Oncology and Translational Medicine
DOI 10.1007/s10330-021-0521-1

February 2022, Vol. 8, No. 1, P43–P54

ORIGINAL ARTICLE

A metabolism-associated gene signature with prognostic value in colorectal cancer*

Lingyan Xiao, Yongbiao Huang, Wan Qin, Chaofan Liu, Hong Qiu, Bo Liu (✉), Xianglin Yuan

Department of Oncology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China

Abstract

Objective In this study, our goal was to explore the role of metabolism-associated genes in colorectal cancer (CRC) and construct a prognostic model for patients with CRC.

Methods Differential expression analysis was conducted using RNA-sequencing data from The Cancer Genome Atlas (TCGA) dataset. Enrichment analyses were performed to determine the function of dysregulated metabolism-associated genes. The protein-protein interaction (PPI) network, Kaplan-Meier curves, and stepwise Cox regression analyses identified key metabolism-associated genes. A prognostic model was constructed using LASSO Cox regression analysis and visualized as a nomogram. Survival analyses were conducted in the TCGA and Expression Omnibus (GEO) cohorts to demonstrate the predictive ability of the model.

Results A total of 332 differentially expressed metabolism-associated genes in CRC were screened from the TCGA cohort. Differentially expressed metabolism-associated genes mainly participate in the metabolism of nucleoside phosphate, ribose phosphate, lipids, and fatty acids. A PPI network was constructed out of 328 key genes. A prognostic model was established based on five prognostic genes (*ALAD*, *CHDH*, *ISYNA1*, *NAT1*, and *P4HA1*) and was demonstrated to predict survival in the TCGA and GEO cohorts accurately.

Conclusion The metabolism-associated prognostic model can predict the survival of patients with CRC. Our work supplements previous work focusing on determining prognostic factors of CRC and lays a foundation for further mechanistic exploration.

Received: 19 September 2021
Revised: 23 December 2021
Accepted: 20 January 2022

Key words: colorectal cancer (CRC); prognostic; metabolism; RNA-seq; The Cancer Genome Atlas (TCGA)

Affecting over one million people globally, colorectal cancer (CRC) is among the top three cancers diagnosed most frequently in men and women [1]. Although numerous novel technologies and strategies for CRC diagnosis and treatment have been developed, approximately 10% of cancer-related deaths are still caused by CRC, and the overall survival of patients with CRC remains poor [2]. Many prognostic factors, such as various gene mutations, non-coding RNAs, expression of PD-L1, the neutrophil-to-lymphocyte ratio, and anatomic stage have been demonstrated to predict the survival of patients with CRC over the past decade [3–4]. However, only a few prognostic factors are effective because of the large extent of heterogeneity in CRC, which calls for identifying other prognostic factors.

Alterations in metabolic activities can help cells obtain and maintain malignant properties, facilitating tumor initiation, growth, or progression. Extensive studies on metabolic alterations in cancer cells began with the observation of the Warburg effect. These studies have highlighted that reprogrammed metabolism is a hallmark of cancer [5–6]. The exploration of cancer metabolism offers a new perspective on tumorigenesis. Furthermore, metabolism-associated genes have been shown to have prognostic value in various tumors. For example, a mutation in the gene coding the metabolic enzyme

✉ Correspondence to: Bo Liu. Email: boliu888@hotmail.com
* Supported by grants from the National Natural Science Foundation of China (No. 81773360 and 81902619) and the Nature Science Foundation of Hubei Province (No. 2020CFB591).
© 2022 Huazhong University of Science and Technology

isocitrate dehydrogenase (*IDH*) may indicate a favorable prognosis for gliomas [7]. The prognostic value of a signature reflecting glucose metabolism has been validated in patients with breast cancer through integrative analysis [8]. High expression levels of genes involved in glycolysis may indicate shorter median survival in patients with pancreatic cancer, but the high expression levels of genes involved in cholesterol synthesis may have the opposite effect [9].

Several studies have shown that metabolism is closely related to colorectal oncogenesis [10–12]. Furthermore, other studies have also identified a few prognostic metabolism-associated genes in colorectal cancer [13]. Nevertheless, since various metabolic alterations, such as the biosynthesis and metabolism of glucose, lipids, amino acids, and triphosadenine, play a role in tumor initiation and progression, the metabolism-associated genes involving in prognosis of CRC patients are far from fully explored. In this study, we analyzed CRC RNA-sequencing data from The Cancer Genome Atlas (TCGA) database from different perspectives and discovered five metabolism-associated genes that are independently related to survival in CRC patients. Additionally, a prognostic model was generated, and its prognostic value was confirmed in GSE39582 and TCGA.

Materials and methods

Data collection and processing

RNA-sequencing data files and corresponding clinical and pathological characteristics of patients with CRC were collected from the TCGA database, including a total of 44 normal samples and 568 tumor samples. Microarray data (GSE39582) with 585 samples from the Gene Expression Omnibus (GEO) database were downloaded as the validation cohort. Patients who were followed up for less than a month were excluded. We obtained metabolism-associated genes from the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway using the Gene Set Enrichment Analysis (GSEA) website (https://www.gsea-msigdb.org). The R package "Limma" was used to conduct differential gene expression analysis (version 3.6.2). Metabolism-associated genes that met the "adjusted *P*-value < 0.05 and |log fold change| > 0.5" thresholds were selected for further analysis. Volcano plots and heatmaps were generated to visualize the differentially expressed genes (DEGs).

Gene Ontology (GO) and KEGG enrichment analyses of DEGs

To gain insight into the possible biological functions of the differentially expressed metabolism-associated genes, the R package "clusterProfiler" was used to perform GO enrichment and KEGG pathway analyses with a threshold of both a *P*- and *Q*-value < 0.05.

Protein-protein interaction (PPI) network construction

Because proteins mediate most of the biological functions, a PPI network was constructed using STRING (http://string-db.org) to elucidate protein interactions. Cytoscape, a visualization tool, was used to construct the PPI network. Proteins that did not interact with any other proteins were considered relatively useless and were removed from the network. The metabolism-associated genes participating in the PPI network were identified as key genes.

Identification and validation of prognostic genes

The log-rank test and univariate Cox regression analysis were conducted to identify candidate prognostic genes from the key genes screened from the PPI network. Multivariate Cox regression analysis was also performed to determine whether the candidate prognostic genes could be independent prognostic indicators. Genes with a *P*-value < 0.05 in all of the above analyses were ultimately considered prognostic genes. Differential expression of these five genes was confirmed from different perspectives. Unpaired samples were discarded, and differential expression analysis was performed between 44 paired tumor and peritumoral tissues for these five genes in the TCGA cohort to avoid the effect caused by the large difference between the number of tumor and normal samples. The prognostic genes were also verified in GSE39582 using GraphPad Prism 7.0 software.

Construction and analysis of the prognostic model

The prognostic metabolism-associated genes identified from the above analyses were analyzed using LASSO Cox regression analysis with the R package "glmnet" to generate the prognostic model. The established model was presented as a formula, and the risk score of each sample was calculated using regression coefficients and mRNA expression levels of the prognostic genes. Patients were assigned to the high- and low-risk groups, with the median risk score used as the classification criterion. Kaplan-Meier survival curves were drawn to compare the outcomes of the high- and low-risk groups. The heat map, survival state diagram, and risk curve were generated according to the risk score. Then, univariate and multivariate Cox proportional regression analyses were conducted to determine the role of the risk score in outcome prediction. The "survminer" and "survival" R packages were utilized to perform the above survival analyses. To evaluate the ability of the model to predict survival, receiver operating characteristic (ROC) curves

were created using the R package "survival ROC". A nomogram was then generated based on the prognostic genes to predict patient survival using the "rms package" in the R software, and calibration curves were used to assess the deviation of predicted from actual survival.

