Original Article

Construction and validation of four-metabolism related-long non-coding RNAs as potential signature in prognosis of colon cancer

Guanxuan Chen[1,2], Yinjun Dong[3], Wenna Shi[4], Lei Li[5], Wanqi Zhu[6], Li Xie[7], Xianrang Song[7]

[1]Cheeloo College of Medicine, Shandong University, Jinan, China; [2]Department of Intensive Care Unit, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan, China; [3]Department of Thoracic Surgery, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan, China; [4]Pharmaceutical Department, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan, China; [5]Weifang Medical University, Weifang, China; [6]Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Taian, China; [7]Department of Clinical Laboratory, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan, China

Contributions: (I) Conception and design: X Song, L Xie; (II) Administrative support: X Song; (III) Provision of study materials: Y Dong, W Zhu; (IV) Collection and assembly of data: W Shi, L Li; (V) Data analysis and interpretation: G Chen; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Xianrang Song. Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, 440 Ji-Yan Road, Jinan 250117, China. Email: xrsong@sdfmu.edu.cn.

Background: Emerging evidence suggests that metabolism plays important roles in the initiation and progression of colon cancer (CC) and the outcomes of CC patients. Long non-coding RNAs (lncRNA) are key regulators of regulatory molecules linking to a wide variety of cancer cellular functions. This study aims to develop a metabolic lncRNA signature to help better predict prognosis for CC patients.

Methods: In the current study, the transcriptome data and clinical data of CC was downloaded from The Cancer Genome Atlas (TCGA). Metabolism-related gene sets were downloaded from the Molecular Signatures Database (MSigDB). Differential lncRNAs related to metabolism was obtained by performing the correlations between differential expression profile of metabolic genes and lncRNAs. To construct a prognostic model of CC based on metabolism-related lncRNAs, we divided patients, whose clinical data were available, into a training set and a validation set at a ratio of 7:3. The prognostic metabolism related-lncRNA signature was established using the training set by univariate and multivariate Cox regression analysis, and the validation set was used to test the capacity of the prognostic model. The correlation between risk score and clinicopathological features, immune function GO and KEGG analysis was investigated using the entire set. Finally, GSEA pathway enrichment analysis was carried out on the entire set samples for the high- and low- risk groups.

Results: We identified 604 differential lncRNAs and 252 genes related to metabolism. After univariate and multivariate Cox regression analysis, four lncRNAs were finally identified to build a signature, which was verified the effectiveness by the TCGA validation set. The multivariate Cox regression analysis showed that the risk score, age of diagnosis and T stage were independent prognostic factor for CC patients. It is shown that some immunopathogenesis, GO items and KEGG pathways demonstrated difference between high- and low- risk group.

Conclusions: We developed a four-metabolism related-lncRNA signature for prognostic prediction of CC, which may help select high-risk subpopulation patients who require more aggressive therapy or intervention.

Keywords: Colon cancer (CC); metabolism; long non-coding RNAs (lncRNA); signature; prognosis

Submitted Sep 17, 2021. Accepted for publication Dec 22, 2021.
doi: 10.21037/tcr-21-2184
View this article at: https://dx.doi.org/10.21037/tcr-21-2184

Introduction

Colon cancer (CC) is one of the common digestive tract malignancies, which has severely threatened human health, and its morbidity and mortality rank third and fourth, respectively (1). The primary tumor, lymph node, distant metastasis (TNM) staging system is the standard classification for receiving different treatment regimens in patients with CC. However, patients with the same TNM stage, pathological classification, and treatment regimen can have different prognoses, which owing to its high heterogeneity (2,3). A large number of research findings have showed that patients with microsatellite instability (MSI) or mismatch repair protein deficiency (dMMR) present a good prognosis and can't benefit from adjuvant chemotherapy of 5-florouracil (5-FU), which constitutes 15–20% of these cancers (4). Cetuximab is one of the EGFR inhibitors that used for the treatment of metastatic CC patients with EGFR-positive, moreover, it was reported that KRAS status is a predictive marker of response to cetuximab (5). Thus, molecular subtypes are needed to predict the prognosis for individual CC patient.

The pathogenesis of CC has not yet been clearly defined, abnormal metabolism may play important roles in cancer development and death. Abnormal metabolisms, including abnormal glucose metabolism, nucleotide metabolism and lipid metabolism, are the main characteristics of CC. Abnormal activation of glycolysis pathway in cancer cells is recently recognized as a hallmark of cancer, cancer cells display activation of glycolysis with more production of lactic acid, which is exported to lead a decrease in extracellular pH during glycolytic metabolism (6), and low pH value in the microenvironment has been associated with poor prognosis in CRC patients (7). Nucleotide metabolism has been reported to be transcriptionally regulated by both tumor suppressor genes and oncogenes, importantly, tumor cells also manifest a larger nucleotide pool and a more active nucleotide anabolic pathway (8,9). Alterations in lipid metabolism are associated with changes in glucose metabolism. Accumulating evidence suggest that cancer cells depend on altered lipid metabolism for unrestrained growth and survival (10). For example, amounts of omega-3 EPA and the arachidonic acid-mediated inflammatory pathways are well-established to influence colonic carcinogenesis (11). Results suggest that modulating the lipid metabolism with a pharmaceutical agent or through diet may be a valid strategy to influence CIN (12).

Long non-coding RNA (lncRNA) whose transcription with no protein coding potential is a non-coding RNA with exceeded 200 nt length, it is highly conserved across the evolution and can regulate gene expression at levels of epigenetic regulation, transcriptional regulation and posttranscriptional regulation (13,14). LncRNA plays a vital role in the development and progression of CC, and study revealed that lncRNA have spatiotemporal and tissue specific expression patterns, thus they are expected to be biomarkers for early diagnosis, predicting the risk of recurrence and metastasis of CC (15). Currently, it is almost impossible to exactly predict the outcomes of CC by a single gene, in contrast, signature of multiple genes combination may be able to effectively predict the prognosis of CC patients, then guide selection in treatment options and intervening to extend their survival time.

However, little is known about the metabolism related-lncRNA signature features driving high or low risk in CC, thus, it is of great value to discover more potential prognostic and predictive markers. In this study, to screen the metabolism related-lncRNAs closely related to patient prognosis, we integrated the gene expression profiles of 473 patients with CC from the TCGA database, and constructed and validated a risk score model to predict the prognosis of CC, the model's accuracy and reliability were also demonstrated. In addition, we conducted a comprehensive analysis of this model with clinical characteristics, immune functions, GO and KEGG analysis. We present the following article in accordance with the TRIPOD reporting checklist (available at https://tcr.amegroups.com/article/view/10.21037/tcr-21-2184/rc).

