Identification and validation of a two-gene metabolic signature for survival prediction in patients with kidney renal clear cell carcinoma

Xudong Guo
Department of Urology, Shandong Provincial Hospital, Cheeloo College of Medicine, Shandong University, Jinan, Shandong, 250021, China. Department of Urology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, 250

Zhuolun Sun
Department of Urology, Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong, 510630, China

Shaobo Jiang
Department of Urology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, 250021, China

Xunbo Jin
Department of Urology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, 250021, China

Hanbo Wang (TigerS323@163.com)
Shandong Provincial Hospital

Research

Keywords: kidney renal clear cell carcinoma, metabolism, nomogram, prognostic signature, the Cancer Genome Atlas.

Posted Date: August 4th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-51839/v1

Version of Record: A version of this preprint was published at Aging on March 3rd, 2021. See the published version at https://doi.org/10.18632/aging.202636.

Abstract

Background: Kidney renal clear cell carcinoma (KIRC) is one of the most common malignant tumors worldwide. Deregulated tumor cell metabolism is emerging as a common feature of tumorigenesis. The expression pattern and clinical significance of metabolism-related genes (MRGs) in KIRC remains unclear.

Methods: We downloaded the RNA sequencing data and corresponding clinical information for KIRC from the Cancer Genome Atlas (TCGA) database and identified the differently expressed MRGs between tumors and normal tissues. According to the Cox regression analysis and least absolute shrinkage and selection operator (LASSO), we identified target genes for prognostic signature construction. We also analyzed the correlations of the signature risk score with clinicopathological features. The robustness of the signature was further examined by stratified survival analysis. A predictive nomogram was built for the optimal strategy to predict the survival possibility of KIRC patients. The expression levels of target genes were validated in multiple datasets. Gene set enrichment analyses (GSEA) were performed to unveil several significantly enriched pathways.

Results: A total of 123 differentially expressed MRGs were identified, including 60 up-regulated genes and 63 down-regulated genes. Next, RRM2 and ALDH6A1 were identified as prognosis-related genes and used to construct a prognostic signature. The signature was proved to be an independent prognostic factor for KIRC survival by multivariable Cox regression analysis. Subgroup analysis indicated that this signature could serve as a classifier for the evaluation of low- and high-risk groups. Up regulation of RRM2 and down regulation of ALDH6A1 were associated with unfavorable prognosis in patients suffering from KIRC. Nomogram including the signature suggested some clinical net benefit for overall survival prediction. In addition, the calibration curves indicated the nomogram performed well in predicting 3- and 5-year OS compared with the ideal model. The expression level of RRM2 was significantly up-expressed, while ALDH6A1 were significantly down-expressed in KIRC samples compared with the normal samples in multiple datasets. Furthermore, RRM2 and ALDH6A1were significantly enriched in different pathways.

Conclusion: Our study identified a two-gene metabolic signature that had important clinical implications in KIRC prognosis prediction, which might provide potential biomarkers and targets of metabolic therapeutic relevance.

Background

Renal cell carcinoma (RCC) accounts for 2% to 3% of all adult malignancies, among which, kidney renal clear cell carcinoma (KIRC) is the most common histological subtype and comprises 80% to 90% of the total cases of RCC [1, 2]. The incidence of RCC has risen steadily within the last decades, moreover, RCC exhibits the highest mortality rate among all urologic malignancies and causes 100,000 deaths worldwide annually [3]. Given the clinical manifestations are diverse and lack specificity, up to 30% of patients with KIRC are typically not diagnosed until the advanced stage [4]. Patients have few treatment options other than surgery due to RCC is frequently resistant to chemotherapy and radiotherapy [5].

Despite the quality of medical care is gradual improving, the prognosis of metastatic RCC patients is poor with a 5-year overall survival (OS) rate of less than 10%. Nonetheless, it should be noted that the 5-year OS rate of patients with early-stage KIRC may be as high as 90% following surgery [6]. Therefore, identifying an early diagnostic biomarker to improve early diagnosis and prolong the survival time of patients with KIRC is crucial.

The metabolic reprogramming of tumor cells has been a focus in recent studies on cancers. Metabolic alteration is considered to be an adaptation mechanism to support increasing energy demands of rapidly growing and proliferating cancer cells [7]. Some metabolic changes, such as the alterations of glycolysis, appeared to be necessary for the malignant transformation [8]. In addition, recent studies have revealed that significant glutamine and lipid metabolism dysfunctions were associated closely with tumor progression and carcinogenesis [9, 10]. Deregulated tumor cell metabolism is emerging as a novel hallmark of cancer based on these basic findings. Targeting dysregulated metabolism has shown significant benefit against solid tumors in both preclinical and clinical studies [11, 12]. In recent years, accumulating investigators have applied integrated transcriptomics and metabolomics to explore underlying molecular mechanisms and potential metabolic biomarkers for cancer. For example, Ma et al. uncovered the hepatocellular carcinoma metabolism characteristics and four significantly differential genes that can be established as promising biomarkers via the integration of metabolomics and transcriptomics [13]. More importantly, comprehensive molecular characterization of KIRC highlighted the critical role of metabolic alteration in kidney cancer progression [14]. Thus, identifying the effective metabolism-related biomarkers may not only be of great significance to improve the early diagnosis, treatment and prognosis, but also shed new light on the candidate therapeutic target in KIRC.

In the current study, we explored the prognostic significance of metabolism-related genes (MRGs) in KIRC patients through the integration of metabolites and transcriptomics data obtained from The Cancer Genome Atlas (TCGA) database. We discovered that several MRGs play important roles in the progression of KIRC, and then constructed a prognostic signature composed of 2 MRGs, which was probed to be able to independently and accurately predict the prognostic of patients. Moreover, a novel promising prognostic nomogram was established as a quantitative prediction tool to provide patients with best-individualized treatment strategies by integrating the prognostic signature and clinical variables. In summary, our results suggested that this metabolism-related prognostic signature may have important clinical implications and provide new biomarker and therapeutic target for early diagnosis and treatment of KIRC patients.

Methods

Study cohort

RNA sequencing transcriptome data and clinical information were downloaded for 539 KIPC and 72 normal tissues from TCGA (The Cancer Genome Atlas, https://cancergenome.nih.gov/). The extracted data were subjected to be normalized and processed by *log2* transformation. We deleted the data on all

KIRC samples with incomplete data and survival times of less than 30 days. The metabolism-related gene (MRG) sets were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways which were provided by The Molecular Signatures Database (MSigDB, https://www.gsea-msigdb.org/gsea/msigdb/) [15]. There was a total of 944 genes ready for further analysis after removing duplicates. Informed consent and ethical approval were not needed due to the TCGA database was an open public database.

Identification of differentially expressed MRGs and enrichment analysis

The differentially expressed MRGs between cancer samples and normal control samples was identified by LIMMA package in R language [16] with the following thresholds: |log2 fold change (FC)| > 2.0 and a false discovery rate (FDR) corrected p-value < 0.05 (calculated by Benjamini & Hochberg procedure). A series of gene functional enrichment analyses were performed to explore potential molecular mechanisms of the differentially expressed MRGs using gene ontology (GO) and KEGG. The ggplot2 and enrichplot R packages were used to supply the visual enrichment maps of annotation analysis results to help interpretation. A value of P < 0.05 was regarded as statistically significant.

Construction and evaluation of prognostic signature

The relationship between the expression levels of the differentially expressed MRGs and OS were first examined via univariate Cox regression analysis. Those genes with significant prognostic value ($P < 0.05$) were included for subsequent research. To obtain the most optimal MRGs and to control the complexity of the model, we carried out the least absolute shrinkage and selection operator (LASSO) Cox regression analysis with the glmnet package. Lasso Cox regression is a robust model building method that reduces the dimension and eliminates the over-fitting phenomenon[17]. After that, several target genes were obtained for the prognostic signature development. The calculation of the risk score for each patient based on the signature using the following formula: Risk Score= , where represents the number of modules genes, denotes the estimated regression coefficient by LASSO analysis and indicates the relative expression level of each MRG.

The median risk score was taken as the cut-off point to separate all KIRC patients into high-risk group and low-risk group. The Kaplan–Meier method and the log-rank test were used to compare the difference of OS between high- and low-risk groups. The time-dependent receiver operating characteristic (ROC) curve analysis was conducted to assess the predictive accuracy of the prognostic signature and the value of area under the ROC curve (AUC) ranged from 0.5 (no predictability) to 1 (perfect predictability).

Furthermore, whether the metabolism-related signature could be an independent predictor of OS for the TCGA KIRC cohort was determined by univariate and multivariate Cox regression analyses. The age, gender, grade, American Joint Committee on Cancer (AJCC) stages, T stage, M stage were used as covariates. N stage was not analyzed for a large amount of missing data.