Results

Identification of differentially expressed metabolism-associated genes

The workflow of this study is shown in Fig. 1. The RNA-sequencing data collected from TCGA included 568 tumor samples and 44 adjacent normal samples. After extracting the expression values of 961 metabolism-associated genes, we identified 332 DEGs that contained 160 downregulated genes and 172 upregulated genes (Fig. 2).

Functions of differentially expressed metabolism-associated genes and the PPI network

GO function and KEGG pathway enrichment analyses were performed on dysregulated genes to investigate their biological functions and lay the foundation for further mechanistic exploration. Bar and circle plots were also generated. The top 30 enriched GO terms and pathways are presented. The upregulated genes were mainly related to the biosynthesis and metabolism of nucleoside phosphate, ribose phosphate, and purine (Fig. 3). The downregulated genes were mostly involved in the metabolism of various lipids and acids (Fig. 4). Because the interactions between proteins are essential in most biological functions, a PPI network was constructed to determine the significant metaboslism-associated genes in biological processes. The PPI network comprised 328 nodes and 3574 edges after removing disconnected nodes (Fig. 5). The mean node degree of the network was 22, and the maximum node degree of protein nodes in the network was 85.

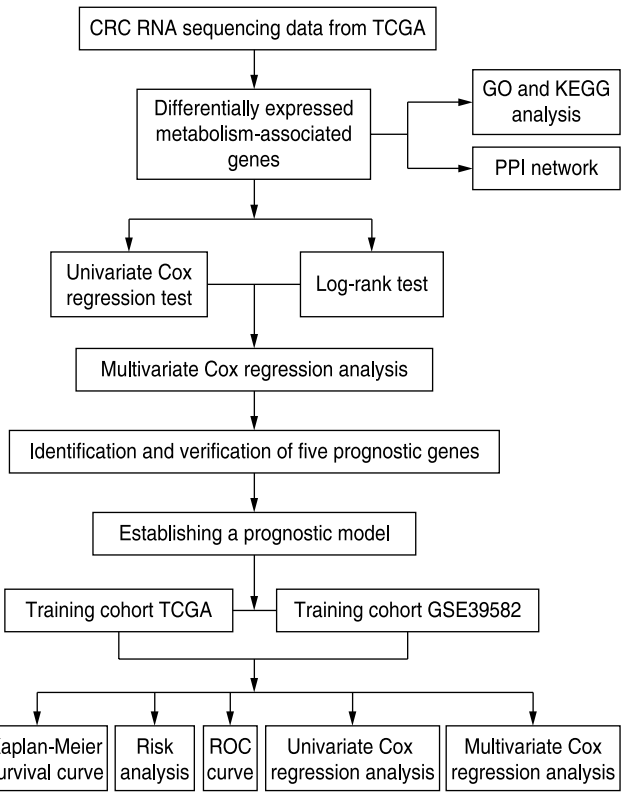


Fig. 1 Workflow for this study

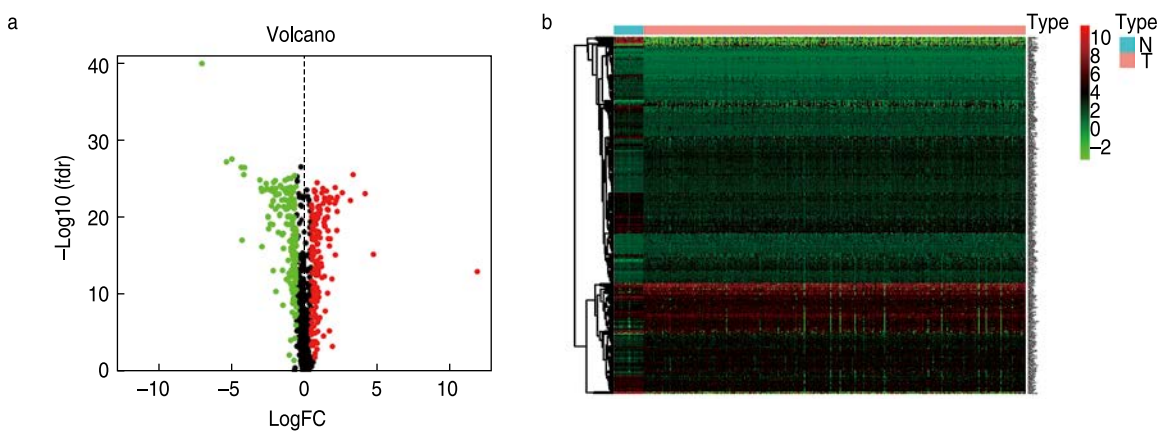


Fig. 2 Analyses of differently expressed genes. (a) The volcano plot of differentially expressed metabolism-associated genes between colorectal cancer and normal tissues in the TCGA database. A total of 160 downregulated genes are displayed in green, and 172 upregulated genes are displayed in red. (b) Heat map of differentially expressed metabolism-associated genes between colorectal cancer and normal tissues in the TCGA database. TCGA = The Cancer Genome Atlas