Methods

Data collection

The dataset of CC samples used in our study was downloaded in the TCGA database (tcga-data.nci.nih.gov/tcga), including transcriptome data and clinical data of 473 patients, then 447 cases with overall survival (OS) data were collected, and 431 of them had complete clinical parameters including age, sex, clinical stage, T-stage, N-stage, M-stage. The metabolism related genes were downloaded from the Molecular Signatures Database (MSigDB http://software.broadinstitute.org/gsea/msigdb/), including 532 Glucose-metabolism, 1,034 lipid-metabolism and 13 glutamine-metabolism related genes (16), these genes were analyzed by differentially expressed genes (DEGs) between tumor and adjacent normal tissue groups using limma package

in R software (version 4.0.5, https://www.r-project.org/). After DEGs analysis of lncRNAs in the TCGA database, the metabolism related-lncRNAs were identified based on the correlation analysis between differentially expressed metabolism related genes and differentially expressed lncRNAs. The filter criteria of DEGs were screened according to log2|fold change| >1, FDR <0.05, the correlation analysis was calculated according to Pearson correlation coefficient >0.4 and P value <0.001.

Risk-score signature establishment and validation

Patients with clinical information in TCGA were randomly divided into training set (312 cases) and validation set (135 cases), with a ratio of approximately 7:3. In the training set, metabolism related-lncRNAs were subjected to univariate Cox proportional hazard regression analysis, then the statistical significant ones were tested in the multivariate Cox regression analysis to generate a coefficient of each lncRNAs. To validate the signature, risk score was calculated according to the lncRNAs expression for every patient in the training set, then all patients were separated into high- and low- risk group by the median risk scores, which was also as cutoff in the validation set. Risk analyses were done using the survival (version 3.1.12), survminer (version 0.4.6), survival ROC (version 1.16.1) and pheatmap (version 0.7.7) package in R. The PCA analyses were done and plotted by R package vegan (version 2.5.7).

Correlation analysis between risk-score signature and clinical characteristics, functional analysis, immune functions, signaling pathways

To assess the contribution of each variable as an independent prognostic factor for patient survival, the Cox proportional hazards model was conducted. Chi-square test was used to analyze clinicopathological characteristics between the two groups. The pathway/ontology enrichment analysis of genes was performed using the Metascape (can be GO/KEGG terms, canonical pathways, hall mark gene sets, etc., based on the default choices under Express Analysis, http://metascape.org). Immune assay was performed for the low risk and the high risk groups, by computing single sample Gene Set Enrichment Analysis (ssGSEA) scores through gsva R package (version 1.38.2). The differences of gene function and pathway between low risk and the high risk groups was performed by Gene-set Enrichment Analysis (GSEA) tool (http://software.broadinstitute.org/gsea/index.jsp).

Statistical analysis

All statistical analyses were performed in R (version 4.0.5, https://www.r-project.org/). Univariate cox proportional hazards regression was used to estimate the HRs. Coefficients of the prognostic signature were calculated by the multivariate Cox regression analysis. The confidence interval (CI) was set at 95%, and a P value <0.05 was considered to indicate a significant difference in the statistical analyses. The survival curve was generated by the Kaplan-Meier method. Pearson test was conducted for correlation analysis.

Ethical statement

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This study was also approved by the ethics committee of Shandong First Medical University and Shandong Academy of Medical. (No. 2022001009) and individual consent for this retrospective analysis was waived.

Results

Identification of metabolism related-lncRNAs signature for predicting prognosis of CC patients

A flow diagram (*Figure 1A*) showed the flow chart about how the four lncRNAs of the signature were extracted. To identify metabolism-related genes, we performed differential gene expression analysis of the metabolism related genes between the CC samples and normal colon samples. Altogether, 252 genes were differentially expressed between the two groups, in which 126 were upregulated and 126 were downregulated [*Figure 1B*, the supplementary table (metabolism-related diff genes) at https://cdn.amegroups.cn/static/public/tcr-21-2184-1.xlsx]. Then to perform lncRNA expression profile between the CC tissues and normal colon tissues, and 604 lncRNAs were differentially expressed in CC, including 532 upregulated and 72 downregulated [*Figure 1C*, the supplementary table (diff lncRNA) at https://cdn.amegroups.cn/static/public/tcr-21-2184-1.xlsx]. Finally, after analyzing the correlation between these differentially expressed metabolism-related genes