In addition, to detect the prognostic value of risk score in different subgroups, stratified survival analysis was carried out according to clinical characteristics related to the prognosis including age (≤ 65 and > 65), gender (female and male), grade (G1-2 and G3-4), AJCC stage (Ⅰ/Ⅱ and Ⅲ/Ⅳ), T stage (T1-2 and T3-4) and M (M0 and M1) stage. The relationship between the expression level of each target gene and clinical parameters was also compared to further understand the impact of the individual target gene in our prognostic signature on KIRC patients.

Development of predictive nomogram

A nomogram was developed as a quantitative prediction tool to evaluate the probability of 1-, 3-, and 5-year OS for TCGA KIRC patients, incorporating several clinical variables (age, gender, grade, AJCC stage, T stage and M stage) and risk score calculated from the prognostic signature. Subsequently, the concordance index (C-index) and the calibration plots were computed to evaluate the discrimination and predictive ability of the nomogram. The C-index values ranged from 0.5 to 1.0, with a C-index value of 1.0 being perfect concordance and 0.5 being the opposite. A reference line with a slope of one in the calibration plots represents perfect calibration. Decision curve analysis (DCA) was performed to evaluate the clinical utility of the signature-based nomogram model by quantifying the net benefits under different threshold values.

External verification of the prognostic genes in the prognostic signature

To further evaluate the reliability of the results mentioned above, we investigated The expression levels of the RRM2 and ALDH6A1 between tumor and normal tissues using Oncomine (https://www.oncomine.org/resource/main.html) database and The Tumor Immune Estimation Resource (TIMER, https://cistrome.shinyapps.io/timer/) database. Moreover, Gene Expression Profiling Interactive Analysis (GEPIA, http://gepia.cancer-pku.cn/) online tool was explored to verify the prognostic value of those target genes through the survival analysis. The expression of the prognostic genes in the gene signature was further validated at the protein level using The Human Protein Atlas (HPA, https://www.proteinatlas.org/) database. Then, the expression levels of these target genes between tumor tissues and normal tissues were verified in Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/).

Gene set enrichment analysis (GSEA)

GSEA was performed using GSEA software (http://www.broadinstitute.org/gsea). This permutation test was conducted 1000 times for each task to identify the significantly changed pathways. The minimum and maximum criteria for selection of gene sets from the collection were 15 and 500 genes, respectively. The significant related genes were defined with A nominal P value < 0.05 and FDR value < 0.25.

Statistical analyses

All statistical analyses were performed with R 3.6.2 (https://www.r-proje ct.org/). All analyses were two-sided and statistical significance was defined as $P < 0.05$.

Results

Differentially expressed MRGs and functional enrichment analysis

We first compared the expression values of 944 MRGs between KIRP samples and normal control samples using the Wilcoxon signed-rank test in R. A total of 123 differentially expressed MRGs, including 60 up-regulated genes and 63 down-regulated genes, were eventually identified based on the criteria of |log2FC| > 2 and FDR < 0.05. Then, the volcano plot and heatmap were utilized to visualize the expression patterns of these differentially expressed MRGs in tumor samples and normal samples (Figure 1A and 1B).

Functional enrichment analysis was performed to explore potential molecular mechanisms of the differentially expressed MRGs. The most enriched GO terms in the biological processes (BP) category were small molecule catabolic process, cellular amino acid metabolic process and organic acid catabolic process. Significantly enriched GO terms related to cellular components (CC) category included microbody part, peroxisomal part and peroxisomal matrix. In the molecular function (MF) category, the differentially expressed MRGs were highly enriched in the terms of oxidoreductase activity, coenzyme binding and iron ion binding (Figure 1C). In addition, KEGG pathway analysis revealed that these genes were notably associated with pathways in retinol metabolism, metabolism of xenobiotics by cytochrome P450 and drug metabolism − cytochrome P450 (Figure 1D).

Identification of metabolism-related prognostic signature

To explored the prognostic value of MRGs in renal cancer progression, we performed univariate Cox regression analysis to examine the relationships between the expression levels of those 123 genes and OS. Results demonstrated that 15 genes were significantly associated with OS (P < 0.01) (Figure 2A). Among the 15 MRGs, P4HA3, IL4I1, RRM2, ITPKA, PSAT1, TYMP, HK3, PLCB2 and AANAT were considered as risk gene with HR > 1, while AGMAT, GATM, HAO2, FBP1, ADH6 and ALDH6A1 were considered as protective genes with HR < 1. We then used LASSO Cox regression on the above-mentioned 15 MRGs to identify the most optimal risk score model for predicting survival in KIRC patients (Figure 2B and 2C). Eventually, RRM2 and ALDH6A1 were retained as target genes, and the coefficient of each target gene was calculated to construct the metabolism-related prognostic signature.

Subsequently, we constructed the OS prognostic signature based on the expression of the 2 target genes and its prognostic coefficients using the following formula: Risk score= (0.0361 × expression level of RRM2) + (−0.0184 × expression level of ALDH6A1). According to the median risk score, a total of 253 and 254 KIRC patients were sorted into the high-risk group and low-risk group, respectively. The Kaplan-Meier curve displayed that a significant difference in OS between the high- and the low-risk groups, and patients with high-risk scores exhibited lower survival than patients with low-risk scores (5-year survival rate, 49.3% vs. 72.6%, P <0.001) (Figure 2D). We applied the ROC curve to evaluate the predictive accuracy of the signature, and the area under the curve of the ROC curve was 0.705, suggesting a moderate

prognostic value (Figure 2E). In addition, the distribution of risk scores and the survival status of patients were ranked according to the risk scores (Figure 2F and 2G).

Determination of the prognostic signature as an independent prognostic factor

Furthermore, we performed the univariate and multivariate Cox regression analyses to further determine whether the prognostic signature could serve as an independent prognostic factor. Univariate analysis revealed that the age, grade, AJCC stage, T stage, M stage and risk score were significantly associated with OS (Figure 2H). Subsequent results showed that age (P < 0.001), grade (P = 0.024), AJCC stage (P = 0.043) and risk score (P < 0.001) were still significantly related with OS in multivariate analyses (Figure 2I). Therefore, the metabolism-related prognostic signature was an independent prognostic factor for KIRC patients.

Clinical utility of prognostic signature

We further explored the relationships between metabolism-related prognostic signature and various clinical parameters. The expression levels of the 2 identified target MRGs in high- and low-risk groups were demonstrated in the heatmap (Figure 3A). The results showed that RRM2 and ALDH6A1were represented high and low expression levels in the high-risk group, respectively. Additionally, we observed that there were significant differences between low- and high-risk groups in grade (*P* = 4.841e−08), AJCC stage (*P* = 1.648e−08), T stage (*P* = 3.826e−07), M stage (*P* = 5.918e−05) (Figure 3B-3E).

To better evaluate the survival outcomes and detect the broad applicability of the prognostic signature, we next performed survival analyses stratified by age, gender, tumor grade, AJCC stage, T stage and M stage. As shown in Figure 4, high-risk group had significantly shorter OS than those in the low-risk group for the cases with age ≤ 60 (*P* = 1.236e−02), age > 60 (*P* = 2.404e−05), female patients (*P* = 1.139e−04), male patients (*P* =3.673e−04), G1-2 (*P* = 3.163e−02), G3-4 (*P* = 2.236e−02), AJCC stage □&□ (*P* =1.166e−02), T1-2 (*P* = 2.623e−04) and M0 (*P* = 5.778e−03). However, no significant differences were observed for OS between high- and low-risk groups for the KIRC patients with AJCC stage □&□ (*P* = 7.808e−02), T3-4 (*P* = 1.673e−01) and M1 (*P* = 7.726e−02).

We then analyzed the relationship between the expression level of each target gene of the prognostic signature and clinicopathological features to assess the function of those 2 genes in disease progression. The results indicated that enhanced RRM2 expression was significantly associated with advanced tumor stage and high-grade tumor, suggesting that RRM2 was a poor prognostic factor. The highest expression in RRM2 was found in the most progressive clinicopathological stage, that is, G4 and stage IV, T4 and M1 (Figure 5A). On the contrary, ALDH6A1 expression is gradually decreasing in the progression of KIRC (Figure 5B), which suggested that ALDH6A1 was a protective factor for KIRC.

Construction and validation of the predictive nomogram

A predictive nomogram was constructed by incorporating prognostic signature and several clinical parameters to generate individual numerical probabilities of OS (Figure 6A). The C-index of the developed

nomogram was 0.774. The DCA showed demonstrates that using this nomogram to predict OS had higher net benefit if the threshold probability was larger than 3% (Figure 6B). Additionally, there is an obviously higher net benefit occurred in the nomogram than the tumor grade and AJCC stage. The calibration curves indicated the nomogram performed well in predicting 1-, 3- and 5-year OS compared with the ideal model (Figure 6C-E). Taking together, these results showed that the reliability and predictability of the nomogram.