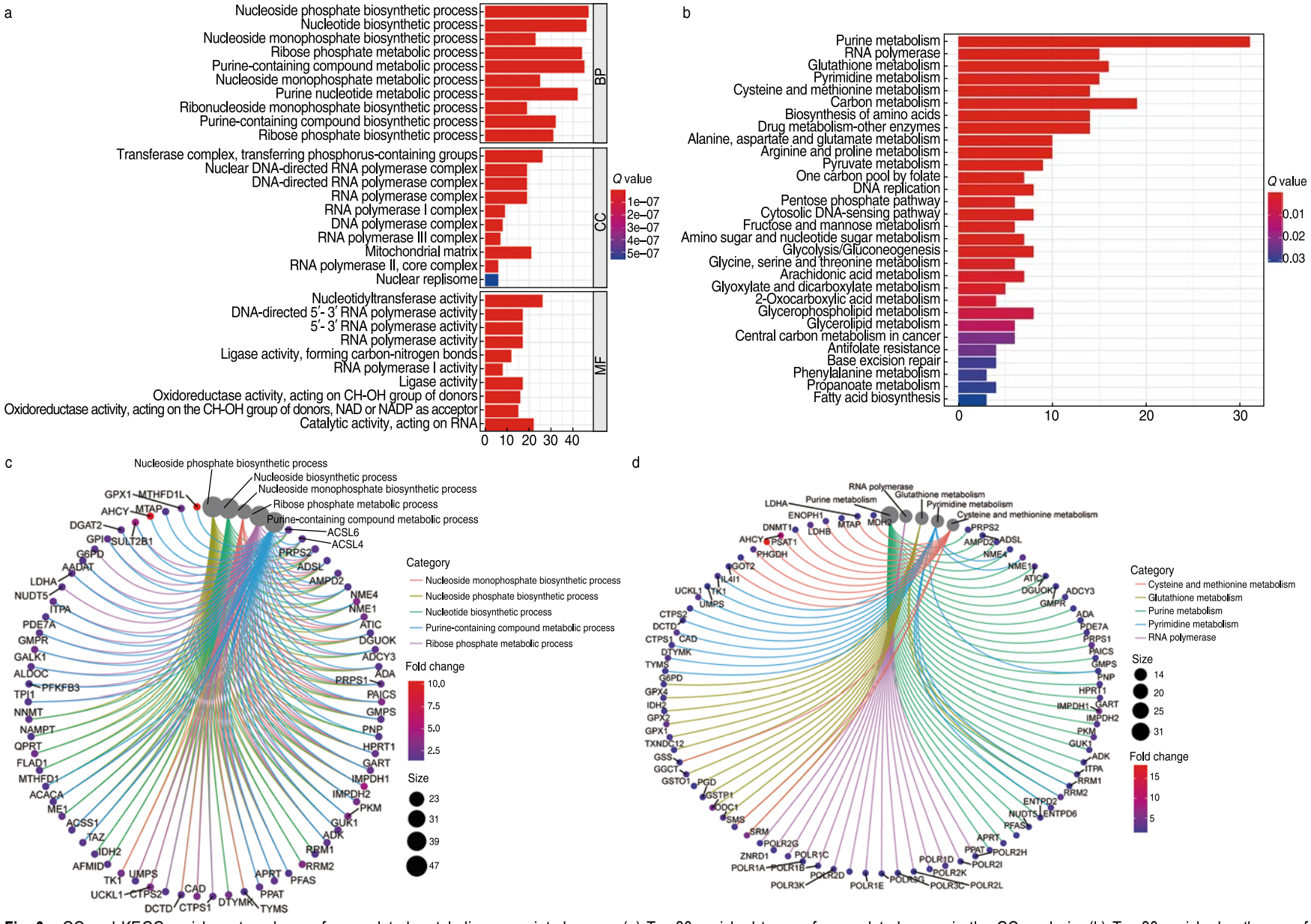


Fig. 3 GO and KEGG enrichment analyses of upregulated metabolism-associated genes. (a) Top 30 enriched terms of upregulated genes in the GO analysis. (b) Top 30 enriched pathways of upregulated genes in the KEGG analysis. (c) Upregulated genes involved in the top five enriched terms in the GO analysis. (d) Upregulated genes involved in the top five enriched pathways in the KEGG analysis. GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes

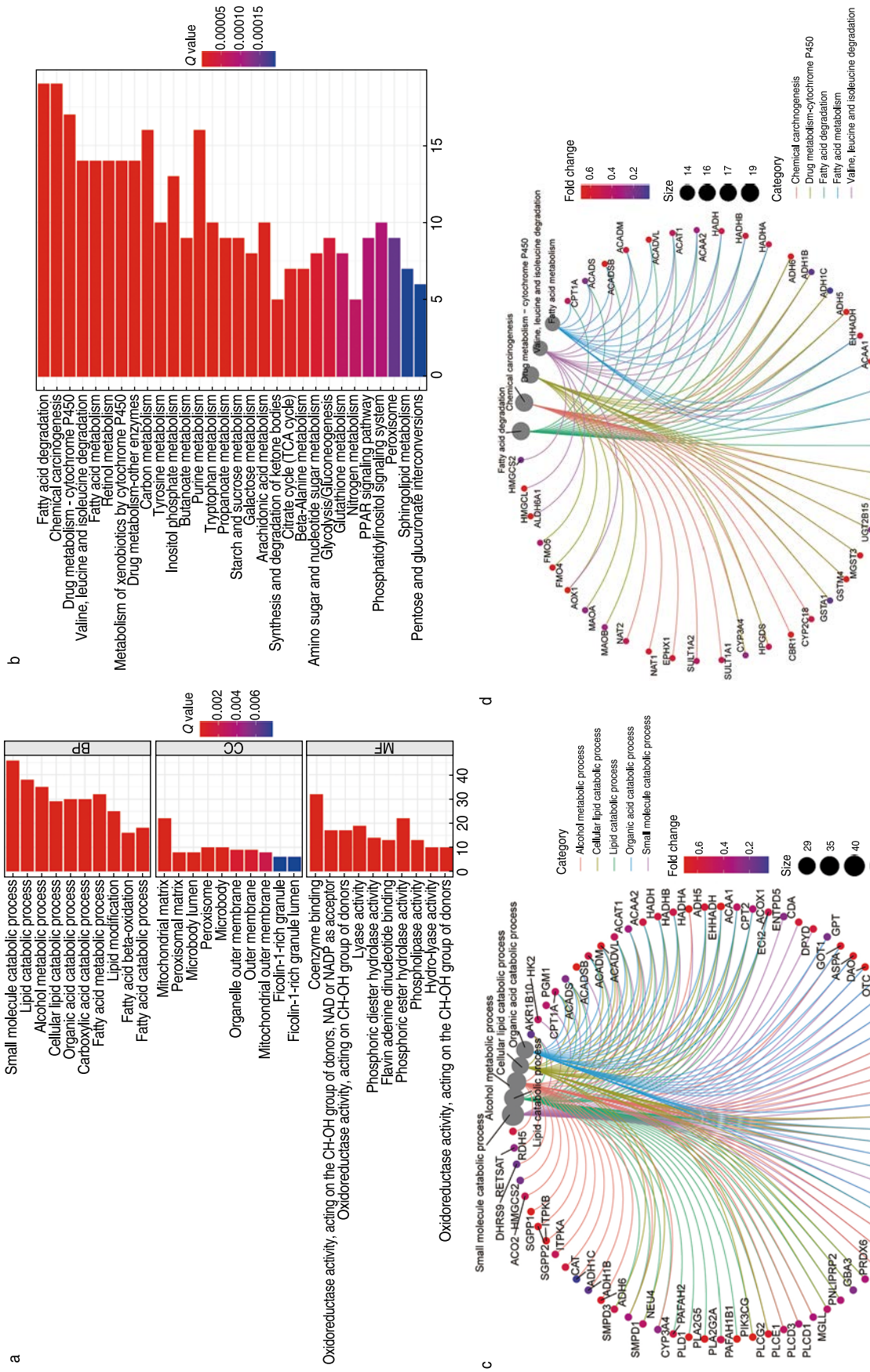


Fig. 4 GO and KEGG enrichment analyses of downregulated metabolism-associated genes. (a) Top 30 enriched terms of downregulated genes in the GO analysis. (b) Top 30 enriched pathways of downregulated genes in the KEGG analysis. (c) Downregulated genes involved in the top five enriched terms in the GO analysis. (d) Downregulated genes involved in the top five enriched pathways in the KEGG analysis. GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes

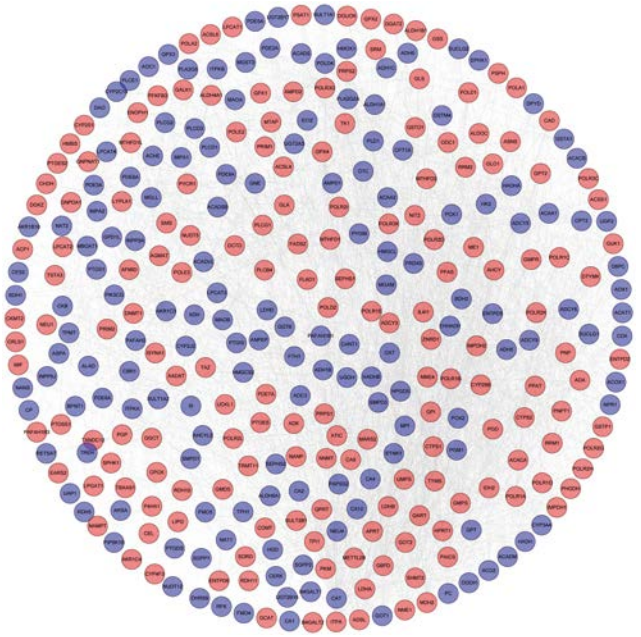


Fig. 5 PPI network analysis. Protein-protein interaction network of differentially expressed metabolism-associated genes. Green dots represent downregulated genes with a fold change of less than 0.5. Red dots represent upregulated genes with fold changes greater than 0.5. PPI: protein-protein interaction

