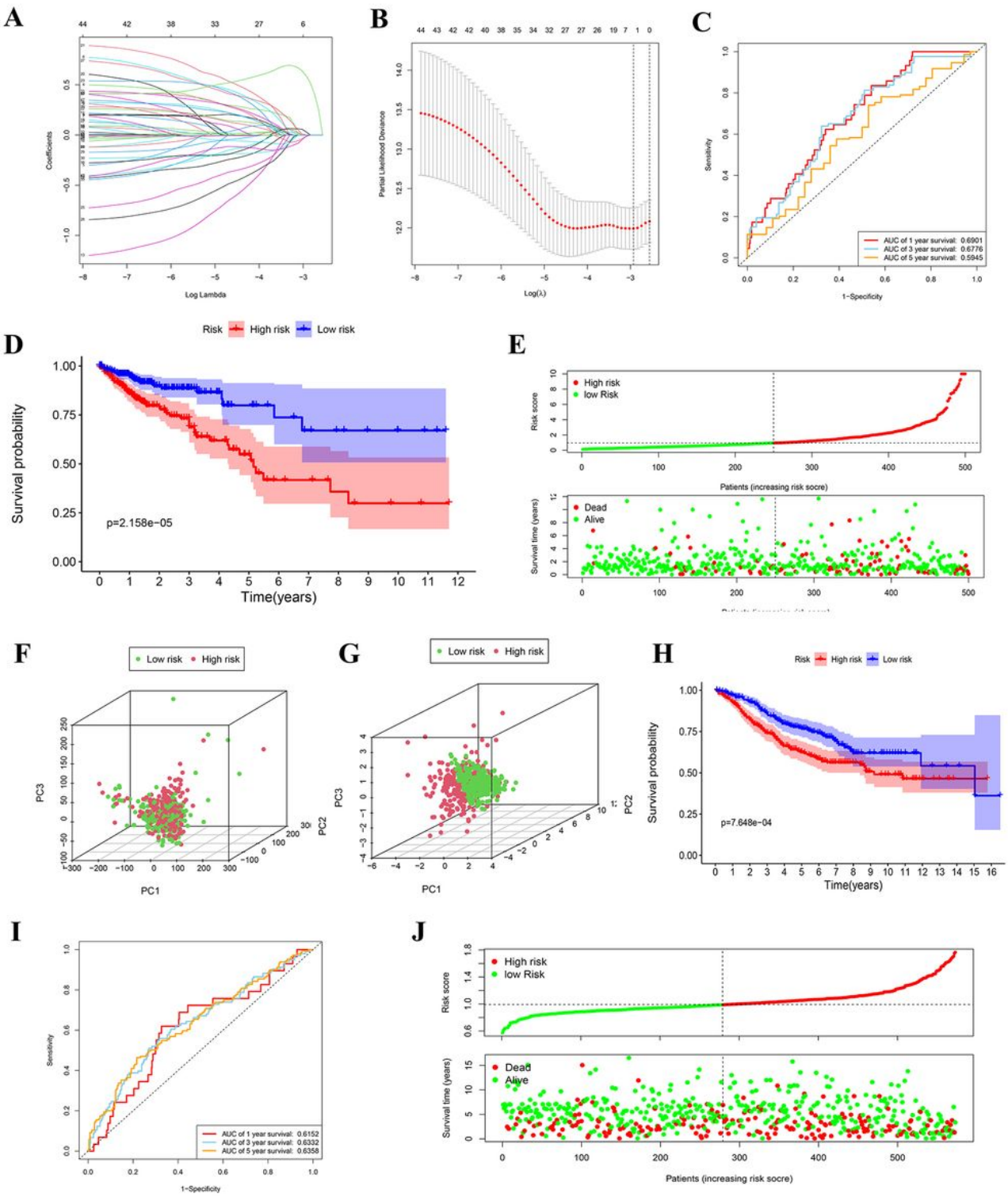

Figure 3

Development and validation of the risk signature. (A) The LASSO COX regression coefficient profiles of the 47 LMRGs. (B) 10-time cross-validation for tuning parameter selection in the LASSO mode. (C) ROC at 1, 3 and 5 years (AUCs = 0.6901, 0.6776 and 0.5945). (D) The survival curve in the high and low risk groups showed significant differences (P< 0.001). (E) Risk score distribution(above) and survival status(bottom) for patients in different risk groups. (F) The PCA showed that the distribution in different