External validation the expression level and prognosis value of RRM2 and ALDH6A1

The expression level of RRM2 was significantly up-expressed, while ALDH6A1 was significantly down-expressed in KIRC samples compared with the normal samples in the Oncomine database (Figure 7A), TIMER database (Figure 7B) and GEPIA database (Figure 7C), which were consistent with our results. Besides, the aberrant expression of these 2 genes was found to be frequently observed in various types of cancer patients. Interestingly, RRM2 and ALDH6A1 consistently maintained over-expression and under-expression in various cancer, respectively. In addition, the prognostic values of the two genes were further confirmed by the Kaplan Meier plotter in the GEPIA database. The results indicated that RRM2 low-expression group and ALDH6A1 high-expression group had favorable prognosis (Figure 7D).

To determine the protein expression levels of the RRM2 and ALDH6A1, the present study referred to the HPA database. According to the results, we discovered that ALDH6A1 expression in kidney cancer tissues was significantly lower than that in normal tissues (Figure 7E). However, no significant difference was found in RRM2 between kidney cancer tissues and normal tissues. Taking together, the aberrant expression of the two genes was further validated in kidney cancer. Such evidence revealed the robustness and reliability of our results to some extent.

In addition, the expression level of RRM2 and ALDH6A1 were further verified with two independent cohorts (GSE53757, Figure 7F and GSE66270, Figure 7G) in GEO database. Results showed that the expression level of RRM2 was significantly up-expressed, while ALDH6A1 was significantly down-expressed in KIRC samples compared with the normal samples in multiple datasets, which was consistent with our previous result.

GSEA

The top 50 significant genes of positive and negative correlation were obtained by GSEA. Subsequently, we performed hallmark analysis for RRM2 and ALDH6A1. Results indicated that the most significant pathways of RRM2 included ANTIGEN PROCESSING AND PRESENTATION, ALLOGRAFT REJECTION, SYSTEMIC LUPUS ERYTHEMATOSUS, LEISHMANIA INFECTION and AUTOIMMUNE THYROID DISEASE (Figure 8A). In addition, the heatmap showed transcriptional expression profiles of the top 50 features for each phenotype in RRM2 (Figure 8B). Correspondingly, GSEA enrichment analysis showed that the significantly enriched pathways for ALDH6A1 included VALINE LEUCINE AND ISOLEUCINE DEGRADATION, FATTY ACID METABOLISM, PROPANOATE METABOLISM, CITRATE CYCLE TCA CYCLE

and PYRUVATE METABOLISM (Figure 9A). In addition, the heatmap showed transcriptional expression profiles of the top 50 features for each phenotype in ALDH6A1 (Figure 9B).

Discussion

Dysregulation of cellular metabolism is now regarded as a hallmark of cancer [18, 19]. King et al. found that alterations of metabolic genes that could provide a direct genetic link to altered metabolism [20]. Metabolism can trigger tumor development through different mechanisms, and conversely, mutations of oncogenes or/and tumor suppressor genes drive cell proliferation and survival programs [21]. Emerging evidence indicated KIRC is a disease that is affected by complex gene interactions with diverse underlying mechanisms including dysregulated cellular metabolism [22]. Recently, several known kidney cancer genes have been reported to be involved in metabolic stress response including VHL, MET, FLCN, TSC1, TSC2, FH, and SDH [23]. Sunitinib is currently the most commonly used drug focusing on the VHL pathway, but the partial response rate for the advanced KIRC is only 31% [24]. Therefore, identifying an effective metabolic-related signature responsible for tumor progression not only assist us to understand the potential molecular mechanism but provide KIRC patients with appropriate target treatments.

In our study, we investigate the relationships between the expression profiling of MRGs and the prognosis of KIRC patients by integrating analysis of metabolomic and transcriptomic data, which is a valuable paradigm for cancer biology to ensure the reliability of metabolic biomarkers. We first examined differences in the metabolic status of KIRC tissues and normal tissues from TCGA and noticed that 123 out of 944 MRGs were differentially expressed in KIRC patients, including 60 up-regulated and 63 down-regulated genes. Furthermore, GO and KEGG analyses of the differentially expressed MRGs were applied to explore the molecular and biological pathways enriched. The results of GO functional annotation in terms of biological processes suggested that those genes were highly enriched in metabolism. In the KEGG pathway analysis, the most significant pathway was highly correlated with retinol metabolism, which was consistent with previous a major discovery that retinol and related metabolites could regulate the growth and differentiation of a wide variety of cell types, including KIRC [25]. According to univariate Cox regression and LASSO regression analyses, we then constructed a metabolism-related signature consisting of 2 prognostic MRGs (RRM2 and ALDH6A1). The ROC curves and AUCs indicated that the prognostic signature performed well. Subsequent clinical application analysis further demonstrated that the signature could accurately discriminate the prognostic differences between high- and low-risk groups.

Our study further demonstrated that the prognosis was an independent prognostic factor for OS in KIRC patients, suggesting that metabolism status might be a useful prognostic indicator. We also found that the gene signature-based risk score for each patient was an important clinical variable in the constructed nomogram model, which indicated the signature was of great significance in predicting the prognostic value of the KIRC patients. Our results were consistent with what is already a well-known fact that patients with old ages, higher grade and advanced stage were closely related to the poor prognostic. Additionally, we demonstrate that the signature-based nomogram model had higher net benefits than did the tumor grade and AJCC stage by performing DCA. The results from calibration plots and the C-index

showed that the generated nomogram performed well in terms of discriminating varied clinical outcomes of KIRC patients. Taken together, the two-gene signature could greatly improve the prognosis of patients with KIRC.

Among the two genes in the signature, RRM2 has been reported to be closely related to tumorigenesis and tumor progression [26]. For example, in *vitro* and *vivo* experiments showed that overexpression of RRM2 promoted epithelial-mesenchymal transition, whereas knockdown of RRM2 inhibited its oncogenic function in prostate cancer [27]. Sun et al. found that RRM2 was a positive regulator in the progression of glioma cancer that promoted glioma cell proliferation and migration via ERK1/2 and AKT signaling [28]. Cancer cells generally require more dNTP supply for their malignant cell growth than their corresponding normal cells. Prior research showed that RRM2 was essential for DNA synthesis and DNA repair, and was highly overexpressed in multiple solid tumors, which led to the recognition of RRM2 as an effective target of anticancer therapies [26, 29]. ALDH6A1, a mitochondrial methylmalonate semialdehyde dehydrogenase, was demonstrated to be involved in the lipid metabolism [30] and valine catabolic pathway [31]. It has been previously reported that ALDH6A1 expression was markedly downregulated in KIRC tissues. Overexpression of ALDH6A1 significantly decreased cell proliferation and migration as well as impaired oncologic metabolism of KIRC [32]. Results from another study showed that hepatic neoplastic transformation seemed to inhibit the expression of ALDH6A1 through extensive quantitative proteomic profiling analysis and molecular characterization [33]. Considering the close affiliation of ALDH6A1 suppression and abnormal cancer cell growth, ALDH6A1 may serve as a novel biomarker of and potential therapeutic target. Two-dimensional gel electrophoresis and mass spectrometry uncovered that ALDH6A1 was highly specific to metastatic tumor cells and was significantly reduced in metastatic prostate cancer [34]. In summary, the expression levels of RRM2 and ALDH6A1 in our results were in line with the mentioned above studies.

Metabolomics is a crucial tool in the progression of tumor cells, which is a post-genomic discipline studying the metabolome. However, a single metabolic gene may not accurately reflect a pattern because the metabolic state is always shifting. In the present study, we investigated the prognostic role of multiple altered metabolic genes in the TCGA-KIRC cohort but not confined to a single metabolic gene. Despite the potentially significant clinical implications of our results, some of the limitations should also be considered. First, this study was a retrospective data collection and analysis, and may cause the possibility of some inevitable bias. Second, the expression levels of the two genes were only verified in the public database, proving the prognostic value of this signature in independent cohorts shall be warranted to expand the application of our signature. Third, it is necessary to perform further experimental verification in *vivo* and *vitro* to illustrate the mechanisms underlying the regulation of the predictive metabolic genes.

Conclusion

In summary, through a comprehensive analysis of metabolomic and transcriptomic data from the TCGA-KIRC cohort, we extracted two metabolic-related genes and constructed a novel metabolic signature with

the ability to accurately and independently predict the prognosis of KIRC patients. Furthermore, the gene signature-based risk score for each patient was proved to be an important clinical variable in constructed nomogram model, which could predict a 3- and 5-year survival probability for individual KIRC patients by integrating prognostic signature and clinical parameters. Therefore, our findings indicated that the metabolic-related signature might be of great clinical significance, and has the potential to become a valuable diagnostic and prognostic biomarker as well as a promising therapeutic target for KIRC.

Abbreviations

AJCC: The American Joint Committee on Cancer; AUC: area under the ROC curve; BP: biological processes; CC: cellular components; C-index: concordance index; DCA: decision curve analysis; GEO: Gene Expression Omnibus; GO: Gene Ontology; GSEA: Gene set enrichment analysis; HR: hazard ratio; KEGG: Kyoto Encyclopedia of Genes and Genomes; KIRC: kidney renal clear cell carcinoma; LASSO: least absolute shrinkage and selection operator; MF: molecular function; MRG: metabolism-related genes; OS: overall survival; PCA: principal component analysis; RCC: renal cell carcinoma; ROC: receiver operating characteristic; TCGA: The Cancer Genome Atlas.