Identification and validation of prognostic metabolism-associated genes

Seventeen genes were acquired using the log-rank test and univariate Cox regression analysis (Fig. 6a). These 17 genes could be significant prognostic factors. However, it is unknown whether their influence on survival is unaffected by other vital characteristics, such as age and stage. Therefore, multivariate Cox regression analysis was conducted, and we obtained five genes (*ALAD*, *CHDH*, *ISYNA1*, *NAT1*, and *P4HA1*) that independently affected overall survival (Fig. 6b). High expression of *CHDH* and *NAT1* was observed to be associated with a lower risk of death with hazard ratios < 1 in both univariate and multivariate Cox analyses, whereas high expression of *ALAD*, *ISYNA1*, and *P4HA1* had negative effect on survival. Thus, we hypothesized that *CHDH* and *NAT1* are tumor suppressor genes, whereas *ALAD*, *P4HA1*, and *ISYNA1* are oncogenes.

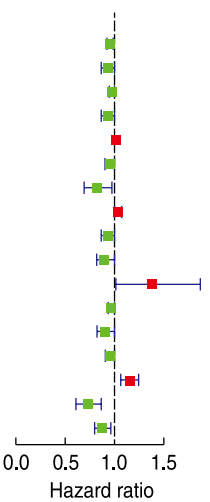
The results of differential expression analysis performed between paired tumor and peritumoral tissues in the TCGA cohort for these five genes confirmed our initial findings (Fig. 7a–7e), demonstrating that the initial differential expression analysis was unaffected by differences in the total sample number between tumor and normal tissues. The expression patterns of these genes were verified using another database. In accordance with the TCGA results, the expression levels of *CHDH*, *P4HA1*, and *ISYNA1* in the validation cohort GSE39582 were significantly elevated in colorectal carcinomas compared to peritumoral tissues, whereas the expression levels of *ALAD* and *NAT1* were lower in tumor tissues (Fig. 7f–7j).

Construction and analysis of prognostic models

A LASSO Cox regression model consisting of regression coefficients and mRNA expression levels of prognostic genes was constructed. The following formula was used to calculate the risk scores: (–0.1025 × Exp *CHDH*) + (0.0242 × Exp *P4HA1*) + (0.1748 × Exp *ALAD*) + (–0.3568 × Exp *NAT1*) + (0.0226 × Exp *ISYNA1*). Patients were assigned

a

	P value	Hazard ratio
ADH5	0.020	0.964 (0.935–0.994)
ACOX1	0.036	0.934 (0.875–0.996)
GOT2	0.024	0.978 (0.959–0.997)
CHDH	0.026	0.933 (0.877–0.992)
P4HA1	0.033	1.018 (1.001–1.035)
GNPNAT1	0.035	0.954 (0.913–0.997)
NANAP	0.021	0.822 (0.695–0.971)
ISYNA1	< 0.001	1.038 (1.018–1.058)
GPD1L	0.046	0.936 (0.877–0.999)
PAFAH2	0.024	0.900 (0.821–0.986)
ACACB	0.038	1.381 (1.017–1.876)
SUCLG2	0.006	0.975 (0.958–0.993)
HMGCL	0.032	0.908 (0.831–0.992)
NAMPT	0.030	0.958 (0.921–0.996)
ALAD	< 0.001	1.156 (1.073–1.245)
NAT1	< 0.001	0.728 (0.613–0.864)
NAT2	0.005	0.877 (0.801–0.961)



b

	P value	Hazard ratio
Age	< 0.001	1.055 (1.031–1.079)
Gender	0.653	0.903 (0.578–1.411)
Stage	0.264	1.507 (0.734–3.096)
T	0.103	1.538 (0.917–2.578)
M	0.062	2.641 (0.952–7.325)
N	0.849	1.045 (0.661–1.654)
ADH5	0.469	1.013 (0.978–1.050)
ACOX1	0.140	0.933 (0.852–1.023)
GOT2	0.563	1.008 (0.982–1.034)
CHDH	< 0.001	0.841 (0.765–0.924)
P4HA1	0.021	1.029 (1.004–1.053)
GNPNAT1	0.512	0.978 (0.914–1.046)
NANP	0.229	0.860 (0.673–1.099)
ISYNA1	0.005	1.033 (1.010–1.056)
GPD1L	0.421	1.046 (0.937–1.169)
PAFAH2	0.285	1.090 (0.931–1.277)
ACACB	0.114	1.405 (0.922–2.140)
SUCLG2	0.246	1.019 (0.987–1.052)
HMGCL	0.060	0.877 (0.766–1.005)
NAMPT	0.213	0.970 (0.924–1.018)
ALAD	0.013	1.129 (1.026–1.241)
NAT1	0.019	0.685 (0.499–0.940)
NAT2	0.867	0.986 (0.835–1.164)

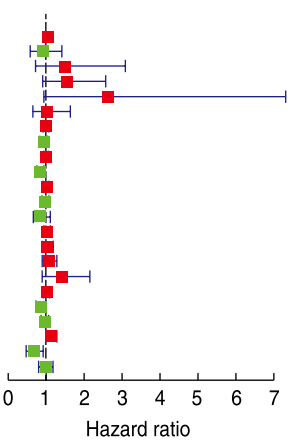


Fig. 6 Univariate Cox regression analysis and multivariate Cox regression analysis of key genes. (a) Seventeen candidate prognostic genes with a *P*-value < 0.05 in both the log-rank test and univariate Cox regression analysis. (b) Results of multivariate Cox regression analysis of 17 candidate prognostic genes. High-risk genes are shown in red, and low-risk genes are in green