directions were general on the basis of the whole gene sets. (G) The PCA showed that the risk signature could distinguish patients in different risk groups markedly. (H) The survival curves showed significant differences in validation sets (P< 0.001). (I) ROC at 1, 3 and 5 years in validation set (AUCs = 0.6152, 0.6332 and 0.6358). (E) Risk score distribution(above) and survival status(bottom) for patients in different risk groups of the validation set. (LASSO: Least Absolute Shrinkage and Selection Operator; the PCA: the principal components analysis; AUCs: the area under curves; ROC: receiver operating characteristic)

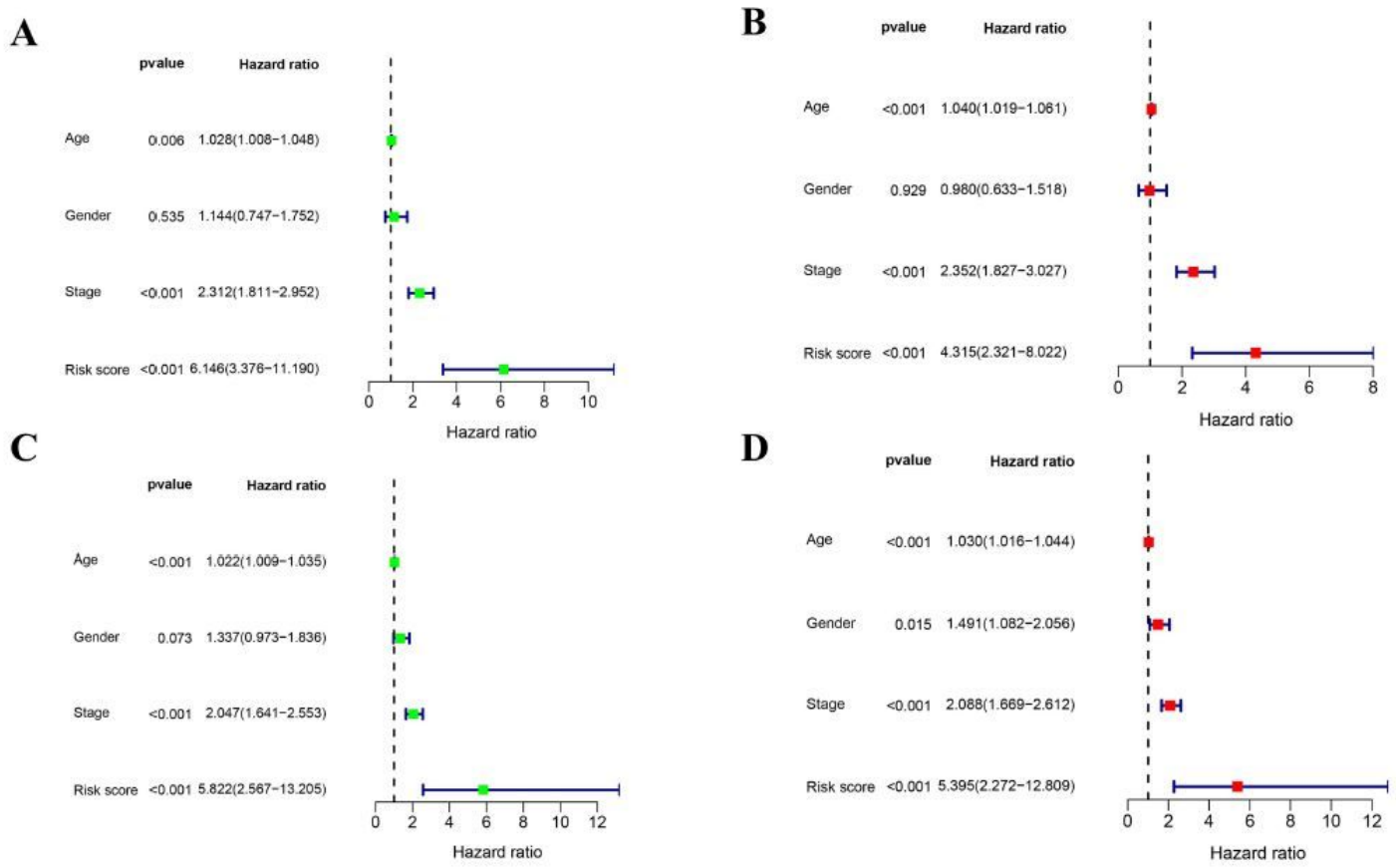

Figure 4

Univariate and multivariate Cox regression analysis showed risk score was an independent prognostic factor both in training and validation set. (A)The result of univariate Cox regression analysis in training set. (B) The result of multivariate Cox regression analysis in training set. (C) The result of univariate Cox regression analysis in validation set. (D) The result of multivariate Cox regression analysis in validation set.

