



OPEN

## Integrated single-cell and bulk RNA sequencing analysis identifies a prognostic signature related to ferroptosis dependence in colorectal cancer

Xiaochen Xu<sup>1,4</sup>, Xinwen Zhang<sup>3,4</sup>, Qiumei Lin<sup>2</sup>, Yuling Qin<sup>2</sup>, Yihao Liu<sup>3</sup> & Weizhong Tang<sup>1</sup>✉

Ferroptosis is an iron-dependent form of cell death induced by lipid oxidation with an essential role in diseases, including cancer. Since prognostic value of ferroptosis-dependent related genes (FDRGs) in colorectal cancer (CRC) remains unclear, we explored the significance of FDRGs in CRC through comprehensive single-cell analysis. We downloaded the GSE161277 dataset for single-cell analyses and calculated the ferroptosis-dependent gene score (FerrScore) for each cell type. According to each cell type-specific median FerrScore, we categorized the cells into low- and high-ferroptosis groups. By analyzing differentially-expressed genes across the two groups, we identified FDRGs. We further screened these prognosis-related genes used to develop a prognostic signature and calculated its correlation with immune infiltration. We also compared immune checkpoint gene efficacy among different risk groups, and qRT-PCR was performed in colorectal normal and cancer cell lines to explore whether the signature genes could be used as clinical prognostic indicators. In total, 523 FDRGs were identified. A prognostic signature including five signature genes was constructed, and patients were divided into two risk groups. The high-risk group had poor survival rates and displayed high levels of immune infiltration. Our newly developed ferroptosis-based prognostic signature possessed a high predictive ability for CRC.

Of the most common cancer types, colorectal cancer (CRC) accounts for 9.4% of all cancer deaths worldwide<sup>1</sup>. Currently, surgery and chemotherapy have significantly improved the survival rates of CRC patients<sup>2</sup>. However, owing to a lack of efficient clinical treatment and prognostic biomarkers, the overall prognosis of CRC is poor. Moreover, because of tumor heterogeneity, the clinical and histopathological features of tumors cannot currently be used to accurately predict the course of CRC. Therefore, it is critical to identify new prognostic factors and treatment targets for CRC.

Over the past few years, tumor heterogeneity has been shown to be a significant challenge in the treatment and prognosis of cancer. Recently, single-cell RNA sequencing (scRNA-seq) has attracted considerable attention. It allows for the genome-wide analysis of individual cells and makes it possible to understand cellular heterogeneity<sup>3,4</sup>. Li et al.<sup>5</sup> compared the intra-tumor cell heterogeneity between carcinoma and normal tissues in CRC using scRNA-seq. Poonpanichakul et al.<sup>6</sup> used a droplet-based scRNA-seq method to profile intra-tumor cell heterogeneity in CRC ascites. However, few studies have been conducted on the cellular heterogeneity of CRC during its evolution from adenoma to carcinoma. In this study, we analyzed the cellular heterogeneity of adenomas and carcinomas by single-cell analysis, and used this information to develop an effective treatment strategy for CRC.

Ferroptosis, an iron-dependent cell death process, is characterized by lipid peroxidation. It is morphologically and mechanistically distinct from the other types of cell death<sup>7</sup>. Increasing evidence suggests that ferroptosis plays a role in various cancers<sup>8</sup>. Lu et al.<sup>9</sup> found that downregulation of KLF2 inhibits ferroptosis by reducing

<sup>1</sup>Department of Gastrointestinal Surgery, Guangxi Medical University Cancer Hospital, Nanning 530021, Guangxi Zhuang Autonomous Region, China. <sup>2</sup>Department of Clinical Laboratory, Guangxi Medical University Cancer Hospital, Nanning 530021, Guangxi Zhuang Autonomous Region, China. <sup>3</sup>Department of Gastrointestinal Surgery, Affiliated Hospital of Guilin Medical University, Guilin 541001, Guangxi Zhuang Autonomous Region, China. <sup>4</sup>These authors contributed equally: Xiaochen Xu and Xinwen Zhang. ✉email: tangweizhong@gxmu.edu.cn

the transcriptional repression of GPX4 and promoting the invasive activity of renal cell carcinoma. Moreover, because iron metabolism and homeostasis are associated with tumor immunity, they also play a significant role in immunity<sup>10</sup>. Wang et al.<sup>11</sup> demonstrated that the activation of CD8<sup>+</sup> T cells could increase ferroptosis and the efficacy of immunotherapy. However, the mechanisms underlying ferroptosis in CRC, and the role that ferroptosis-dependent related genes (FDRGs) of CRC remain unclear. Therefore, it is necessary to understand the pathophysiology and underlying mechanisms of FDRGs in CRC.