Declarations

Data Availability Statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

The study was approved by Shandong Key Research and Development Program, China (2019GSF108263).

Authors' contributions

GXD designed the study and was a major contributor to writing the manuscript. SZL conducted the bioinformatics and statistical analysis. JSB and JXB supervised and coordinated the whole research. WHB revised and finalized the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors gratefully thank the academic editor and the anonymous reviewers for their insightful comments and suggestions to improve this manuscript.

References

1. Chow WH, Dong LM, Devesa SS: **Epidemiology and risk factors for kidney cancer.** *Nat Rev Urol* 2010, **7:**245-257.
2. Siegel RL, Miller KD, Jemal A: **Cancer Statistics, 2017.** *CA: a cancer journal for clinicians* 2017, **67:**7-30.
3. Rini BI, Campbell SC, Escudier B: **Renal cell carcinoma.** *Lancet* 2009, **373:**1119-1132.
4. Lambea J, Anido U, Etxániz O, Flores L, Montesa Á, Sepúlveda JM, Esteban E: **The Wide Experience of the Sequential Therapy for Patients with Metastatic Renal Cell Carcinoma.** *Curr Oncol Rep* 2016, **18:**66.
5. Basu B, Eisen T: **Perspectives in drug development for metastatic renal cell cancer.** *Target Oncol* 2010, **5:**139-156.
6. Chaffer CL, Weinberg RA: **A perspective on cancer cell metastasis.** *Science* 2011, **331:**1559-1564.
7. Pavlova NN, Thompson CB: **The Emerging Hallmarks of Cancer Metabolism.** *Cell Metab* 2016, **23:**27-47.
8. Petersen MC, Vatner DF, Shulman GI: **Regulation of hepatic glucose metabolism in health and disease.** *Nat Rev Endocrinol* 2017, **13:**572-587.
9. Altman BJ, Stine ZE, Dang CV: **From Krebs to clinic: glutamine metabolism to cancer therapy.** *Nat Rev Cancer* 2016, **16:**619-634.
10. 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:**76.
11. Mondanelli G, Iacono A, Carvalho A, Orabona C, Volpi C, Pallotta MT, Matino D, Esposito S, Grohmann U: **Amino acid metabolism as drug target in autoimmune diseases.** *Autoimmun Rev* 2019, **18:**334-348.
12. Amoedo ND, Obre E, Rossignol R: **Drug discovery strategies in the field of tumor energy metabolism: Limitations by metabolic flexibility and metabolic resistance to chemotherapy.** *Biochim Biophys Acta*

Bioenerg 2017, **1858:**674-685.

13. Liu GM, Xie WX, Zhang CY, Xu JW: **Identification of a four-gene metabolic signature predicting overall survival for hepatocellular carcinoma.** *J Cell Physiol* 2020, **235:**1624-1636.
14. **Comprehensive molecular characterization of clear cell renal cell carcinoma.** *Nature* 2013, **499:**43-49.
15. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M: **KEGG as a reference resource for gene and protein annotation.** *Nucleic Acids Res* 2016, **44:**D457-462.
16. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies.** *Nucleic Acids Res* 2015, **43:**e47.
17. Friedman J, Hastie T, Tibshirani R: **Regularization Paths for Generalized Linear Models via Coordinate Descent.** *J Stat Softw* 2010, **33:**1-22.
18. Yang HB, Xu YY, Zhao XN, Zou SW, Zhang Y, Zhang M, Li JT, Ren F, Wang LY, Lei QY: **Acetylation of MAT IIα represses tumour cell growth and is decreased in human hepatocellular cancer.** *Nat Commun* 2015, **6:**6973.
19. Hakimi AA, Reznik E, Lee CH, Creighton CJ, Brannon AR, Luna A, Aksoy BA, Liu EM, Shen R, Lee W, et al: **An Integrated Metabolic Atlas of Clear Cell Renal Cell Carcinoma.** *Cancer Cell* 2016, **29:**104-116.
20. King A, Selak MA, Gottlieb E: **Succinate dehydrogenase and fumarate hydratase: linking mitochondrial dysfunction and cancer.** *Oncogene* 2006, **25:**4675-4682.
21. Dang CV: **Links between metabolism and cancer.** *Genes Dev* 2012, **26:**877-890.
22. Cairns RA, Harris IS, Mak TW: **Regulation of cancer cell metabolism.** *Nat Rev Cancer* 2011, **11:**85-95.
23. Linehan WM, Srinivasan R, Schmidt LS: **The genetic basis of kidney cancer: a metabolic disease.** *Nat Rev Urol* 2010, **7:**277-285.
24. Motzer RJ, Hutson TE, Tomczak P, Michaelson MD, Bukowski RM, Oudard S, Negrier S, Szczylik C, Pili R, Bjarnason GA, et al: **Overall survival and updated results for sunitinib compared with interferon alfa in patients with metastatic renal cell carcinoma.** *J Clin Oncol* 2009, **27:**3584-3590.
25. Zhan HC, Gudas LJ, Bok D, Rando R, Nanus DM, Tickoo SK: **Differential expression of the enzyme that esterifies retinol, lecithin:retinol acyltransferase, in subtypes of human renal cancer and normal kidney.** *Clin Cancer Res* 2003, **9:**4897-4905.
26. Chen G, Luo Y, Warncke K, Sun Y, Yu DS, Fu H, Behera M, Ramalingam SS, Doetsch PW, Duong DM, et al: **Acetylation regulates ribonucleotide reductase activity and cancer cell growth.** *Nat Commun* 2019, **10:**3213.
27. Mazzu YZ, Armenia J, Chakraborty G, Yoshikawa Y, Coggins SA, Nandakumar S, Gerke TA, Pomerantz MM, Qiu X, Zhao H, et al: **A Novel Mechanism Driving Poor-Prognosis Prostate Cancer: Overexpression of the DNA Repair Gene, Ribonucleotide Reductase Small Subunit M2 (RRM2).** *Clin Cancer Res* 2019, **25:**4480-4492.
28. Sun H, Yang B, Zhang H, Song J, Zhang Y, Xing J, Yang Z, Wei C, Xu T, Yu Z, et al: **RRM2 is a potential prognostic biomarker with functional significance in glioma.** *Int J Biol Sci* 2019, **15:**533-543.

29. Aye Y, Li M, Long MJ, Weiss RS: **Ribonucleotide reductase and cancer: biological mechanisms and targeted therapies.** *Oncogene* 2015, **34:**2011-2021.
30. Knebel B, Goeddeke S, Poschmann G, Markgraf DF, Jacob S, Nitzgen U, Passlack W, Preuss C, Dicken HD, Stühler K, et al: **Novel Insights into the Adipokinome of Obese and Obese/Diabetic Mouse Models.** *Int J Mol Sci* 2017, **18**.
31. Sass JO, Walter M, Shield JP, Atherton AM, Garg U, Scott D, Woods CG, Smith LD: **3-Hydroxyisobutyrate aciduria and mutations in the ALDH6A1 gene coding for methylmalonate semialdehyde dehydrogenase.** *J Inherit Metab Dis* 2012, **35:**437-442.
32. Lu J, Chen Z, Zhao H, Dong H, Zhu L, Zhang Y, Wang J, Zhu H, Cui Q, Qi C, et al: **ABAT and ALDH6A1, regulated by transcription factor HNF4A, suppress tumorigenic capability in clear cell renal cell carcinoma.** *J Transl Med* 2020, **18:**101.
33. Shin H, Cha HJ, Lee MJ, Na K, Park D, Kim CY, Han DH, Kim H, Paik YK: **Identification of ALDH6A1 as a Potential Molecular Signature in Hepatocellular Carcinoma via Quantitative Profiling of the Mitochondrial Proteome.** *J Proteome Res* 2020, **19:**1684-1695.
34. Cho SY, Kang S, Kim DS, Na HJ, Kim YJ, Choi YD, Cho NH: **HSP27, ALDH6A1 and Prohibitin Act as a Trio-biomarker to Predict Survival in Late Metastatic Prostate Cancer.** *Anticancer Res* 2018, **38:**6551-6560.