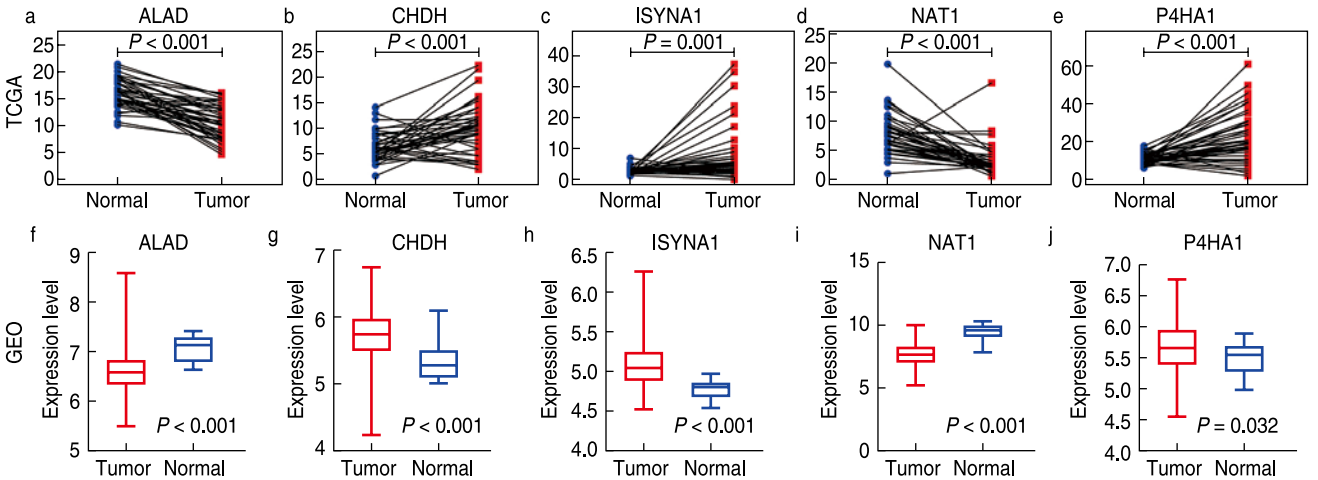


Fig. 7 Validation of five prognostic genes in the TCGA and GEO databases. (a–e) Differential expression of five prognostic genes between paired tumor and normal tissues in the TCGA cohort. (f–j) Differential expression of five prognostic genes in the CRC samples and control samples in GSE39582, and gene expression underwent log2 transformation. GEO: Gene Expression Omnibus database; CRC: colorectal cancer

to the high- or low-risk groups, with the median risk score being the classification criterion. The difference in survival probability between these two groups was statistically significant in both the TCGA ($P < 0.01$; Fig. 8a) and GEO cohorts ($P < 0.01$; Fig. 8b). Patients in the low-risk group were more likely to live longer. In the TCGA cohort, univariate (HR = 3.029, $P < 0.01$; Fig. 8c) and multivariate Cox regression analyses (HR = 2.485, $P < 0.01$; Fig. 8e) showed that the risk score was negatively associated with the overall survival of CRC patients, regardless of confounding factors, such as age, sex, and stage. For the GEO cohort, univariate Cox regression analysis suggested that the overall survival of patients with CRC was significantly related to the risk score (HR = 1.231, $P = 0.026$; Fig. 8d). However, multivariate Cox regression analysis did not yield the same results (HR = 1.076, $P = 0.439$; Fig. 8f). The areas under the ROC curve were 0.744 and 0.569 for the TCGA cohort (Fig. 8g) and GEO cohorts, respectively (Fig. 8h), indicating that the prognostic model was powerful. The difference in the expression levels of the five prognostic genes between the high- and low-risk groups was not statistically significant in the TCGA (Fig. 9a) and GEO cohorts (Fig. 9b). Patients ranked by risk score in the TCGA (Fig. 9c) and GEO (Fig.

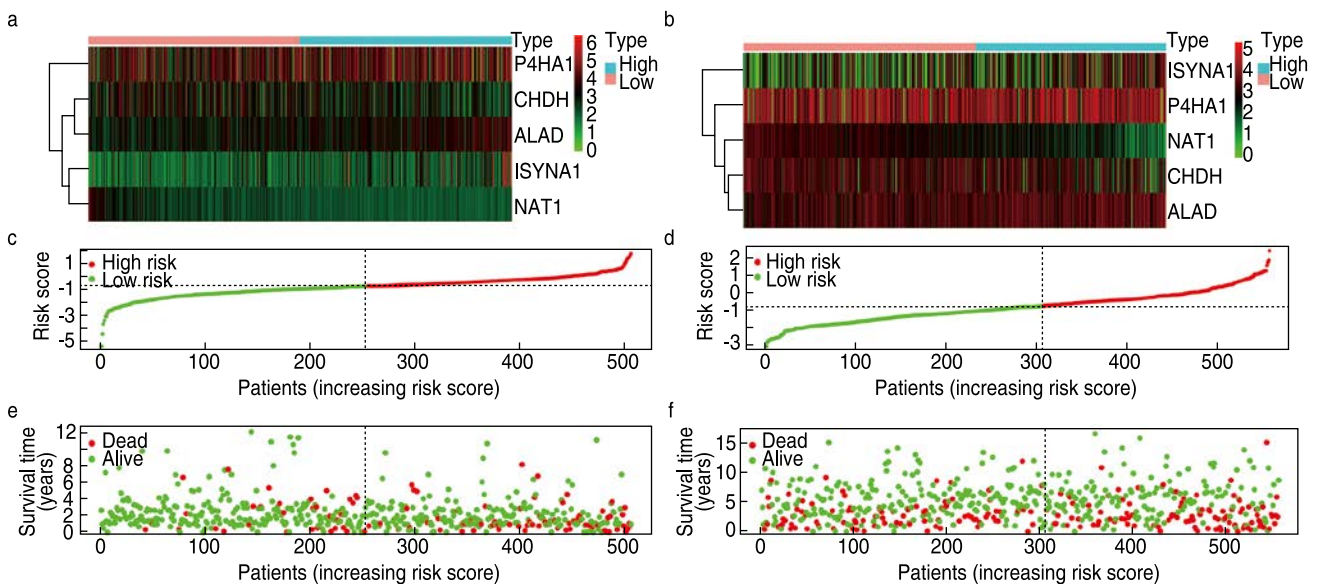


Fig. 9 Risk analyses of the prognostic model. Expression of five prognostic genes in the high- and low-risk groups in the TCGA cohort (a) and GEO cohort (b). Patients ranked by risk scores in the TCGA cohort (c) and GEO cohort (d). Survival status of patients ranked by risk score in the TCGA cohort (e) and GEO cohort (f). TCGA: The Cancer Genome Atlas; GEO: Gene Expression Omnibus database