In this study, we calculated ferroptosis-dependent gene scores (FerrScores) for different CRC cell types and used them to define FDRGs. We also constructed a prognostic signature for CRC. Our findings may provide a novel therapeutic strategy for CRC.

## Methods

**Data collection.** To compare the cell heterogeneity in colorectal adenoma and carcinoma tissues, we downloaded the GSE161277 CRC sequence library from the Gene Expression Omnibus database (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>) and selected four colorectal adenoma and four colorectal carcinoma samples (GSM4904234, GSM4904235, GSM4904236, GSM4904238, GSM4904239, GSM4904242, GSM4904243, and GSM4904245) for single-cell analysis<sup>12</sup>. We also downloaded the RNA-Seq data and relevant clinical information on CRC from the official website of The Cancer Genome Atlas (TCGA) (<https://portal.gdc.cancer.gov/>). Furthermore, we obtained the validation dataset GSE17538 from the GEO database. A list of ferroptosis-dependent genes was compiled using the FerrDb website (<http://www.zhounan.org/ferrdb/current/>).

**Processing of sc-RNAseq data.** We used the “Seurat” R package (version 4.1.1) and integrated downstream analysis of single-cell transcriptome profiles. The data were quality controlled. Cells with fewer than 300 features and genes expressed in fewer than three cells were excluded. In addition, the proportion of mitochondria was limited to less than 20%. We then normalized the data using the LogNormalization method. We also screened the 2000 highly variable genes with the “FindVariableFeatures” function. Uniform manifold approximation and projection (UMAP) was used for data visualization in 2 dimensions<sup>13</sup>. Subsequently, the “FindNeighbors” function was used for cell clustering analysis. Furthermore, we used the FindAllMarkers function to define genes in each cluster. Finally, the “SingleR” package was employed for cell-type annotation. We also validated the annotation via cell markers from a previous study<sup>14</sup>.

We loaded 259 ferroptosis-dependent genes and calculated FerrScore in each cell using the AddModuleScore function. We also compared the FerrScore between adenoma and carcinoma samples. According to each cell type-specific median FerrScore, we divided these cells into low-ferroptosis and high-ferroptosis cell groups. We also used the FindMarker function to screen out the differentially-expressed genes (DEGs) of the two groups (adjusted  $p < 0.05$ ,  $|\log_{\text{twofold}} \text{change}| > 0.5$ ). We defined these genes as FDRGs.

We explored the cell ratios of the adenoma and carcinoma samples of the low and high-ferroptosis cell groups. We also performed gene set variation analysis (GSVA) to analyze the enrichments of different cell types in adenoma and carcinoma samples based on the “GSVA” R package<sup>15</sup>.

**Processing of TCGA data.** We used the Perl programming language to extract RNA-Seq data from the database. The expression data of the FDRGs were then extracted based on the TCGA data. Patients with short survival times (less than 30 days) or those who died were excluded from the clinical data.

**FDRG prognostic signature construction and validation.** Patients were randomly divided into training and test cohorts in a 1:1 ratio (Table 1). For the training cohort, we performed univariate Cox regression analysis to identify genes associated with prognosis. In order to minimize the potential of overfitting, we employed Least Absolute Shrinkage and Selection Operator (LASSO) Cox proportional hazards regression to evaluate prognostic genes<sup>16</sup>. Subsequently, a stepwise multivariate Cox regression analysis was conducted using the genes identified through LASSO Cox regression. This analysis aimed to ascertain the prognostic significance of specific gene signatures. Finally, a risk model was constructed by combining the mRNA expression of the genes with their respective risk coefficients in a linear fashion.