Figures

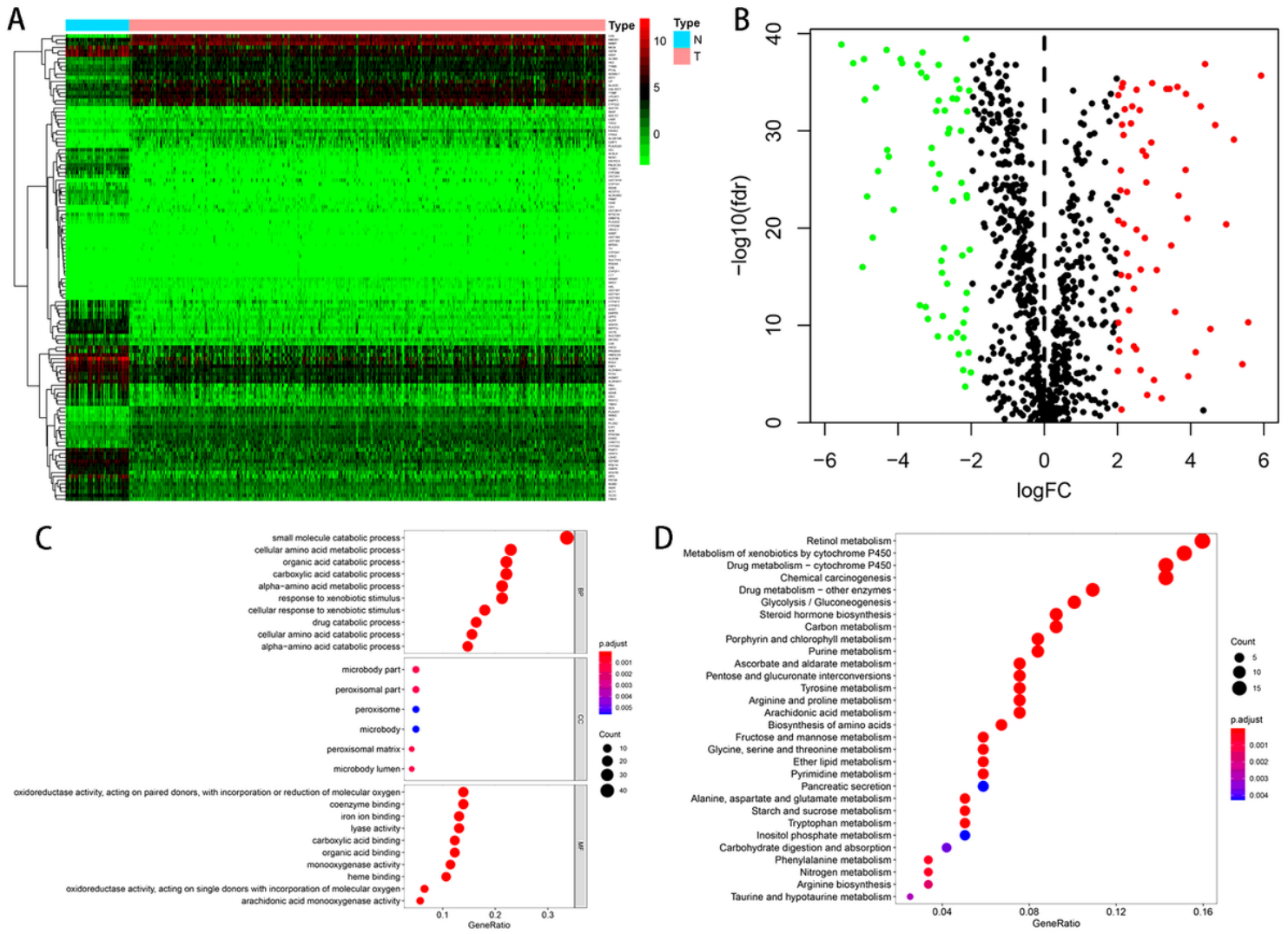

Figure 1

The differentially expressed metabolism-related genes (MRGs) and functional enrichment analysis. (A) The heatmap of the 123 differently expressed metabolism-related genes between kidney cancer and normal tissues. (B) The volcano plot of the 123 differently expressed metabolism-related genes. Red dots indicate significantly up-regulated genes, green dots indicate significantly down-regulated genes and black dots indicate no difference genes. (C) GO analysis shows the significantly enriched biological processes (BP), cellular components (CC) and molecular functions (MF) involved in differential genes. (D) KEGG shows the significantly enriched pathways involved in differential genes.

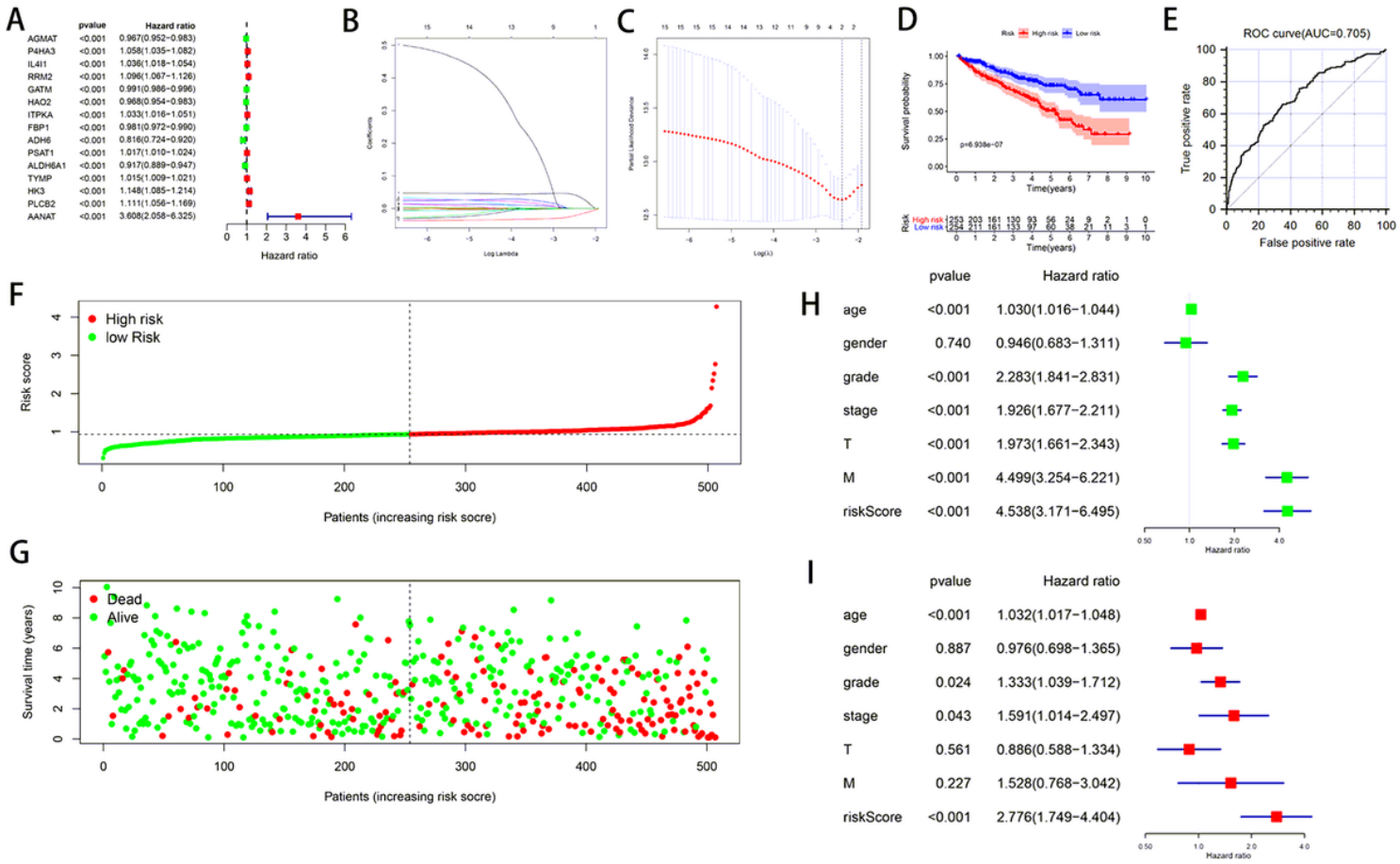

Figure 2

Construction of the prognostic signature based on TCGA-KIRC cohort. (A) Univariate Cox regression analysis demonstrates that 15 MRGs are significantly associated with overall survival. (B and C) Screening of the optimal MRGs used for the final construction of the predictive signature using Lasso regression analysis. (D) Patients in the high-risk group have shorter overall survival. (E) ROC analysis demonstrates the survival prediction accuracy. (F) The distributions of risk scores. (G) The survival status for KIRC patients in high- and low-risk groups. Univariate (H) and (I) multivariate Cox regression analysis verify the metabolism-related prognostic signature was an independent prognostic factor for KIRC patients.