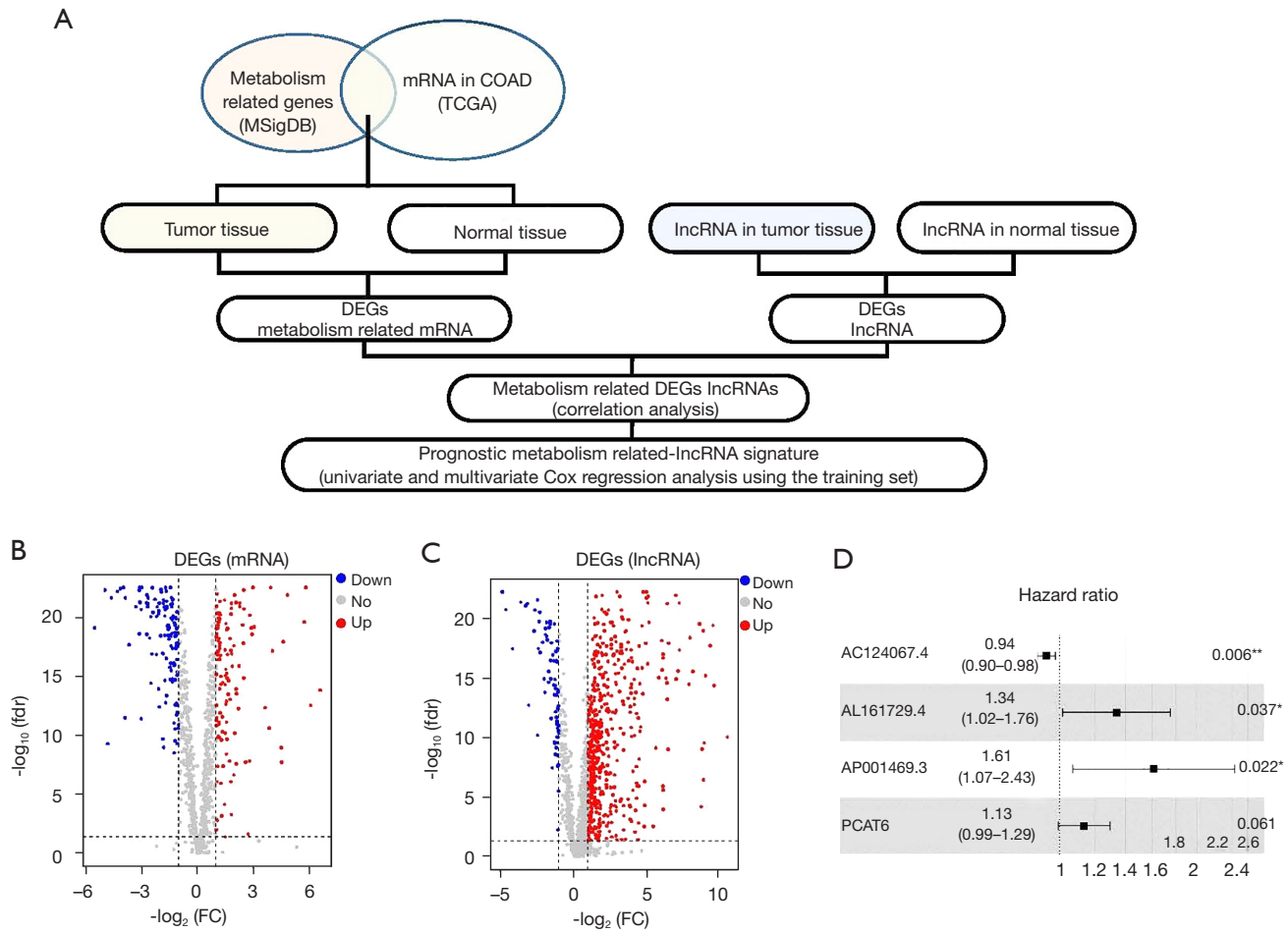

Figure 1 Identification of metabolism-related lncRNAs signature. (A) Flow diagram for the identification. (B) The volcano plot shows the both up- and downregulated differentially expressed genes. (C) The volcano plot shows the differentially expressed lncRNAs. (D) Forest plot of multivariable Cox proportional hazard signature. lncRNAs, long non-coding RNAs. *P<0.05; **P<0.01.

and lncRNAs by applying Pearson correlation analysis, 381 lncRNAs were considered as metabolism-related lncRNAs [the supplementary table (correlation result) at https://cdn.amegroups.cn/static/public/tcr-21-2184-1.xlsx]. To explore prognostic value of lncRNAs in CC, we subjected the 381 differentially expressed lncRNAs to univariable Cox analysis in the training set and screened 12 lncRNAs which were significantly related to prognosis of CC patients [the supplementary table (unicox) at https://cdn.amegroups.cn/static/public/tcr-21-2184-1.xlsx]. Subsequently, those 12 candidate lncRNAs were further analyzed by multivariate Cox regression models in the training set, the final signature was composed of four metabolism-related lncRNAs (AC124067.4, AL161729.4, AP001469.3 and PCAT6). To determine the clinical application, the expression values of lncRNAs of the metabolic signature were transformed into a risk score as follows: risk score = (–0.064 × expression value of AC124067.4) + (0.291 × expression value of AL161729.4) + (0.478 × expression value of AP001469.3) + (0.126 × expression value of PCAT6) [the supplementary table (multicox) at https://cdn.amegroups.cn/static/public/tcr-21-2184-1.xlsx]. Forest plot showed results of multivariate cox analysis, the four lncRNAs in the metabolic signature was mutually independent (*Figure 1D*). To descript of how the four lncRNAs relate to metabolism, we carried out pathway and process enrichment analysis for gene list which related to them. We identified statistically enriched terms as displayed in Figure S1, the heatmap showed that lipids

metabolism appeared significantly enriched.

Evaluation of the metabolic lncRNAs signature for predicting prognosis in the training and validation sets

To further determine whether this risk score model could precisely predict the survival of CC patients, we utilized the training set and validation set to validate the prognosis prediction ability of this model. After calculating risk score for each patient in the training set, we ranked them and the median risk score was the cut-off value, patients with scores below 0.956673094 were classified as the low-risk group, and those with scores equal to or higher than the median were classified as the high-risk group. Then, we divided the patients, in the training set and the validation set, into high-risk group or low-risk group according to the threshold. As shown in the *Figure 2A*, the distribution of risk score, the survival status of the CC patients and these lncRNA expression profiles differences between the two groups were also obtained. With the increase in risk score, the number of dead patients increased and the survival time decreased, which means the mortality rate was higher and the survival time was lower of patients in the high-risk group than that in the low-risk group. The heat map showed the expression level of risk genes, the high-risk group showed different expression of the four lncRNAs. Furthermore, Kaplan-Meier (K-M) analysis indicated significant differences between the high-risk and low-risk groups in the training set and the testing set that the OS in the high-risk group was poor (*Figure 2B*, P<0.001). Next, ROC analyses were performed in both the training set and validation set to evaluate the accuracy of the four-genes signature model in predicting survival at 1, 3, and 5 years, and the areas under the curve (AUC) of the signature were 0.701, 0.708, and 0.646 respectively in the training set, and 0.615, 0.7, and 0.652 in the validation set, which demonstrated that the model had good perform for survival prediction in sensitivity and specificity (*Figure 2C*). To further evaluate the performance of the risk score, the principal component analysis (PCA) was performed, and showed that the high- and low- risk group were scattered separately by the PC2 component (*Figure 2D*).

The associations of the metabolic lncRNAs signature and clinical characteristics

We then assessed the correlation between the risk scores derived from the metabolic lncRNAs signature and the clinical characteristics in CC patients, including TNM stage, race, gender and age in the entire group. The results indicated that there were significant difference between high- and low-risk groups in respect of pathologic N status, patients in high-risk group underwent significant increasing staging compared with those in low-risk group. Additionally, high-risk group showed three higher and one lower expression lncRNAs than did the low-risk group (*Figure 3A*). To examine whether the model was an independent prognostic factor in CC patients, we performed univariate and multivariate Cox regression analyses. In the univariate Cox analysis, pathologic T, N, M, stage status, age, and risk score were independent prognostic factors of CC patients (P<0.001) (*Figure 3B*). After performing multivariate Cox assay, the four metabolic lncRNAs signature was independent of the features (P<0.001), as well as age (P<0.001), T status (P=0.032), stage status (P=0.026) (*Figure 3C*).