The median risk score was used to categorize all patients into high- and low-risk categories. Risk scores were calculated by multiplying gene expression levels by the coefficients of the signature genes. Kaplan–Meier survival curves were plotted using the “survival” R package. A receiver operating characteristic (ROC) curve for the signature was calculated and displayed using the R package “time-ROC”. In addition, we performed univariate and multivariate Cox proportional hazards analyses in the TCGA cohort to determine the predictive value of the riskscore when combined with clinical variables such as age, gender, and stage. We further assessed the association between the riskscore and patient survival outcomes. We also examined the differential expression levels of genes within the signature between the different risk groups. Finally, we conducted nomogram with Risk scores and clinicopathological features to determine the efficacy of Overall Survival (OS) rates at 1-, 2-, and 3-year in CRC patients.

**Analysis of functional enrichment.** To determine the biological functions of the signature, we analyzed the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) data using the ClusterProfiler R package<sup>17</sup>. Enrichment was defined as a  $p$ -value of  $< 0.05$ .

**Correlation between immunity infiltration and signatures.** Single-sample gene set enrichment analysis (ssGSEA) was used to detect the level of immune cell infiltration in correlation with the signature.

Covariates	Type	Total	Test	Train	p-value
Age	< =65	196	97	99	0.9234
	>65	242	122	120	
Gender	Female	197	106	91	0.1787
	Male	241	113	128	
Stage	Stage I	78	38	40	0.8611
	Stage II	160	78	82	
	Stage III	119	61	58	
	Stage IV	66	36	30	
	unknow	15	6	9	
T	T1	13	8	5	0.3971
	T2	81	38	43	
	T3	299	146	153	
	T4	45	27	18	
M	M0	329	162	167	0.4415
	M1	65	36	29	
	unknow	44	21	23	
N	N0	254	124	130	0.6727
	N1	108	54	54	
	N2	75	41	34	
	unknow	1	0	1	

**Table 1.** Clinical characteristics of CRC patients involved in the study.

Immune cell infiltration scores for the low- and high-risk groups were calculated using the “GSVA” R package. Moreover, we compared the immune function between different risk groups.

Based on the TCGA dataset, we calculated immune scores, stromal scores, and ESTIMATE scores using the “estimated” R package<sup>18</sup>. Moreover, we compared the risks in the different groups. Finally, the expression of checkpoint genes<sup>19–22</sup> in the two risk groups was compared.

**CRC cell line culture and quantitative real-time polymerase chain reaction (qRT-PCR).** Human normal intestinal epithelial cells (NCHM460) were obtained from IMMOCELL (Xiamen, China), and human colon cancer cell lines (Caco2, HCT15, HCT116, HT29, Lovo, SW480, SW620) were obtained from iCell (Shanghai, China). All these cells were cultured in media containing 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin (P/S) at 37 °C in a humidified atmosphere of 5% CO<sub>2</sub>. The culture media used were DMEM, MEM, 1640, McCoy’s 5A, McCoy’s 5A, F12K, L15 and L15, which were purchased from Gibco BRL in the USA.

Total cellular and tissue RNA was extracted from tissues or cell lines using Total RNA Extraction Reagent (DP451, Tiangen) according to standard protocols. The obtained RNA was then used for cDNA synthesis using cDNA Synthesis Kit (MR05401S, Monad). Gene expression was quantified using SYBR Green Master Mix (MQ10301S, Monad) on a Roche LightCycler 480 and expression levels were calculated using the 2<sup>-ΔΔCT</sup> method.

GAPDH was used as an internal reference for normalization. All primers used for qRT-PCR were synthesized by Wuhan Jinkairui Bioengineering Co Ltd (Wuhan, China). The sequences of the primers used are listed in Table 2.

**Statistical methods.** We analyzed the data using R software (version 4.1.3) and compared the differences between the groups the Wilcoxon test. Spearman’s correlation coefficient was also calculated; *p* values of < 0.05 were considered statistically significant.