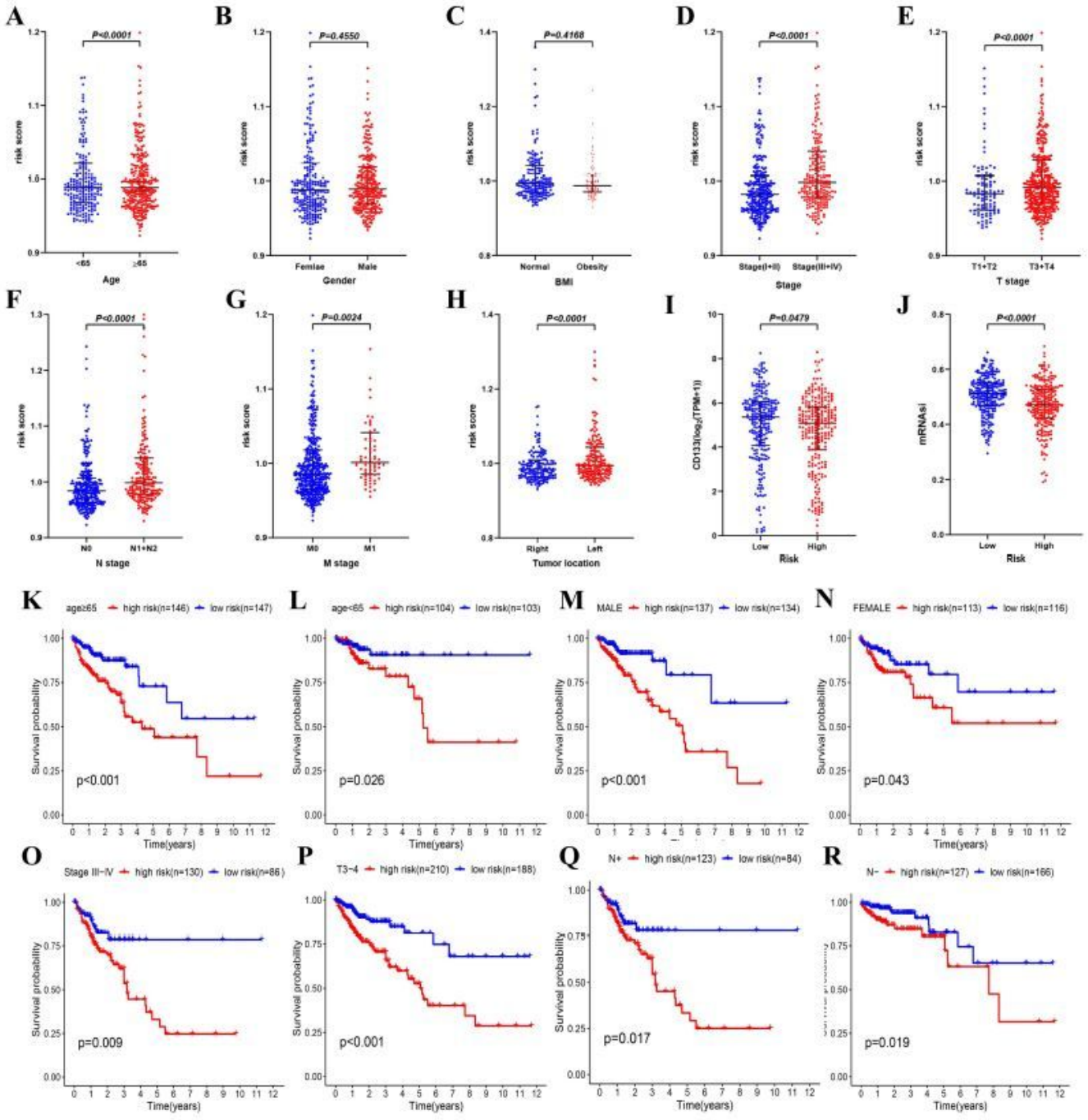

Figure 5

Stratified analysis of the risk signature in the training set. The relationships between the risk signature and age (A), gender (B), BMI (C), tumor stage (D), T stage (E), N stage (F), M stage (G), tumor location (H), CD133 mRNA (I) and mRNAsi (J). A poor prognosis was observed in high risk group in majority of clinical stratification, including age (K-L), gender (M-N), tumor stage (O), T stage (P) and N stage (Q-R). (BMI: Body Mass Index; mRNAsi: the mRNA expression-based stemness index)