Based on univariate and multivariate Cox regression analysis, a nomogram which integrated the metabolic signature, stage status, age, and T status was developed based on the entire group, the C-index of nomogram was 0.762 (95% CI: 0.735–0.789) (*Figure 3D*). The total points which were figured out using each variable could be converted to predict 1-, 3-, and 5-year probability of OS of CC patients. This result showed our signature was effective and would be a good index parameter in CC patients prognostic prediction (*Figure 3E*).

Difference in immune function and GO and KEGG analysis between high- and low risk group

To assess the association between the immune function and risk scores, we analyzed 29 immune gene sets including immune related pathways, immune cell types and immune related functions in patients with CC. As shown in *Figure 4A*, the Treg, iDCs, exhibited a more marked difference between high- and low- risk group, however there were no differences in CD8 T cells and check-point, which were the popular targets for cancer immunotherapy. (*Figure 4A*). Then, all genes in the entire set were used to perform gene-set enrichment analysis (GSEA) to compare the difference between the two groups, and 3985/8350 gene sets are unregulated in high-risk group (the supplementary table at https://cdn.amegroups.cn/static/public/tcr-21-2184-2.xlsx), 2 gene sets are significant at FDR <25%, 41 gene sets are significantly enriched at nominal P value <1%, 177 gene sets are significantly enriched at nominal

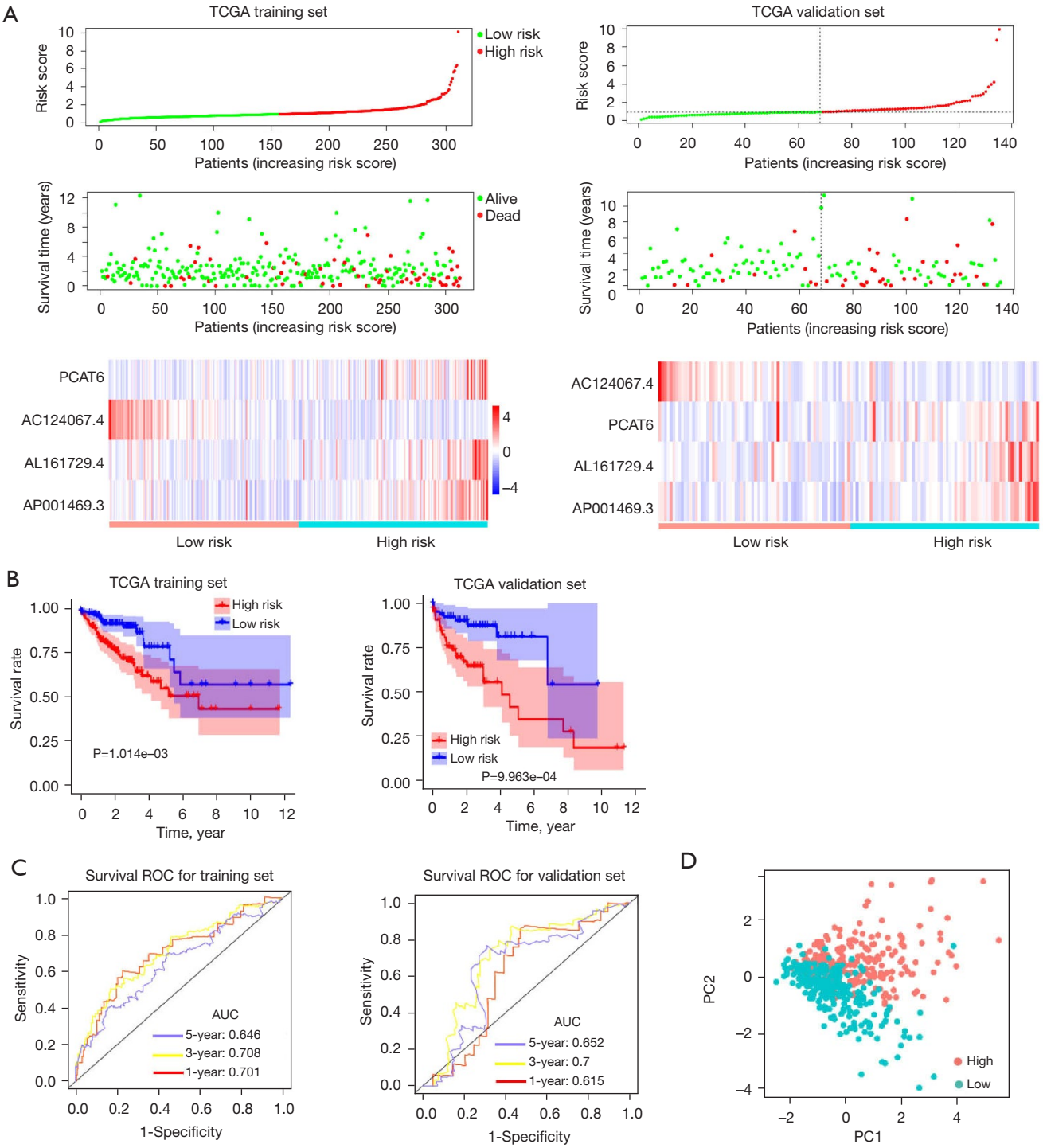

Figure 2 Estimation of the metabolism-related lncRNAs signature in training and validation sets. (A) Risk score distribution, survival status and expression of risk genes in the high-risk and low-risk group of CC patients in the training set (left) and validation set (right). (B) Kaplan-Meier survival curves of OS between high- and low-risk groups in the training set (left) and validation set (right). (C) Time-dependent receiver operating characteristic (ROC) curves at 5-, 3- and 1-year of OS for the signature in the training set (left) and validation set (right). (D) PCA of all samples based on risk scores; each point represents a sample, and different colors distinguish the groups. lncRNAs, long non-coding RNAs; CC, colon cancer; OS, overall survival; PCA, principal component analysis.