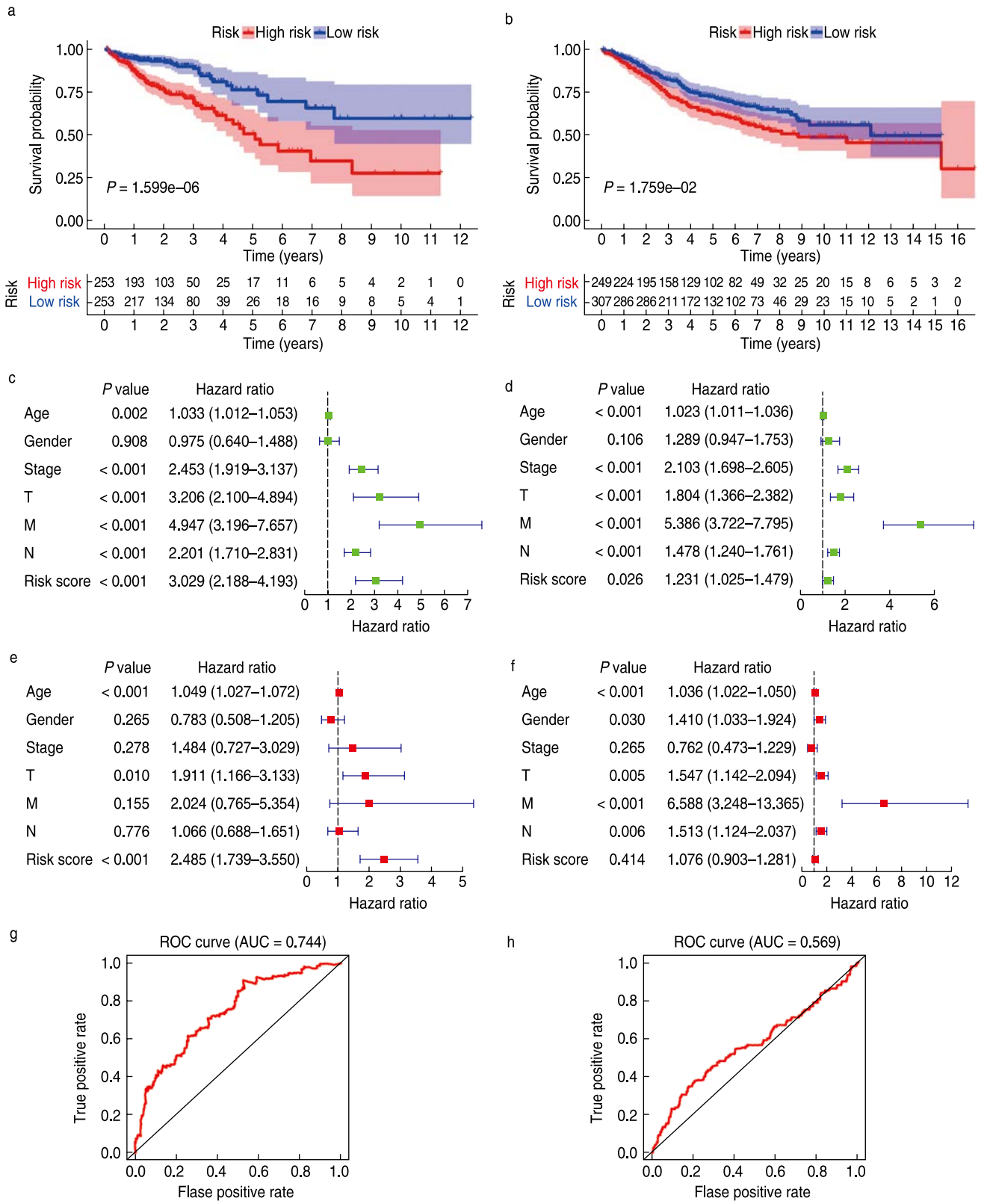


Fig. 8 Construction and verification of the prognostic model. Kaplan-Meier survival curves of the high- and low-risk groups in the TCGA cohort (a) and GEO cohort (b). Univariate Cox regression analysis of risk score and clinicopathological variables in TCGA cohort (c) and GEO cohort (d). Multivariate Cox regression analysis of risk score and clinicopathological variables in the TCGA cohort (e) and GEO cohort (f). ROC curves of the risk score in the TCGA cohort (g) and GEO cohort (h). TCGA: The Cancer Genome Atlas; GEO: Gene Expression Omnibus database

9d) cohorts are displayed. Surviving patients decreased with an increase in the risk score (Fig. 9e and 9f), consistent with the results of the Kaplan-Meier curve and stepwise Cox regression analyses. A nomogram based on the prognostic model was plotted to predict the survival of patients with CRC (Fig. 10a), and calibration curves showed that the predicted survival of the nomogram was consistent with actual survival (Fig. 10b and 10c).

Discussion

CRC accounts for a large porportion of gastrointestinal tumors and poses a huge threat to global health. The overall survival of patients with CRC depends on many risk factors. Recently, numerous prognostic biomarkers have been developed for CRC, but only a few of them have been applied clinically. Therefore, it is necessary to identify more potential prognostic factors. Cancer metabolism is an important segment of the malignant transition. The link between the gut microbiome and colon carcinogenesis may also be mediated by altered metabolism [14]. Numerous studies have confirmed the prognostic value of metabolism-associated genes in various tumors. Therefore, there is a need to explore the metabolism-associated genes that play a role in the outcome of patients with CRC.

In this study, using TCGA data, we conducted an integrative analysis to offer a well-rounded landscape of 961 metabolism-associated genes in CRC. The possible mechanisms underlying oncogenesis were explored using functional enrichment and PPI network analyses. Additionally, we identified five prognostic metabolism-associated genes (*ALAD*, *CHDH*, *ISYNA1*, *NAT1*, and *P4HA1*) through stepwise statistical analyses and

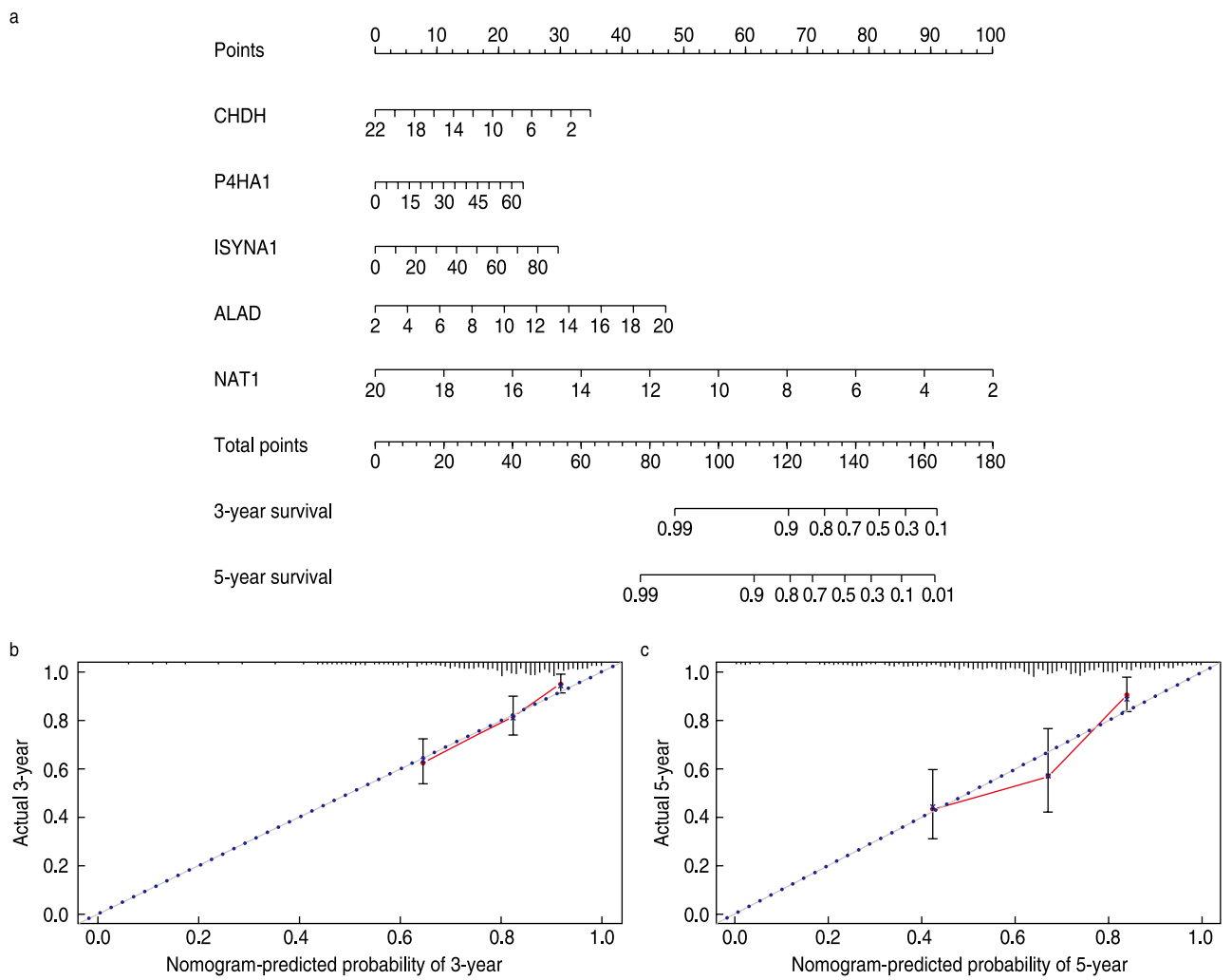


Fig. 10 Nomogram and calibration curves of the prognostic model. (a) Nomogram based on five prognostic genes for predicting the 3-year and 5-year overall survival probability of patients with colorectal cancer. (b) A 3-year calibration plot of the nomogram. (c) A 5-year calibration plot of the nomogram

constructed a prognostic model that performed well in the GEO dataset.