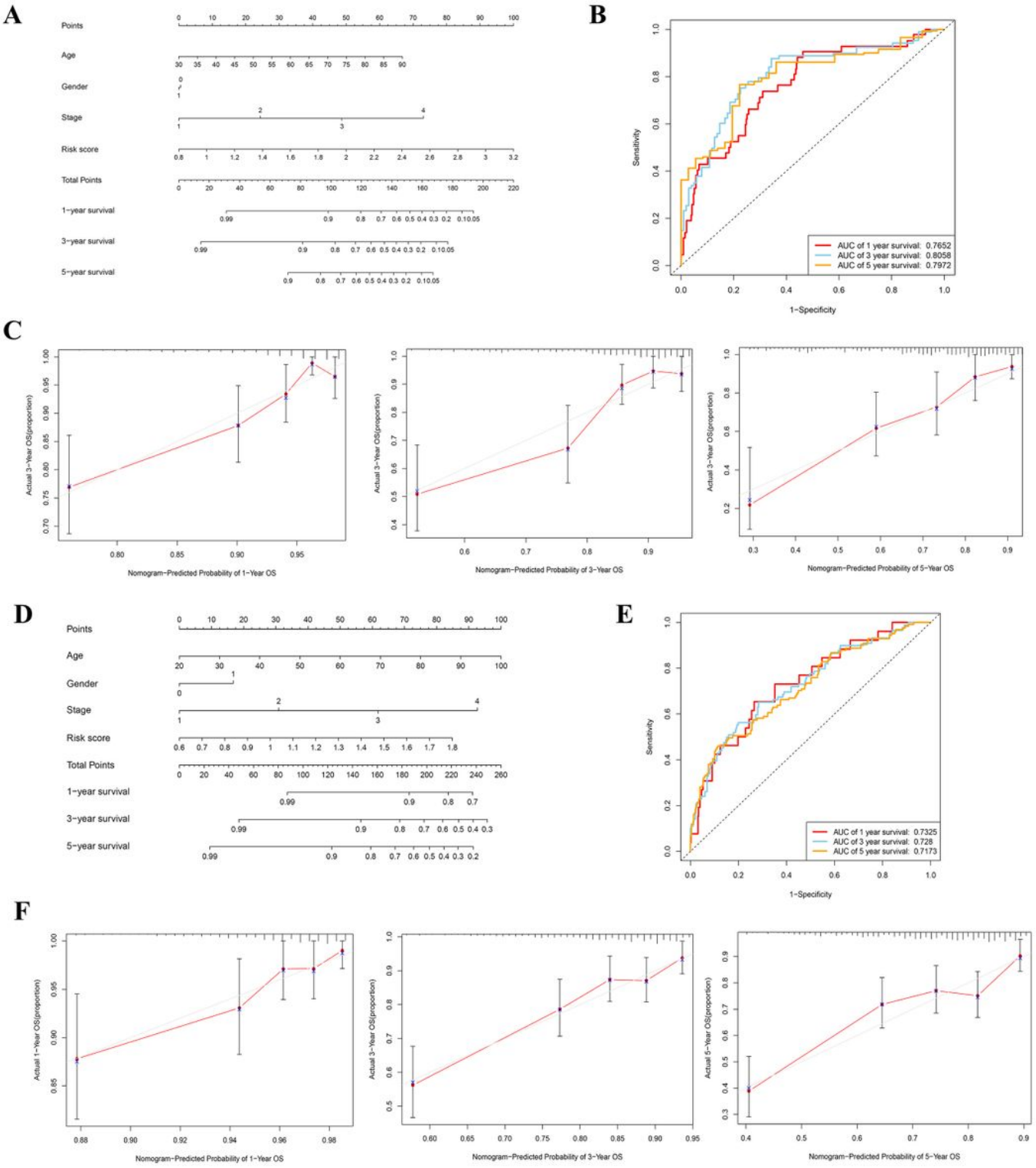

Figure 6

Construction and validation of the nomogram combining clinicopathological features and risk signature. (A) The nomogram in training set. (B) The prognostic value of nomogram for predicting 1-, 3-, and 5-year overall survival rate in training set. (C) Calibration plots of the nomogram in training set. (D) The nomogram in validation set. (E) The prognostic value of nomogram for predicting 1-, 3-, and 5-year overall survival rate in validation set. (F) Calibration plots of the nomogram in validation set.