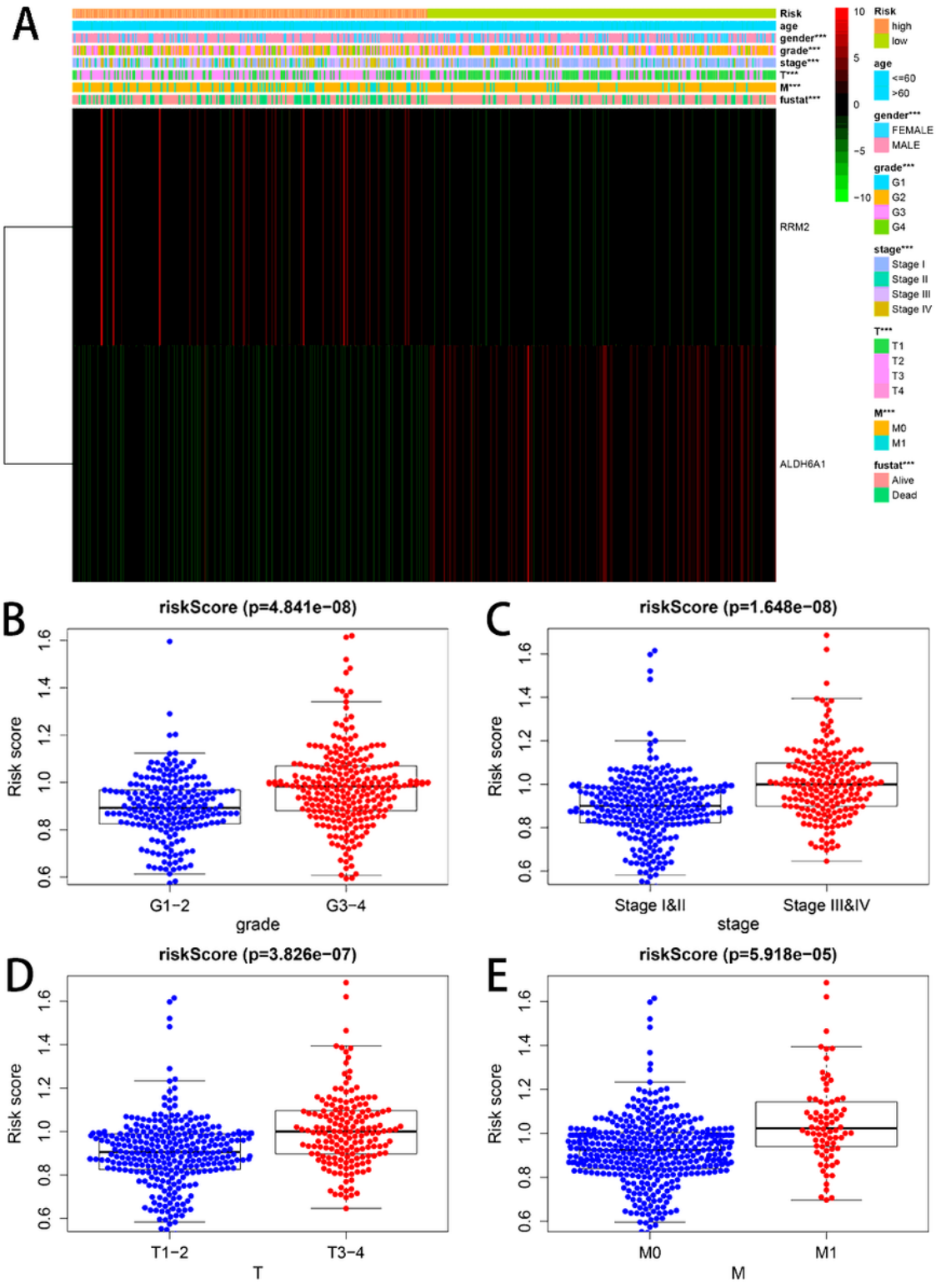

Figure 3

The relationship between the risk score and different clinicopathological features. (A) The heatmap shows the distribution of clinical parameters and the expression of two genes between the low- and high-risk groups. *** P < 0.001. (B), (C), (D) and (E) represent grade, AJCC stage, T stage and M stage, respectively.

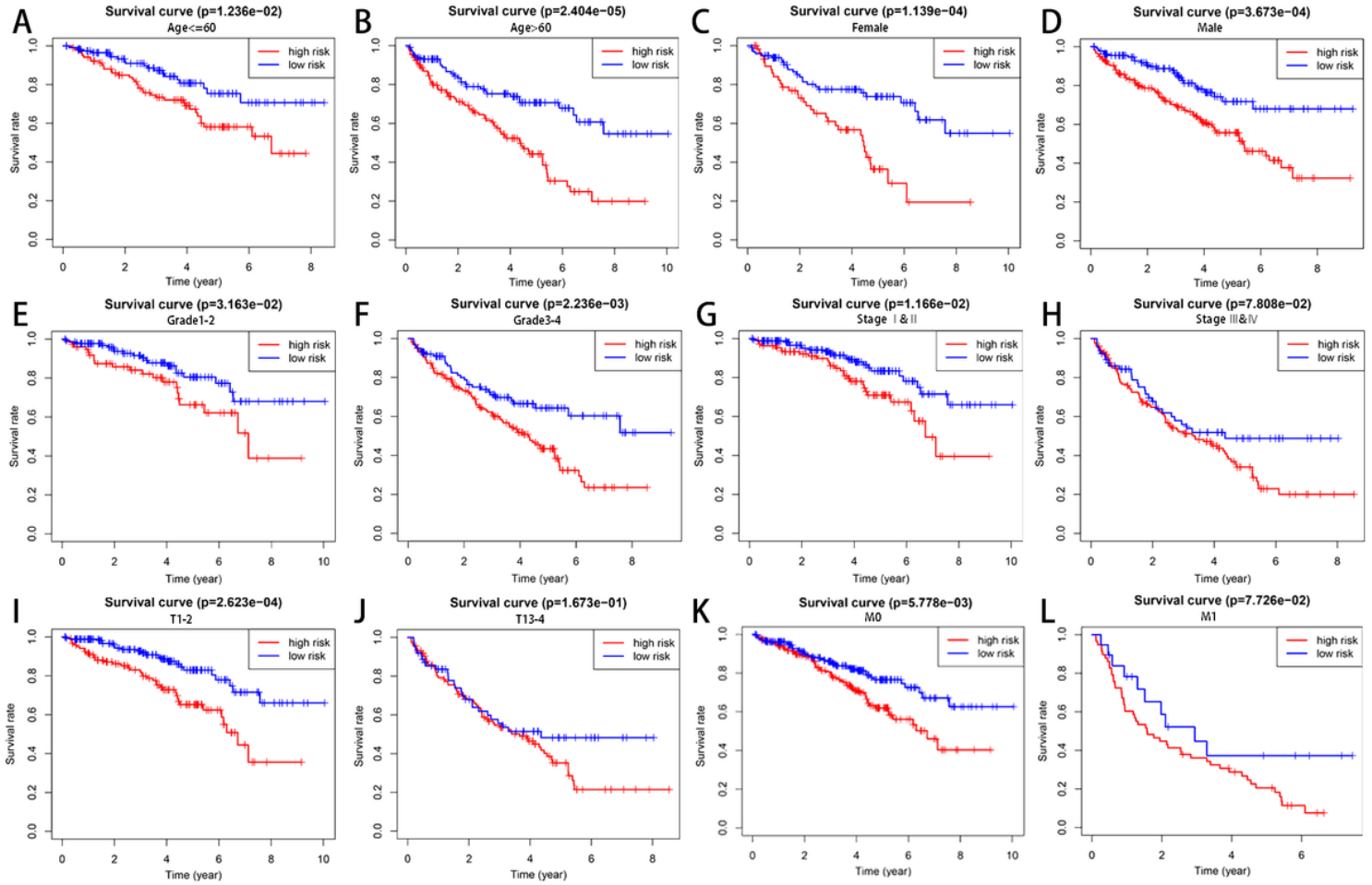

Figure 4

The survival difference between high- and low-risk group stratified by clinical parameters. The difference in OS between high- and low-risk group stratified by age (A and B), gender (C and D), grade (E and F), AJCC stage (G and H), T stage (I and J) and N stage (K and L).