Functional enrichment analysis of differentially expressed metabolism-associated genes revealed that these genes are closely related to the biosynthesis and metabolism of nucleoside phosphate, ribose phosphate, DNA, and RNA polymerase, and the metabolism of lipids and acids. The involvement of nucleotide metabolism has been illustrated in senescence [15], which could determine cancer cell fate. There is increasing evidence that lipid metabolism often affects cancer cells in different ways [16–19]. Cancer cells have an added demand for amino acids and fatty acids to aid their rapid proliferation and increased communication. An earlier study also demonstrated that higher expression levels of genes involved in DNA replication were related to poorer survival in patients with CRC [20]. These findings imply that the functions and pathways discovered in our study are worth exploring.

Given that proteins are the direct mediators of vital biological processes, genes screened from the PPI network are more likely to provide crucial functions and are considered key genes. The node degree of a protein represents the number of proteins with which they interact. None of the five prognostic genes in this study ranked among the genes with the highest node degrees in the PPI network; this could be explained by the paucity of studies on these genes. Thus, future studies must focus on how these genes are involved.

The prognostic genes identified in our study have been shown to impact tumor development in different ways. ALAD is also known as aminolevulinate dehydratase, and its major function is to synthesize heme and inhibit the 26S proteasome. A recent study suggested that ALAD expression level was lower in breast cancer tissues than in normal breast tissues. Increased *ALAD* expression level was correlated with longer disease-free survival of patients with breast cancer, which could be caused by inhibiting the epithelial-mesenchymal transition [21]. Other studies have indicated that genetic variation in *ALAD* is related to the risk of urologic neoplasms and brain tumors [22–23]. In our study, *ALAD* expression negatively affected the overall survival of colorectal cancer patients. Choline dehydrogenase (CHDH), found in the mitochondria, participates in the mitophagy and transformation of betaine aldehyde [24]. A study showed that *ALAD* had positive effects on the overall survival of patients with head and neck squamous cell carcinoma [25]. It could be hypothesized that *ALAD* functions as a tumor suppressor gene. Additionally, *CHDH* variants were correlated with the risk of pancreatic cancer [26].

In our study, inositol-3-phosphate synthase 1 (ISYNA1) was upregulated in colorectal carcinomas compared to that in para-tumor tissues, and its expression was correlated with poorer overall survival. Research evidence has also demonstrated that the mRNA level of ISYNA1 is higher in bladder carcinomas than in para-tumor tissues and that patients with higher expression level of ISYNA1 tended to have higher pathological grades [27]. *ISYNA1* functions as a regulator of proliferation and apoptosis [27]. Additionally, low *ISYNA1* expression level indicated poorer prognoses for patients with pancreatic cancer, which was correlated with p21 inhibition [28]. It is also worth mentioning that *ISYNA1* is associated with the p53 mutation in several tumors, which indicates its significant role in tumorigenesis [29].

NAT1 (N-acetyltransferase 1) can metabolize carcinogens, and its impact on tumor development has been elucidated in numerous studies. Zhao *et al* found that patients with luminal breast cancer had higher expression level of *NAT1* and that *NAT1* could facilitate bone metastasis via a downstream pathway [30]. *NAT1* was also an indicator of response to chemotherapy in patients with breast cancer [31]. Shi *et al* discovered that high expression level of *NAT1* could predict longer overall survival of patients with colon adenocarcinoma through the analyses of the RNA-seq dataset of colon adenocarcinoma (COAD) in TCGA [32]. The positive effect of *NAT1* on the prognosis of patients with CRC was shown by univariate and multivariate Cox regression analyses in our study. Thus, we hypothesized that *NAT1* could function as a tumor suppressor gene in CRC. The functions of *NAT1* in carcinogenesis have also been indicated in bladder cancer and pediatric acute lymphoblastic leukemia [33–34]. However, the underlying mechanisms have not been explored in depth.

As a key gradient of prolyl 4-hydroxylase, P4HA1 (prolyl 4-hydroxylase subunit alpha 1) is essential for collagen synthesis. *P4HA1* is necessary for tumor development. *P4HA1* was demonstrated to regulate the stemness of breast cancer cells and accelerate distant metastasis [35]. Another study on pancreatic cancer showed that the P4HA1 knockdown could reduce stemness in cancer cells and enhance the response to chemotherapy [36]. *P4HA1* has also been shown to be correlated with unfavorable outcomes in patients with high-grade gliomas and head and neck squamous cell carcinoma [37–38]. A recent study demonstrated that the proliferation and invasion of cancer cells could be remarkably promoted by *P4HA1*, and the malignancy of CRC cells could be reduced by *P4HA1* inhibition [39]. However, the prognostic value of *P4HA1* in CRC has not yet been verified. Our study revealed that P4HA1 was upregulated in CRC tissues and that patients with higher P4HA1 expression level had poorer outcomes.

Although all five genes affected overall survival, the effect of a single gene on patient survival was limited. Because it is far from sufficient for one gene to predict patient survival, we constructed a prediction model based on the prognostic genes. Based on the prognostic genes, a

nomogram was used to predict the survival of patients with CRC. The prognostic model we established performed well in both TCGA and GEO cohorts. Overall, we explored the underlying mechanisms of the differentially expressed metabolism-associated genes in CRC, identified five prognostic genes (*ALAD*, *CHDH*, *ISYNA1*, *NAT1*, and *P4HA1*), and constructed a prognostic model via a series of bioinformatics analyses. Although some studies have demonstrated the roles of *CHDH*, *P4HA1*, *ISYNA1*, *ALAD*, and *NAT1* in tumor initiation and progression, few of them have studied the prognostic value of these genes in CRC. The limitation of our study was that all conclusions are drawn from data in public databases, and as such, *in vivo* and *in vitro* experiments were required for further verification and mechanistic exploration. However, our work provides insight into metabolism-associated genes in CRC from multiple perspectives and will lay the foundation for further studies.

Acknowledgments

Not applicable.

Funding

This work was supported by grants from the National Natural Science Foundation of China (No. 81773360 and 81902619) and the Nature Science Foundation of Hubei Province (No. 2020CFB591).

Conflict of interest

The authors indicated no potential conflicts of interest.

Author contributions

Conceptualization: LY Xiao and B Liu; Design and methodology: LY Xiao, YB Huang, W Qin, CF Liu, H Qiu, and B Liu; Data analysis and figure plotting: LY Xiao, YB Huang, and W Qin; Writing-original draft: LY Xiao and B Liu; Writing-review & editing: H Qiu, B Liu, and XL Yuan; Supervision: B Liu and XL Yuan.

Data availability statement

The datasets analyzed in this study are available from TCGA (https://portal.gdc.cancer.gov/) and GEO (http://www.ncbi.nlm.nih.gov/geo).

Ethical approval

Not applicable.