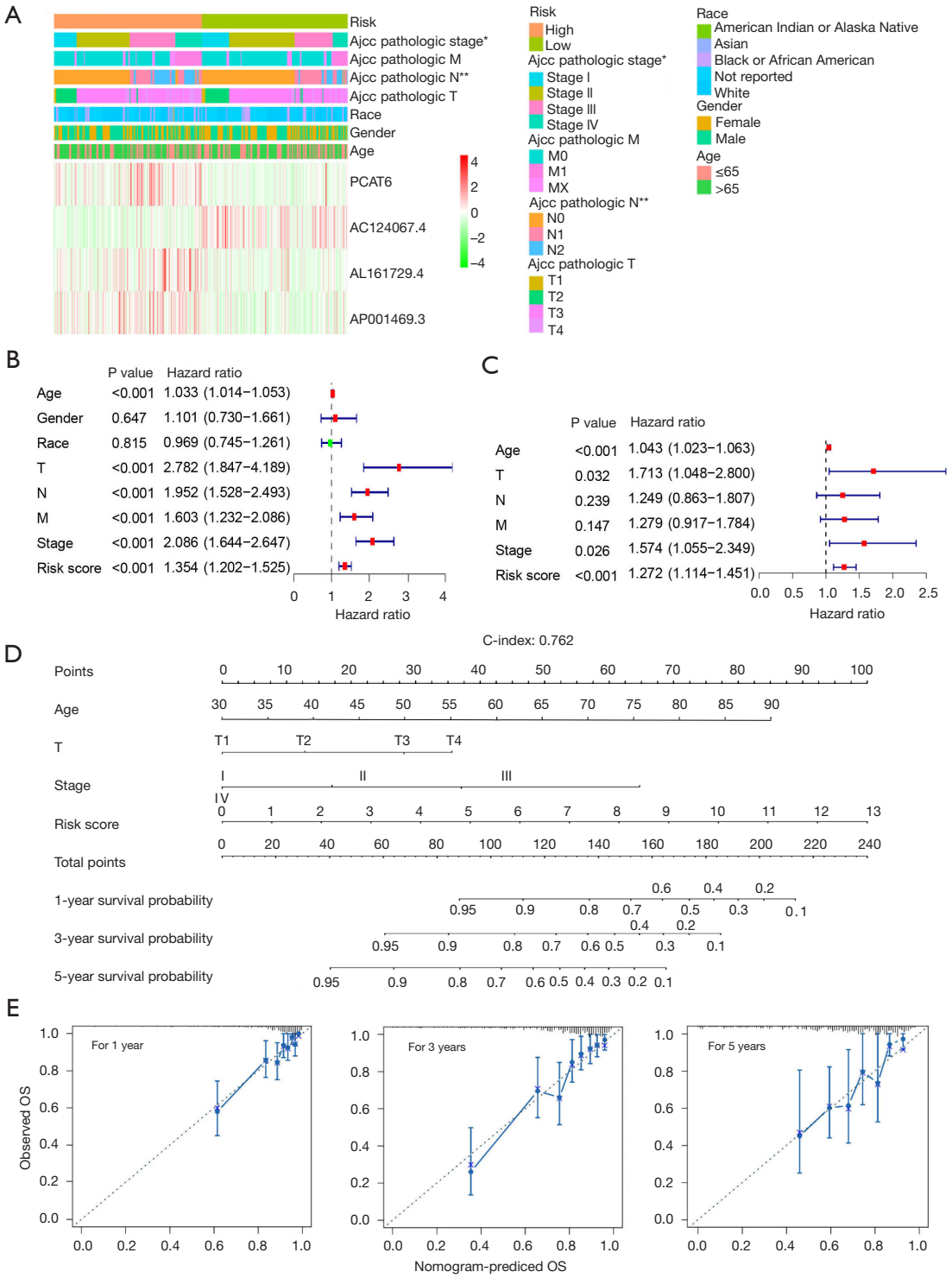

Figure 3 The analysis of the metabolic lncRNAs signature and clinical characteristics in CC. (A) The heatmap revealed the clinicopathologic information of patients in entire database, arranged by high-risk and low-risk group (upper), and differentially expressed four lncRNAs (lower). *P<0.05, **P<0.01. (B) Univariate analysis and (C) multivariate analysis to identify independent prognostic factors of CC patients. (D) Nomogram predicting prognosis of CC patients with 1-, 3- and 5-year OS from entire database. (E) The calibration plot of the nomogram for 1 year, 3 years and 5 years. lncRNAs, long non-coding RNAs; CC, colon cancer; OS, overall survival.