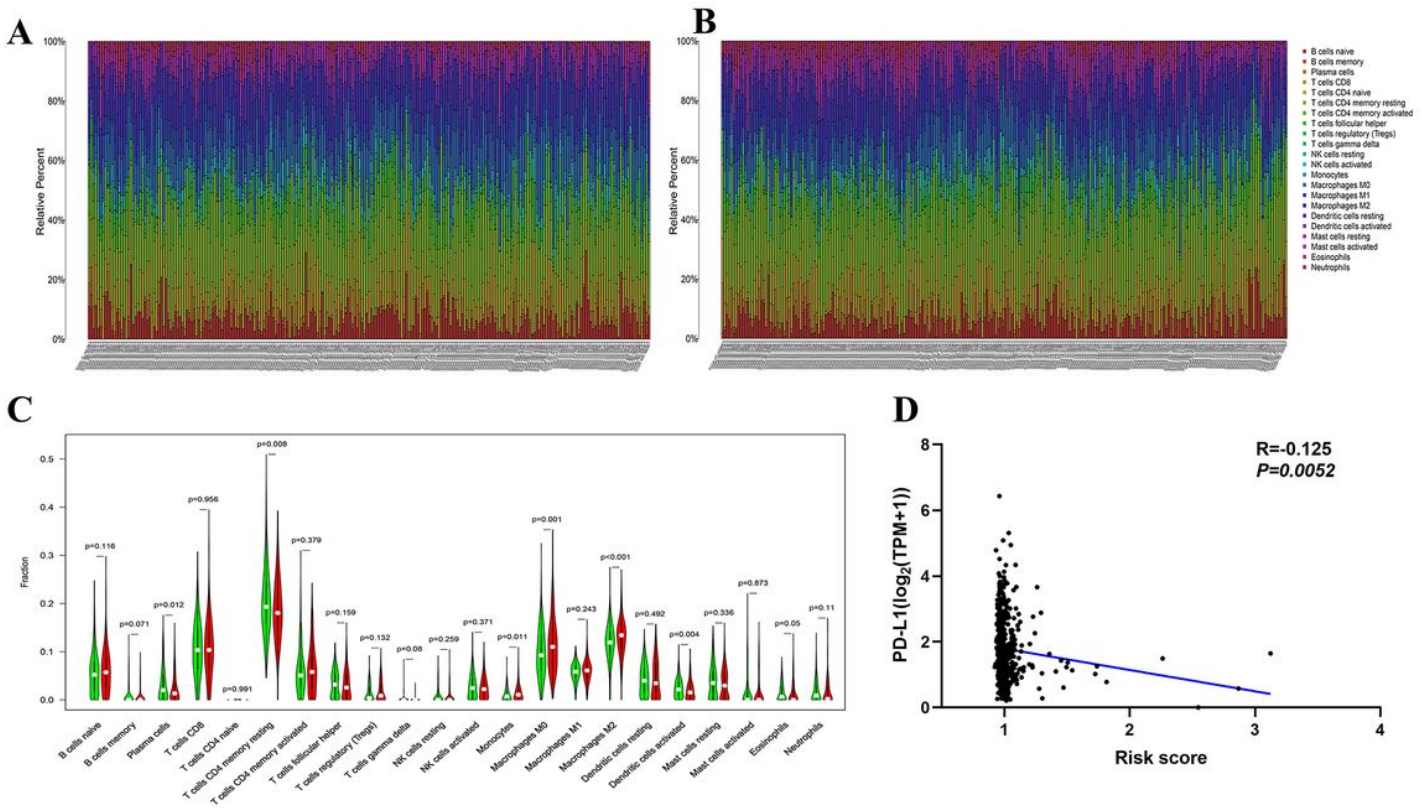

Figure 7

The risk signature was correlated with tumor-infiltrating immune cells. (A) The distribution pattern of immune cells in high risk group. (B) The distribution pattern of immune cells in low risk group. (C) There were significant differences in the abundance of many immune cells in two groups, especially monocytes (P=0.011), M0 (P=0.001) and M2-like macrophages (P<0.001). (Green represented the low risk group, and red meant the high risk group.) (D) Risk score was negatively correlated with PD-L1 mRNA (R=-0.125, P=0.0052).