References

1. Dekker E, Tanis PJ, Vleugels JLA, et al. Colorectal cancer. Lancet. 2019;394(10207):1467-1480.
2. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424.
3. Liu SW, Ma R, Cao HX, et al. Prognostic significance of PD-L1 expression in patients with colorectal cancer: a meta-analysis. Oncol Transl Med. 2019;5(2):68-74.
4. Yu QQ, Qiu H, Zhang MS, et al. Prognostic value of neutrophil-to-lymphocyte ratio in metastatic colorectal cancer. Oncol Transl Med. 2016;2(2):75-79.
5. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-674.
6. Pavlova NN, Thompson CB. The emerging hallmarks of cancer metabolism. Cell Metab. 2016;23(1):27-47.
7. Cohen AL, Holmen SL, Colman H. IDH1 and IDH2 mutations in gliomas. Curr Neurol Neurosci Rep. 2013;13(5):345.
8. Kang S, Kim EH, Hwang JE, et al. Prognostic significance of high metabolic activity in breast cancer: PET signature in breast cancer. Biochem Biophys Res Commun. 2019;511(1):185-191.
9. Karasinska JM, Topham JT, Kalloger SE, et al. Altered gene expression along the glycolysis-cholesterol synthesis axis is associated with outcome in pancreatic cancer. Clin Cancer Res. 2020;26(1):135-146.
10. Montrose DC, Zhou XK, Kopelovich L, et al. Metabolic profiling, a noninvasive approach for the detection of experimental colorectal neoplasia. Cancer Prev Res (Phila). 2012;5(12):1358-1367.
11. Brown DG, Rao S, Weir TL, et al. Metabolomics and metabolic pathway networks from human colorectal cancers, adjacent mucosa, and stool. Cancer Metab. 2016;4:11.
12. Denkert C, Budczies J, Weichert W, et al. Metabolite profiling of human colon carcinoma – deregulation of TCA cycle and amino acid turnover. Mol Cancer. 2008;7:72.
13. Yuan Y, Chen J, Wang J, et al. Development and clinical validation of a novel 4-gene prognostic signature predicting survival in colorectal cancer. Front Oncol. 2020;10:595.
14. Datta S, Chowdhury S, Roy HK. Metabolism, microbiome and colorectal cancer. Aging (Albany NY). 2017;9(4):1086-1087.
15. Aird KM, Zhang G, Li H, et al. Suppression of nucleotide metabolism underlies the establishment and maintenance of oncogene-induced senescence. Cell Rep. 2013;3(4):1252-1265.
16. Abramson HN. The lipogenesis pathway as a cancer target. J Med Chem. 2011;54(16):5615-5638.
17. Santos CR, Schulze A. Lipid metabolism in cancer. FEBS J. 2012;279(15):2610-2623.
18. Röhrig F, Schulze A. The multifaceted roles of fatty acid synthesis in cancer. Nat Rev Cancer. 2016;16(11):732-749.
19. Schulze A, Harris AL. How cancer metabolism is tuned for proliferation and vulnerable to disruption. Nature. 2012;491(7424):364-373.
20. Pillaire MJ, Selves J, Gordien K, et al. A 'DNA replication' signature of progression and negative outcome in colorectal cancer. Oncogene. 2010;29(6):876-887.
21. Ge J, Yu Y, Xin F, et al. Downregulation of delta-aminolevulinate dehydratase is associated with poor prognosis in patients with breast cancer. Cancer Sci. 2017;108(4):604-611.
22. Neslund-Dudas C, Levin AM, Rundle A, et al. Case-only gene-environment interaction between ALAD tagSNPs and occupational lead exposure in prostate cancer. Prostate. 2014;74(6):637-646.
23. Rajaraman P, Stewart PA, Samet JM, et al. Lead, genetic susceptibility, and risk of adult brain tumors. Cancer Epidemiol Biomarkers Prev. 2006;15(12):2514-2520.
24. Park S, Choi SG, Yoo SM, et al. Choline dehydrogenase interacts with SQSTM1/p62 to recruit LC3 and stimulate mitophagy. Autophagy. 2014;10(11):1906-1920.
25. Wu ZH, Tang Y, Zhou Y. A metabolic gene signature to predict overall survival in head and neck squamous cell carcinoma. Mediators Inflamm. 2020;2020:6716908.

26. Chittiboyina S, Chen Z, Chiorean EG, et al. The role of the folate pathway in pancreatic cancer risk. PLoS One. 2018;13(2):e0193298.
27. Guo X, Li HH, Hu J, et al. ISYNA1 is overexpressed in bladder carcinoma and regulates cell proliferation and apoptosis. Biochem Biophys Res Commun. 2019;519(2):246-252.
28. Zhou L, Sheng W, Jia C, et al. Musashi2 promotes the progression of pancreatic cancer through a novel ISYNA1-p21/ZEB-1 pathway. J Cell Mol Med. 2020;24(18):10560-10572.
29. Koguchi T, Tanikawa C, Mori J, et al. Regulation of myo-inositol biosynthesis by p53-ISYNA1 pathway. Int J Oncol. 2016;48(6):2415-2424.
30. Zhao C, Cai X, Wang Y, et al. NAT1 promotes osteolytic metastasis in luminal breast cancer by regulating the bone metastatic niche via NF-κB/IL-1B signaling pathway. Am J Cancer Res. 2020;10(8):2464-2479.
31. Minchin RF, Butcher NJ. Trimodal distribution of arylamine N-acetyltransferase 1 mRNA in breast cancer tumors: association with overall survival and drug resistance. BMC Genomics. 2018;19(1):513.
32. Shi C, Xie LY, Tang YP, et al. Hypermethylation of N-Acetyltransferase 1 is a prognostic biomarker in colon adenocarcinoma. Front Genet. 2019;10:1097.
33. Hernández-González O, Ortiz-Zamudio JJ, Rodríguez-Pinal CJ, et al. Genetic polymorphisms of arylamine N-acetyltransferases 1 and 2 and the likelihood of developing pediatric acute lymphoblastic leukemia. Leuk Lymphoma. 2018;59(8):1968-1975.
34. El Kawak M, Dhaini HR, Jabbour ME, et al. Slow N-acetylation as a possible contributor to bladder carcinogenesis. Mol Carcinog. 2020;59(9):1017-1027.
35. Xiong G, Stewart RL, Chen J, et al. Collagen prolyl 4-hydroxylase 1 is essential for HIF-1α stabilization and TNBC chemoresistance. Nat Commun. 2018;9(1):4456.
36. Cao XP, Cao Y, Li WJ, et al. P4HA1/HIF1α feedback loop drives the glycolytic and malignant phenotypes of pancreatic cancer. Biochem Biophys Res Commun. 2019;516(3):606-612.
37. Hu WM, Zhang J, Sun SX, et al. Identification of P4HA1 as a prognostic biomarker for high-grade gliomas. Pathol Res Pract. 2017;213(11):1365-1369.
38. Li Q, Shen Z, Wu Z, et al. High P4HA1 expression is an independent prognostic factor for poor overall survival and recurrent-free survival in head and neck squamous cell carcinoma. J Clin Lab Anal. 2020;34(3):e23107.
39. Agarwal S, Behring M, Kim HG, et al. Targeting P4HA1 with a small molecule inhibitor in a colorectal cancer PDX model. Transl Oncol. 2020;13(4):100754.

DOI 10.1007/s10330-021-0521-1
Cite this article as: Xiao LY, Huang YB, Qin W, et al. A metabolism-associated gene signature with prognostic value in colorectal cancer. Oncol Transl Med. 2022;8(1):43–54.