Gene	Forward primer	Reverse primer
GAPDH	GAGTCAACGGATTGGTCGT	GACAAGCTTCCCCTTCTCAG
TIMP1	AGTTTTGTGGCTCCCTGGAA	TCCGTCACAAGCAATGAGT
MANF	ATATCGGGCCACAGATGATG	AACTCGGAGCTTCTTCAGGTC
HSPA1A	CCTGTTTGAGGGCATCGACT	TCGTGAATCTGGCCCTTGTC
RPS17	GTTCGCACAAAACCGTGAA	GCTTGTTCGTGTGGAAGTCG
PTMA	CCTGCTAACGGGAATGCTG	CGTCGGTCTTCTGCTTCTG

**Table 2.** The sequences of the qRT-PCR primer used in this study.

## Results

**Single-cell clustering and acquisition of FDRGs.** The overall research design is illustrated in Fig. 1. A total of 24,732 high-quality cells were obtained from the eight samples after filtering and quality control (Supplementary Fig. 1A). There was no significant correlation between the sequencing depth (total number of UniqueMolecular Identifiers, nCount) and mitochondrial percentage (Pearson correlation coefficient  $r=0.18$ ); however, nCount was positively correlated with the number of detected genes (nFeature) ( $r=0.92$ ) (Supplementary Fig. 1B). Subsequently, the top ten highly variable genes out of 2000 were marked in red (Supplementary Fig. 1C). Twenty-three cell clusters were identified and visualized in two dimensions (Supplementary Fig. 2A). The cell identity of each cluster was annotated to six cell types. These cell types were characterized as epithelial cells, monocytes, endothelial cells, B-cells, NK-cells, and T-cells (Fig. 2A).

Bubble plots were used to show the average expression levels of representative marker genes for different cell types in the integrated human CRC data<sup>23</sup> (Supplementary Fig. 2B).

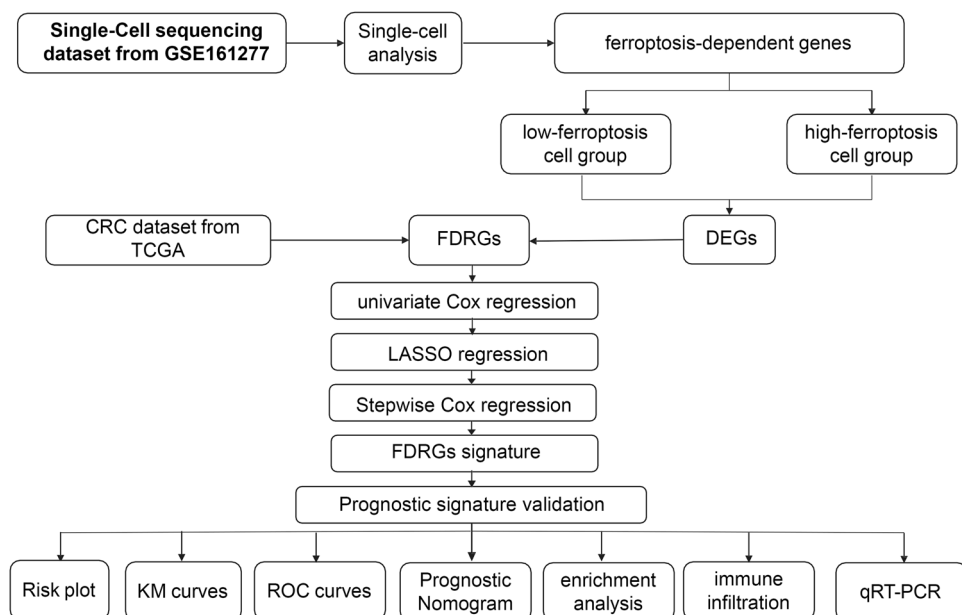
The FerrScores of the different cell types are shown in Fig. 2B. Figure 2C shows the FerrScore of each sample. We also compared the FerrScore among adenoma and carcinoma samples. The differences in FerrScore between adenoma and carcinoma samples were significant. (Fig. 2D,  $***p < 0.001$ ). Figure 2E shows a comparison of the FerrScore for different cell types in adenoma and carcinoma samples. Among these cell types, there were significant differences between adenoma and carcinoma samples, except for endothelial cells ( $***p < 0.001$ ).

For each cell type, we divided these cells into a low-ferroptosis and a high-ferroptosis cell group based on the cell type-specific median FerrScore. Moreover, 523 DEGs were screened in the two groups. We also identified the top five genes with log2FC values (Fig. 2F).