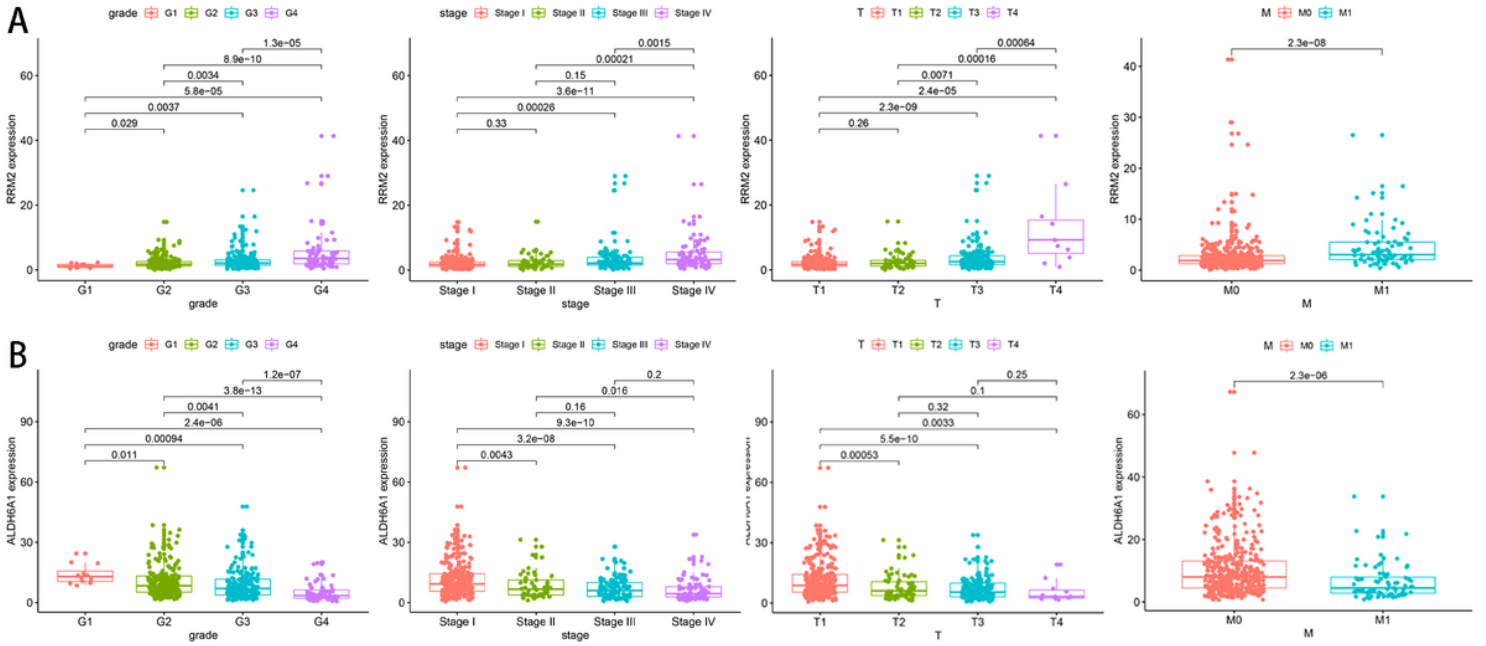

Figure 5

The correlation between the expression level of a single gene in the signature and various clinical parameters. (A) RRM2. (B) ALDH6A1.

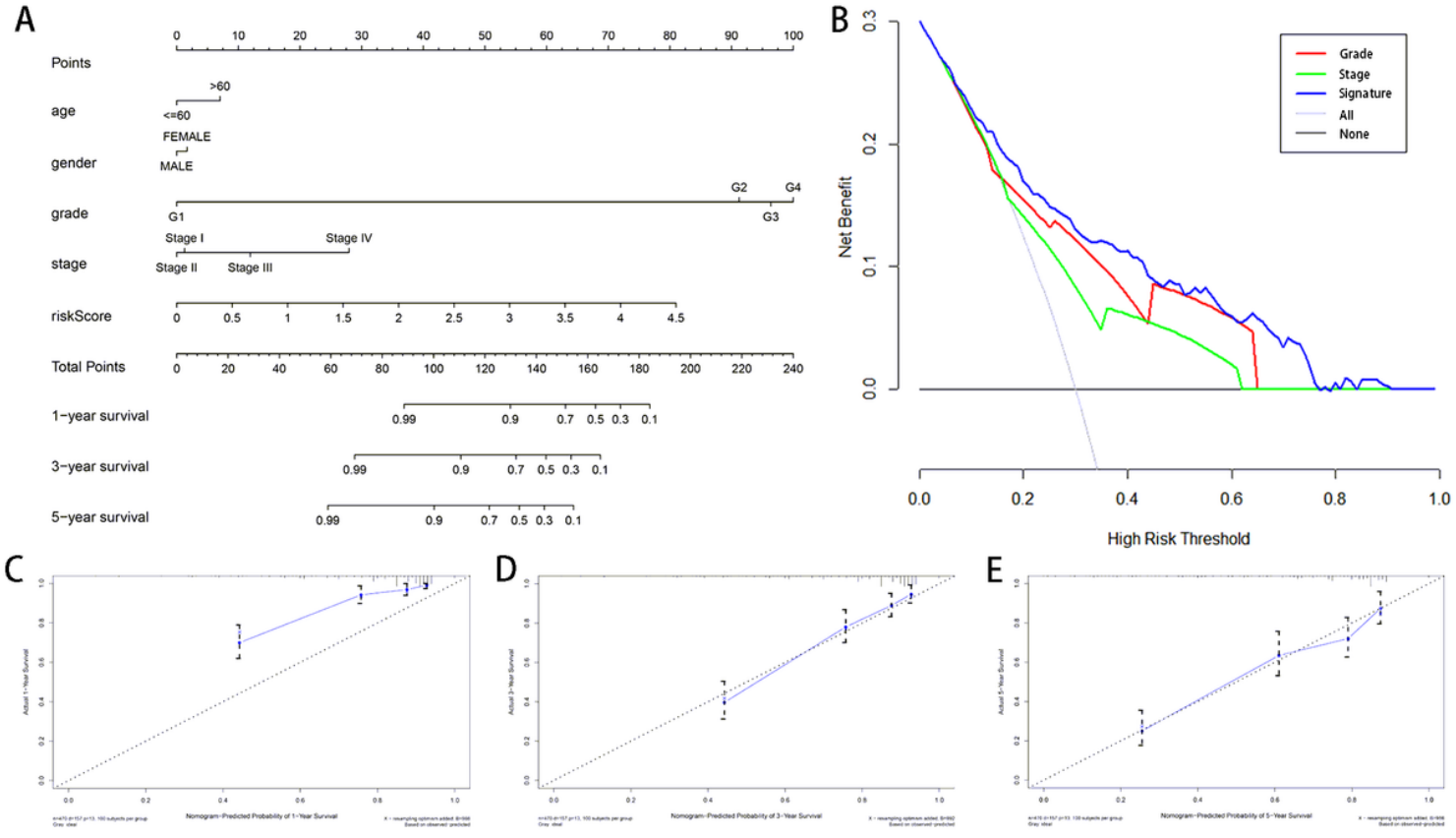

Figure 6

Construction and validation of the predictive nomogram. (A) Nomogram for predicting 1-, 3-, and 5-year OS of KIRC patients in the TCGA cohort. (B) The DCA showed demonstrates that using this nomogram to predict OS had higher net benefit if the threshold probability was larger than 3%. (C-E) The calibration curves indicate the nomogram performed well in predicting 1-, 3- and 5-year OS compared with the ideal model.