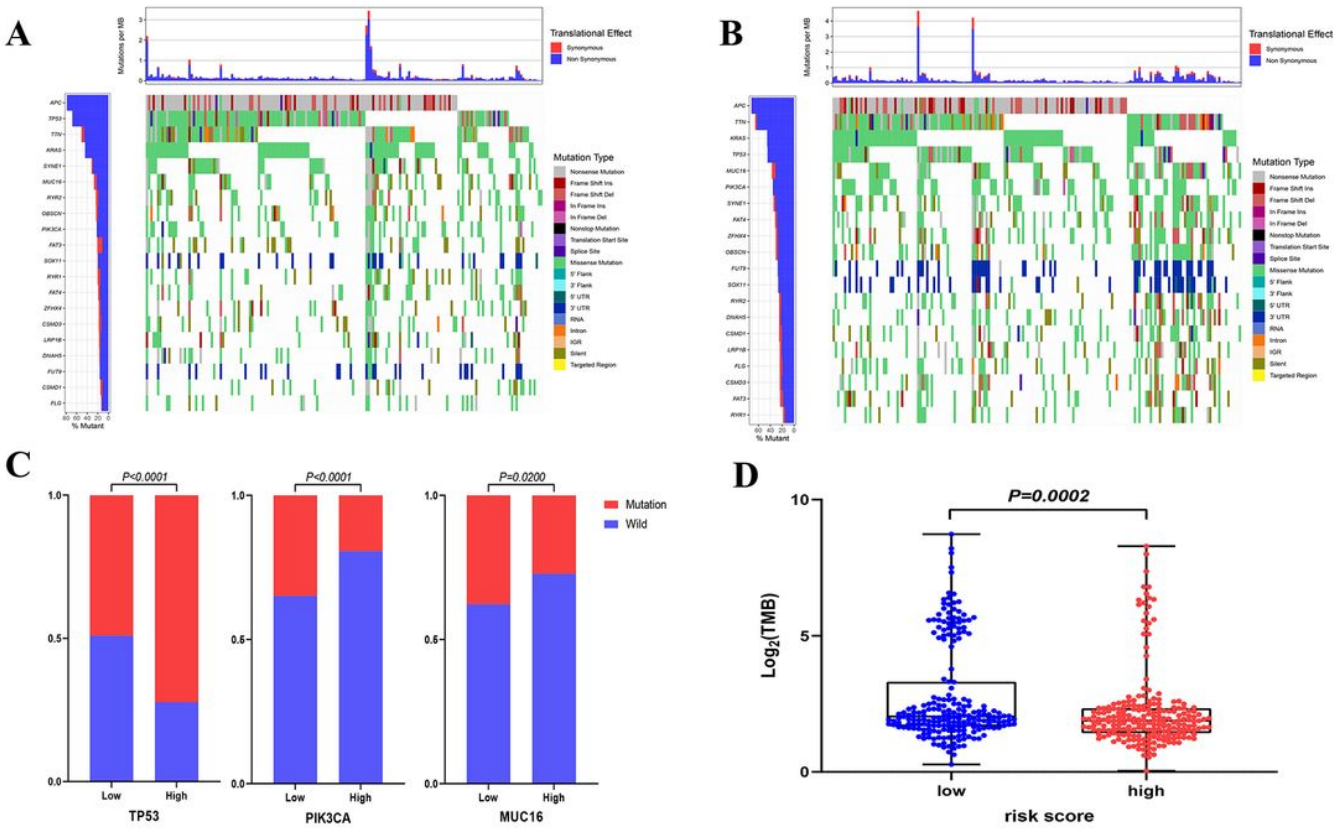

Figure 8

Mutation landscape between the low and high risk. (A) The top 20 genes with the highest frequency of mutation in high group were shown by waterfall plot. (B) The top 20 genes with the highest frequency of mutations in low group were shown by waterfall plot. (C) The key genes, including TP53, PI3K and MUC16, had significant differences in mutation frequency between high and low risk groups. (D) The low risk group had a higher TMB level(P=0.0002). (TMB: tumor mutational burden).