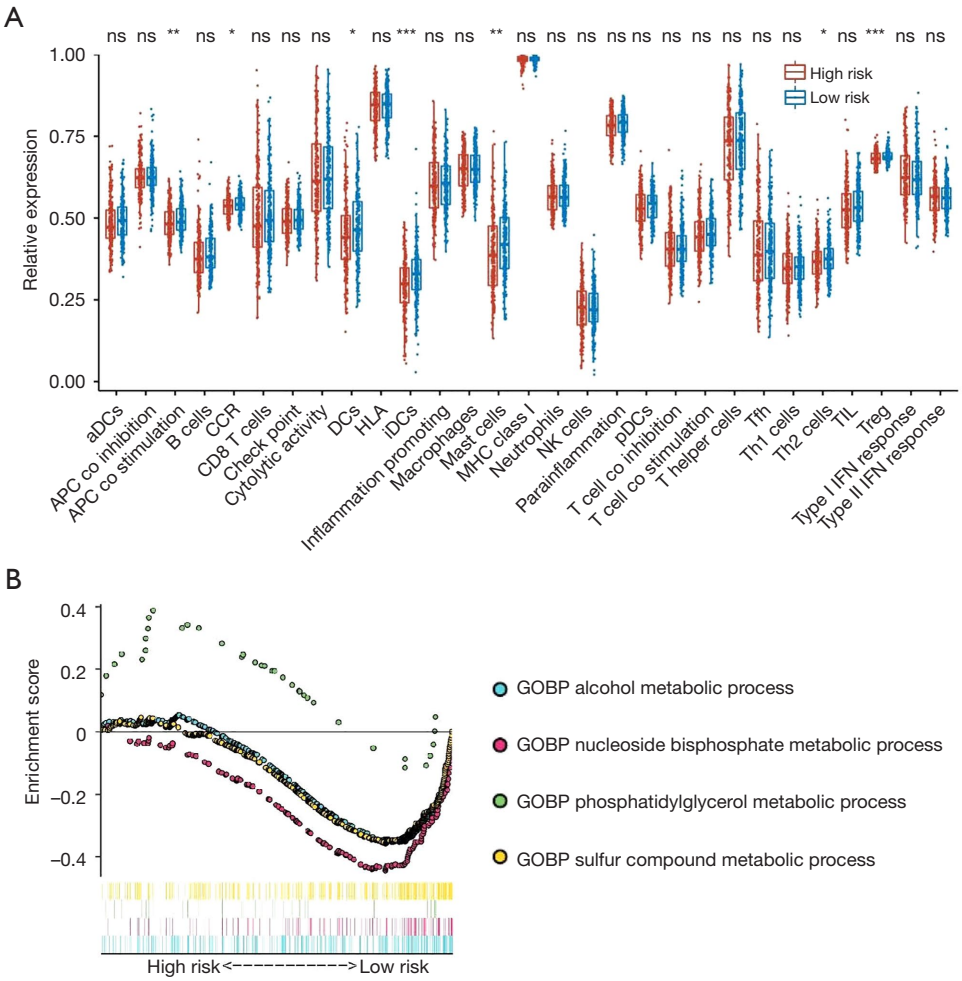

Figure 4 Difference in immune function and GO and KEGG analysis between high- and low risk group according to the metabolic lncRNAs signature. (A) The comparisons of immune function for high- and low risk group. *P<0.05, **P<0.01, ***P<0.001. (B) GO items and KEGG pathways in high- and low risk group were enriched by GSEA, metabolic related pathways was displayed, which at P value <5% in high and the representative pathways with the smallest P values in low risk group. ns, not significant; aDCs, activated dendritic cell; APC, antigen presenting cell; CCR, CC chemokine receptor; DCs, dendritic cells; HLA, human leukocyte antigen; iDCs, immature dendritic cells; MHC class I, major histocompatibility complex I; NK cells, natural killer cell; pDCs, plasmacytoiddendritic cells; Tfh, follicular helper T cell; Th1 cells, T helper 1 cell; Th2 cells, T helper 2 cell; TIL, tumor infiltrating lymphocyte; Treg, regulatory T cells; type I IFN response, type I interferon response; type II IFN response, type II interferon response.

P value <5%. There are 73 metabolic related pathways enriched, and only one biological process (GOBP phosphatidylglycerol metabolic process) significantly enriched at P value <5%, which was displayed in the *Figure 4B*. In terms of low-risk group, 4365/8350 gene sets are upregulated (the supplementary table at https://cdn.amegroups.cn/static/public/tcr-21-2184-2.xlsx), 14 gene sets are significantly enriched at FDR <25%, 116 gene sets are significantly enriched at nominal P value <1%, 385 gene sets are significantly enriched at nominal P value <5%. There are 248 metabolic related pathways enriched, and 51 biological process significantly enriched at P value <5%, then the representative images with the smallest P values (GOBP nucleoside diphosphate metabolic process, GOBP sulfur compound metabolic process, GOBP alcohol metabolic process) are shown in the *Figure 4B*. Meanwhile the genes involved in the above 4 metabolic pathways enrolled in high-risk group and low-risk group

were collected in the supplementary table at https://cdn.amegroups.cn/static/public/tcr-21-2184-3.xlsx.

Discussion

This study was conducted by analysis data of 514 samples (473 cancer samples and 41normal samples) from the TCGA-Colon Adenocarcinoma (TCGA-COAD) database. Differential expression analysis was performed on the metabolic genes and lncRNAs, then the correlation analysis was constructed and 381 differentially expressed lncRNAs related to metabolism genes were obtained. A signature consisting of four prognostic lncRNAs was established as potential prognostic biomarkers for CC patients. We found that the risk score as well as the pathological T, age, stage served as the independent prognostic factors to predict CC prognosis, in addition, there was significant difference in pathological N status between the high- and low-risk groups. Therefore, we explored the immune function and Gene set enrichment analysis (GSEA) involved in the development of CC, and finally, our data analysis showed the further differences between the two groups.

CC is considered a very heterogeneous malignant tumor with different prognosis, to find the prognostic biomarkers is of great significance and currently ongoing. As is known to all, the pathologic T, N, M stage status are associated with CC patient OS. In the near future, small molecular substances, such as ctDNA, snoRNA, mRNA and lncRNA, have emerged as diagnostic biomarkers for cancer detection and monitoring due to the easy access to tissue and liquid biopsy (17). Studies have indicated that the disorder metabolism of glucose and lipid is highly related to the cancer, a study showed that glycometabolism-related gene signature can predict survival in patients with ovarian cancer, another study suggested that a seventeen-gene metabolic signature can predict prognosis with CC (18,19). LncRNAs are key regulators of cellular processes, and associated with tumor progression or suppression of CRC (14). We identified a useful four-metabolism related-lncRNA signature that consists of PCAT6, AP001469.3, AL161729.4 and AC124067.4, and the PCAT6 expression positively correlated with high risk of CC. Previous studies reported that PCAT6 is associated with triple-negative breast cancer, cholangiocarcinoma, bladder cancer and gastrointestinal stromal tumor, etc. (20-22). PCAT6 plays inhibit role in CC cell apoptosis, and targets miR-204 to modulate the chemoresistance to 5-fluorouracil of colorectal cancer cells (23). The present study is consistent with these previous reports.

The metabolism characteristics of tumors are different from those in normal cells, more energy consumption of the tumor cells resulted in the fast metabolism rate than the normal cells. Increasing attention has been given to tumor metabolism in recent studies (24). Researchers has explored the metabolic changes of glucose metabolism in tumors, due to the increased and dependence in anaerobic glycolytic pathway, which has been explained by the "Warburg effect" (25). In addition to glucose metabolism, fatty acid and amino acid metabolism also change in tumor cells. In the present study, alterations in metabolism related pathways displayed in low risk group more frequently, including purine-containing compound metabolic processes, fatty acid metabolic processes, arginine metabolic processes and glucose 6-phosphate etc. This could possibly be because too many alterations of metabolic processes cause disordered metabolism in tumor cells and eventually leads to cellular death.

In current study, immature DCs (iDCs) and regulatory T cell (Treg) were the main significant differences between the high- and low-risk groups, but there were no differences in CD8 T cells and check-point. Recently, immunotherapy has become a mainstream strategy for cancer therapy in clinic, and CD8+ T cells and check-point are considered the main targets for cancer immunotherapy (26). Only a small subset of CRC patients have benefitted from checkpoint blockade immunotherapy, responses to immunotherapy in colorectal cancer have been largely limited to patients with MSI, however, this cancer phenotype represents <5% of advanced stage CRC (27,28). The reasons why the colorectal cancer patients responding to immunotherapy is not well remain unknown, the results of our study may provide a new insight into the possible reason for this.