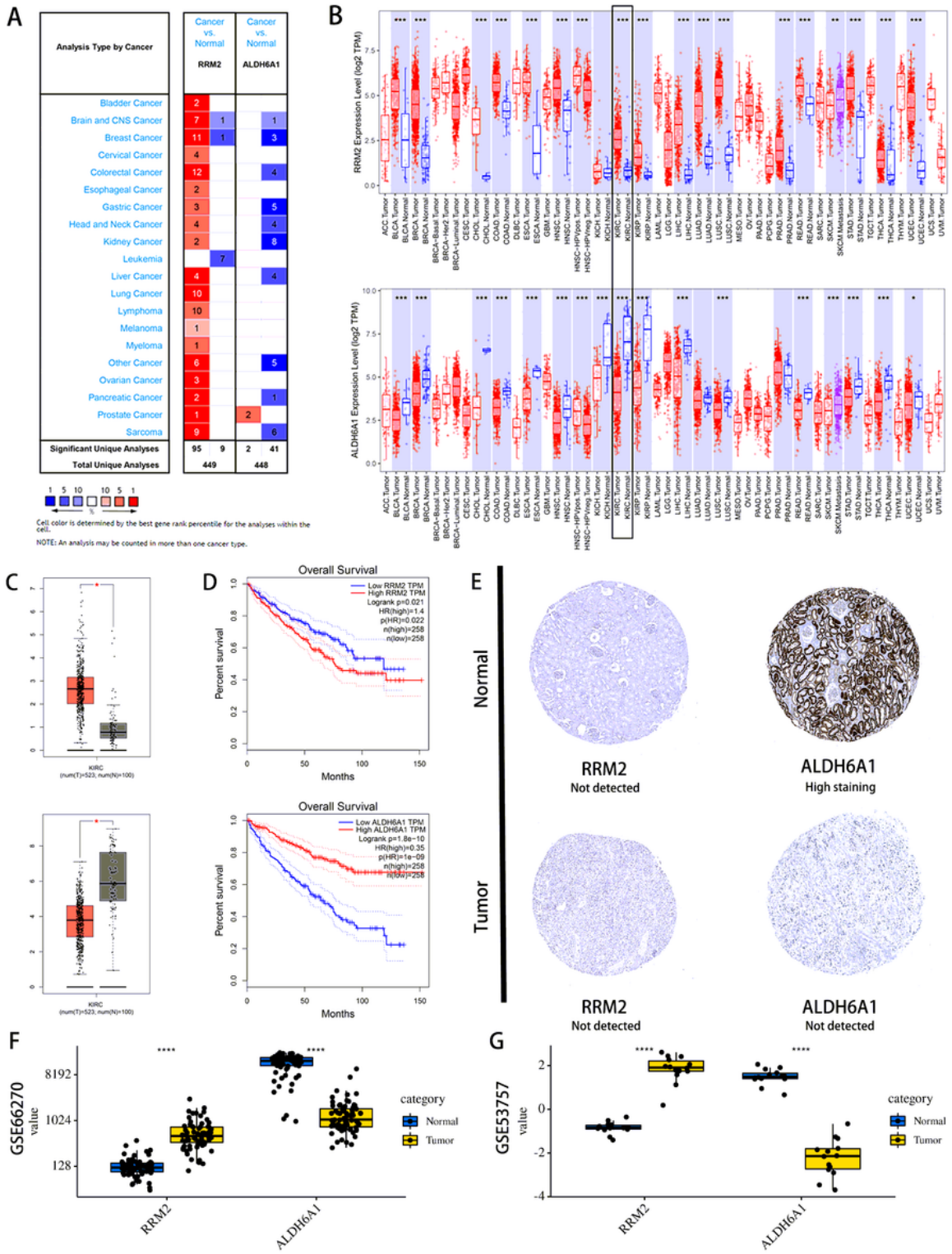

Figure 7

Expression level and prognosis value of the identified two predictive genes. (A) The expression profiles of the RRM2 and ALDH6A1 in the Oncomine database. (B) The expression levels of the RRM2 and ALDH6A1in various cancers based on the TIMER database. (C) The expression profiles of the RRM2 and ALDH6A1 in the GEPIA database. (D) Univariate survival analysis of the RRM2 and ALDH6A1 is performed using the Kaplan-Meier curve. (E) The protein expression levels of the RRM2 and ALDH6A1 in

KIRC and normal kidney tissues. Data are from the Human Protein Atlas database. (F) Verification of RM2 and ALDH6A1 expression in KIRC and normal tissues with the GSE53757 in GEO database. (G) Verification of RM2 and ALDH6A1 expression in KIRC and normal tissues with the GSE66270 in GEO database.

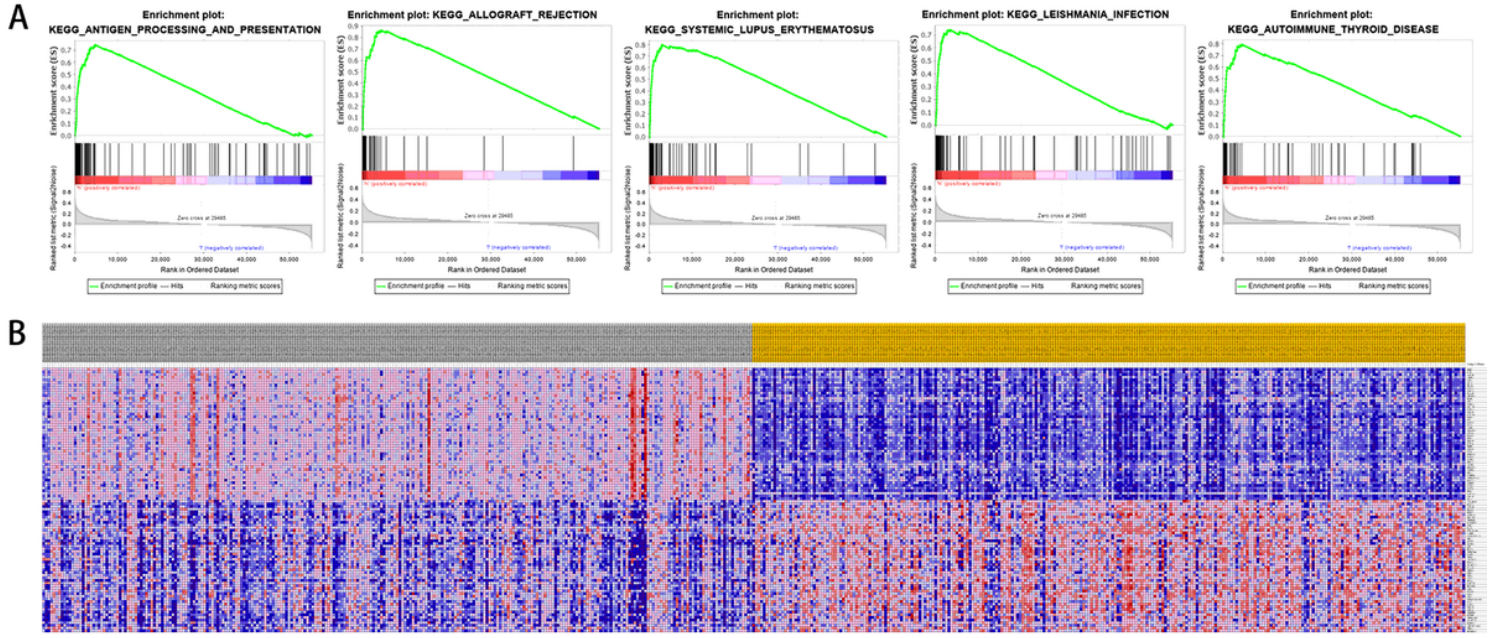

Figure 8

The significant RRM2-related genes and hallmarks pathways in KIRC patients obtained via GSEA. (A) The most significant pathways of RRM2 include ANTIGEN PROCESSING AND PRESENTATION, ALLOGRAFT REJECTION, SYSTEMIC LUPUS ERYTHEMATOSUS, LEISHMANIA INFECTION and AUTOIMMUNE THYROID DISEASE. (B) The heatmap shows transcriptional expression profiles of the top 50 features for each phenotype.

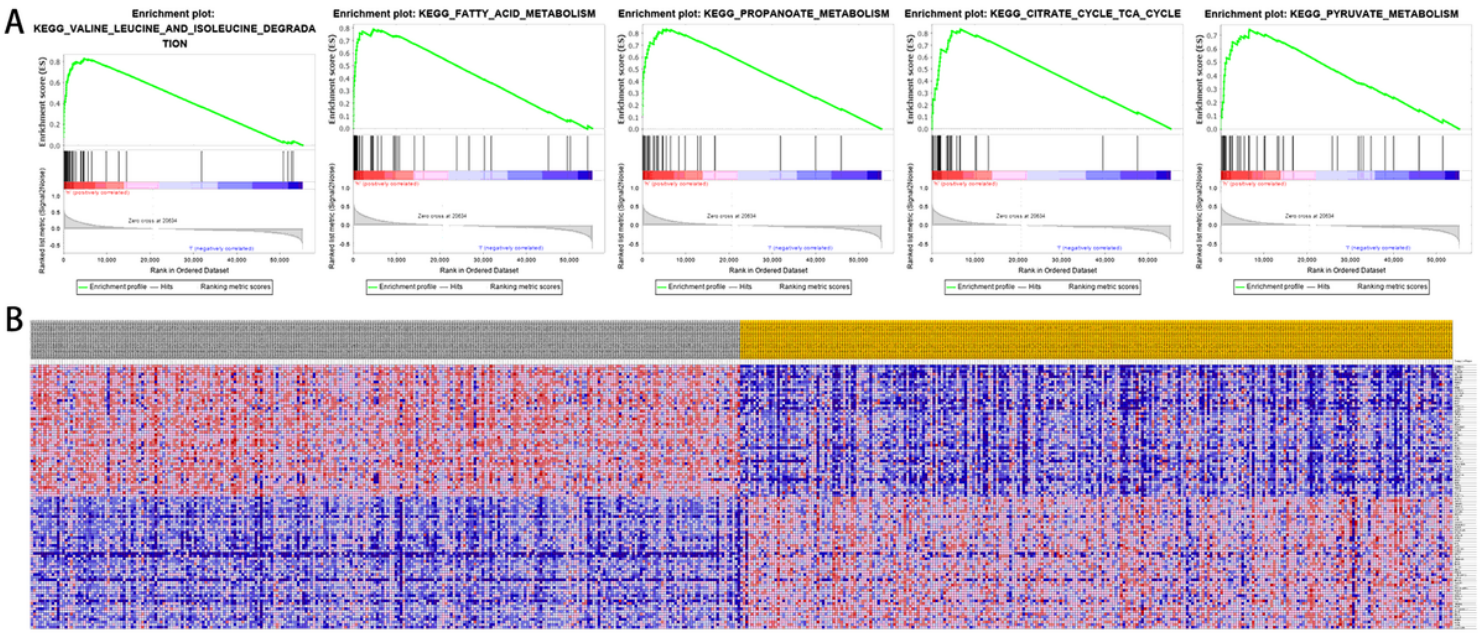

Figure 9

The significant ALDH6A1-related genes and hallmarks pathways in KIRC patients obtained via GSEA. (A) The most significant pathways of RRM2 include VALINE LEUCINE AND ISOLEUCINE DEGRADATION, FATTY ACID METABOLISM, PROPANOATE METABOLISM, CITRATE CYCLE TCA CYCLE and PYRUVATE METABOLISM. (B) The heatmap shows transcriptional expression profiles of the top 50 features for each phenotype.