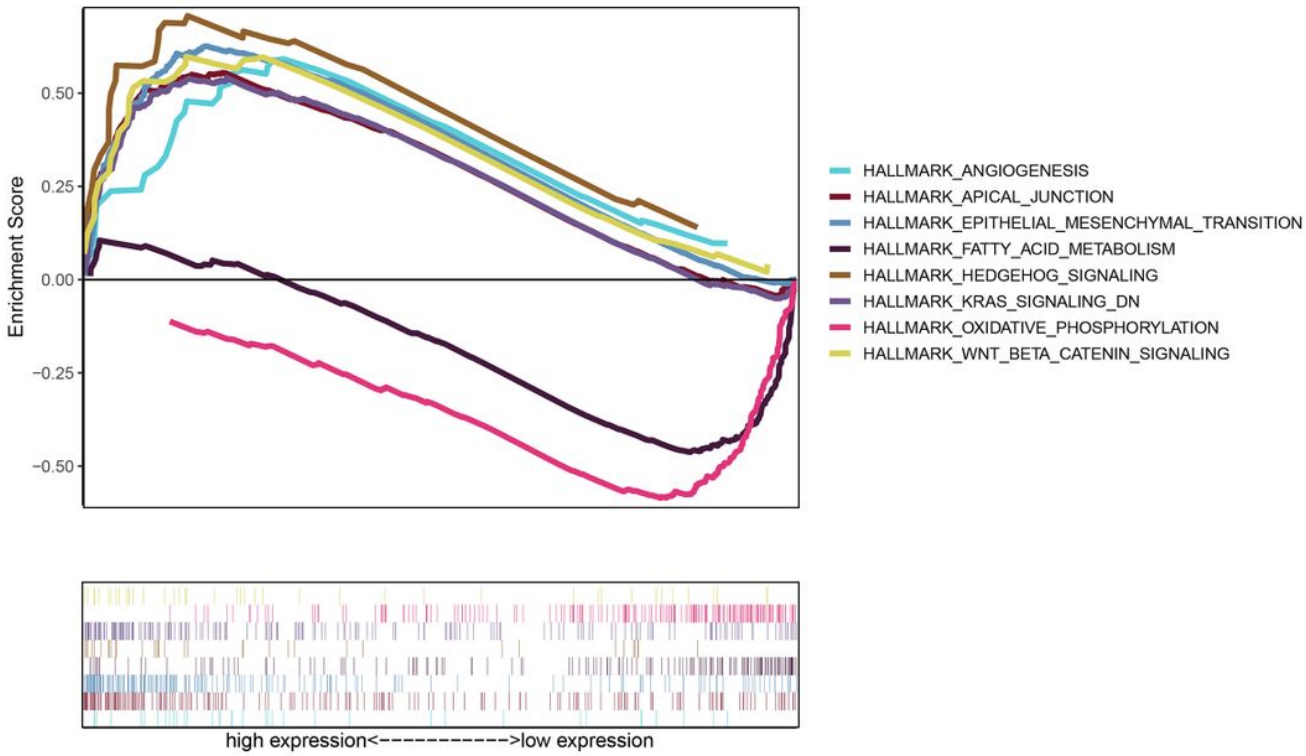

Figure 9

Distinct biological function pathways characterize in two groups. The results showed that the high risk group was associated with Hedgehog signaling, KRAS signaling, Wnt/β catenin signaling, apical junction, epithelial mesenchymal transition and angiogenesis, while functional enrichment in the low risk group focused on energy-metabolism-related functions, including fatty acid metabolism and oxidative phosphorylation.

Supplementary Files

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

- SupplementFigure.doc
- SupplementTables.doc