Phosphatidylglycerol is the major classes of glycerophospholipids comprise, whose metabolic process upregulated in high-risk group in our study. Due to the glycerophospholipids is a major component of cellular membranes, tumor cells often upregulate glycerophospholipid biosynthesis to meet the higher demand for phospholipids, which associated with cell proliferation (29); aside from this, glycerophospholipid supports the increasing energy requirement of tumor cells (30). We speculate that phosphatidylglycerol synthesis is necessary for tumor development and progression, but there is few related study at present, and further research is needed to prove this. Interestingly, alcohol metabolic process unregulated in low-risk group, which mean alcohol-metabolizing genes

would play anti-tumor effects. Alcohol is recognized as a risk-factor for tumors (31), increases alcohol metabolism could increase clearance of alcohol and lower blood alcohol concentrations. In addition to this, alcohol-induced metabolic alterations lead to increased fatty acid synthesis and decrease fat metabolism, which have also been linked to tumorigenesis.

Conclusions

In conclusion, in the current study, we developed a four-metabolism related-lncRNA signature for prognostic prediction of CC based on a computational machine learning framework, which provides a reference for the clinical diagnosis and treatment of CC. The four lncRNAs of this signature may provide some insights for further trials, experiments could explore their potential mechanism in the development and growth regulation of CC.

Acknowledgments

Funding: This work was supported by the National Natural Science Foundation of China (No. 81672104); Shandong Provincial Natural Science Foundation (No. ZR2019LZL016); National Natural Science Foundation of China (No. 82003233); Jinan Science and Technology Plan Project (No. 202019163); Science and Technology Project of traditional Chinese Medicine in Shandong Province (No. 2021M013).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://tcr.amegroups.com/article/view/10.21037/tcr-21-2184/rc

Peer Review File: Available at https://tcr.amegroups.com/article/view/10.21037/tcr-21-2184/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tcr.amegroups.com/article/view/10.21037/tcr-21-2184/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This study was also approved by the ethics committee of Shandong First Medical University and Shandong Academy of Medical (No. 2022001009) and individual consent for this retrospective analysis was waived.

References

1. Siegel RL, Miller KD, Goding Sauer A, et al. Colorectal cancer statistics, 2020. CA Cancer J Clin 2020;70:145-64.
2. Li H, Courtois ET, Sengupta D, et al. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. Nat Genet 2017;49:708-18.
3. Inoue Y, Toiyama Y, Yokoe T, et al. Direct evidence that heterogeneity necessitates and limits the use of multidrug chemotherapy in colon cancer. Mol Med Rep 2008;1:531-5.
4. Alwers E, Jansen L, Bläker H, et al. Microsatellite instability and survival after adjuvant chemotherapy among stage II and III colon cancer patients: results from a population-based study. Mol Oncol 2020;14:363-72.
5. Di Fiore F, Blanchard F, Charbonnier F, et al. Clinical relevance of KRAS mutation detection in metastatic colorectal cancer treated by Cetuximab plus chemotherapy. Br J Cancer 2007;96:1166-9.
6. Helmlinger G, Sckell A, Dellian M, et al. Acid production in glycolysis-impaired tumors provides new insights into tumor metabolism. Clin Cancer Res 2002;8:1284-91.
7. Kato Y, Ozawa S, Miyamoto C, et al. Acidic extracellular microenvironment and cancer. Cancer Cell Int 2013;13:89.
8. Christopherson RI, Lyons SD, Wilson PK. Inhibitors of de novo nucleotide biosynthesis as drugs. Acc Chem Res 2002;35:961-71.
9. Wintzerith M, Ciesielski-Treska J, Dierich A, et al. Comparative investigation of free nucleotides in two neuroblastoma clonal cell lines. J Neurochem 1976;26:205-7.

10. Sulciner ML, Gartung A, Gilligan MM, et al. Targeting lipid mediators in cancer biology. Cancer Metastasis Rev 2018;37:557-72.
11. Janakiram NB, Rao CV. The role of inflammation in colon cancer. Adv Exp Med Biol 2014;816:25-52.
12. Rao CV, Sanghera S, Zhang Y, et al. Systemic Chromosome Instability Resulted in Colonic Transcriptomic Changes in Metabolic, Proliferation, and Stem Cell Regulators in Sgo1-/+ Mice. Cancer Res 2016;76:630-42.
13. Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. Genes Dev 2009;23:1494-504.
14. Bergmann JH, Spector DL. Long non-coding RNAs: modulators of nuclear structure and function. Curr Opin Cell Biol 2014;26:10-8.
15. Wang Z, Yang B, Zhang M, et al. lncRNA Epigenetic Landscape Analysis Identifies EPIC1 as an Oncogenic lncRNA that Interacts with MYC and Promotes Cell-Cycle Progression in Cancer. Cancer Cell 2018;33:706-20.e9.
16. He Z, Wang C, Xue H, et al. Identification of a Metabolism-Related Risk Signature Associated With Clinical Prognosis in Glioblastoma Using Integrated Bioinformatic Analysis. Front Oncol 2020;10:1631.
17. Zhao Y, Yan Y, Ma R, et al. Expression signature of six-snoRNA serves as novel non-invasive biomarker for diagnosis and prognosis prediction of renal clear cell carcinoma. J Cell Mol Med 2020;24:2215-28.
18. Luo D, Shan Z, Liu Q, et al. A Novel Seventeen-Gene Metabolic Signature for Predicting Prognosis in Colon Cancer. Biomed Res Int 2020;2020:4845360.
19. Liu L, Cai L, Liu C, et al. Construction and Validation of a Novel Glycometabolism-Related Gene Signature Predicting Survival in Patients With Ovarian Cancer. Front Genet 2020;11:585259.
20. Shi R, Wu P, Liu M, et al. Knockdown of lncRNA PCAT6 Enhances Radiosensitivity in Triple-Negative Breast Cancer Cells by Regulating miR-185-5p/TPD52 Axis. Onco Targets Ther 2020;13:3025-37.
21. Bai F, Zhang N, Fang W, et al. PCAT6 mediates cellular biological functions in gastro-intestinal stromal tumor via upregulation of PRDX5 and activation of Wnt pathway. Mol Carcinog 2020;59:661-9.
22. Wan L, Zhang L, Fan K, et al. Knockdown of Long Noncoding RNA PCAT6 Inhibits Proliferation and Invasion in Lung Cancer Cells. Oncol Res 2016;24:161-70.
23. Wu H, Zou Q, He H, et al. Long non-coding RNA PCAT6 targets miR-204 to modulate the chemoresistance of colorectal cancer cells to 5-fluorouracil-based treatment through HMGA2 signaling. Cancer Med 2019;8:2484-95.
24. Mayers JR, Vander Heiden MG. Famine versus feast: understanding the metabolism of tumors in vivo. Trends Biochem Sci 2015;40:130-40.
25. Liberti MV, Locasale JW. The Warburg Effect: How Does it Benefit Cancer Cells? Trends Biochem Sci 2016;41:211-8.
26. Kallies A, Zehn D, Utzschneider DT. Precursor exhausted T cells: key to successful immunotherapy? Nat Rev Immunol 2020;20:128-36.
27. Vilar E, Gruber SB. Microsatellite instability in colorectal cancer-the stable evidence. Nat Rev Clin Oncol 2010;7:153-62.
28. Iacopetta B, Grieu F, Amanuel B. Microsatellite instability in colorectal cancer. Asia Pac J Clin Oncol 2010;6:260-9.
29. Hishikawa D, Hashidate T, Shimizu T, et al. Diversity and function of membrane glycerophospholipids generated by the remodeling pathway in mammalian cells. J Lipid Res 2014;55:799-807.
30. Peck B, Schug ZT, Zhang Q, et al. Inhibition of fatty acid desaturation is detrimental to cancer cell survival in metabolically compromised environments. Cancer Metab 2016;4:6.
31. Kim W, Jeong D, Chung J, et al. Development of colorectal cancer predicts increased risk of subsequent hepatocellular carcinoma in patients with alcoholic liver disease: case-control and cohort study. Sci Rep 2019;9:3236.

Cite this article as: Chen G, Dong Y, Shi W, Li L, Zhu W, Xie L, Song X. Construction and validation of four-metabolism related-long non-coding RNAs as potential signature in prognosis of colon cancer. Transl Cancer Res 2022;11(4):725-735. doi: 10.21037/tcr-21-2184

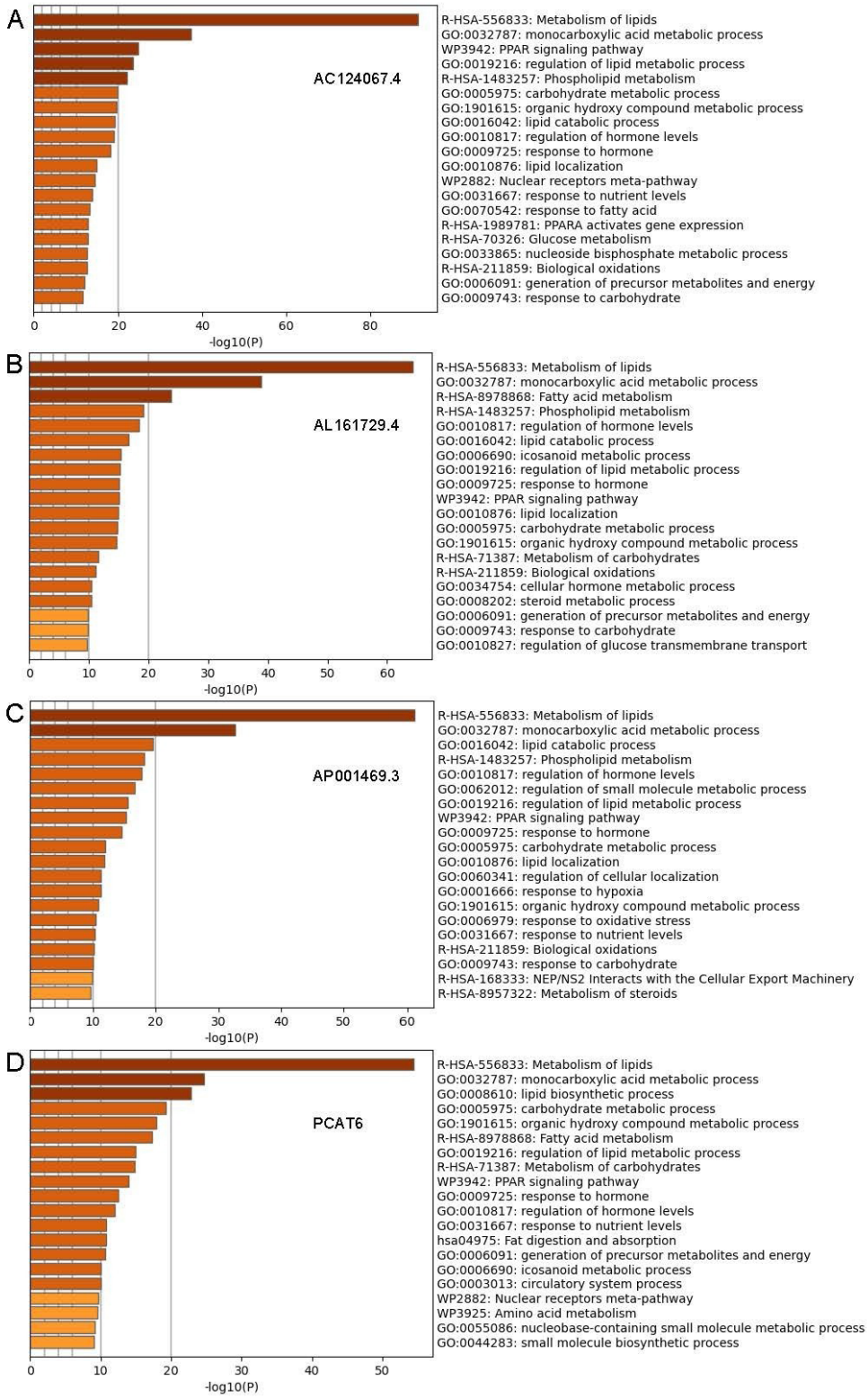

Figure S1 Functional analysis of gene list that correlated with the four lncRNAs respectively. (A) Bar graph of enriched terms across input genes lists that correlated with AC124067.4, colored by P values. (B) Bar graph of enriched terms across input genes that correlated with AL161729.4, colored by P values. (C) Bar graph of enriched terms across input genes that correlated with AP001469.3, colored by P values. (D) Bar graph of enriched terms across input genes that correlated with PCAT6, colored by P values. lncRNAs, long non-coding RNAs.

https://dx.doi.org/10.21037/tcr-21-2184