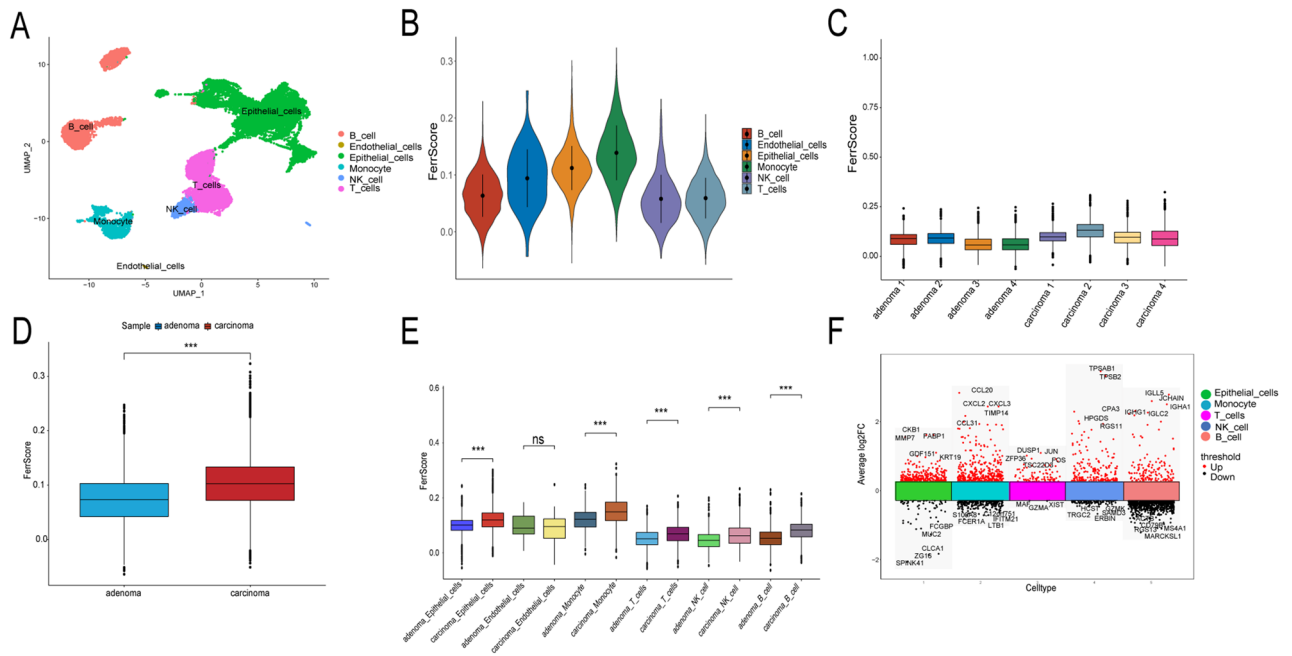
**Cell ratios of the samples and functional enrichment by GSVA.** The cell ratios for each sample are shown in Fig. 3A. We also show the ratios of the different cell types in terms of adenomas vs. carcinomas, in Fig. 3B. The cell ratios of the low-ferroptosis and high-ferroptosis cell groups are displayed in Fig. 3C. Figure 3D shows the results of gene set enrichment in the different cell types between adenoma and carcinoma samples.

**Construction of five-gene prognostic signature.** We developed a prognostic signature based on FDRGs. Using univariate Cox regression analysis, we identified 21 genes associated with OS in patients with CRC (Fig. 4A). Then, LASSO regression analysis was performed on the 21 previously screened genes and obtained 14 genes by the optimal  $\lambda$  (Fig. 4B, C). Finally, we constructed a 5-gene (HSPA1A, MANF, PTMA, RPS17, and TIMP1) signature based on multifactorial stepwise Cox regression analysis. According to the formula, the riskscore =  $HSPA1A * 0.274668948995861 + MANF * (-1.21739069354923) + PTMA * 1.07281342991973 + RPS17 * 0.820155083563555 + TIMP1 * 1.17800493596346$  (Table 3). Figure 4D shows the expression of signature genes in the cell types between adenoma and carcinoma samples.

**Evaluation and validation of prognostic signature.** As we validated the prognostic power of the signature, The Kaplan–Meier curves demonstrated that patients in the low-risk group had a significantly better overall survival (OS) compared to those in the high-risk group in both the TCGA and GEO cohorts (Fig. 5A–B). In the TCGA cohort, the area under the receiver operating characteristic (ROC) curves (AUCs) for 1-, 2-, and 3-year overall survival (OS) were 0.706, 0.729, and 0.707 respectively, as shown in Fig. 5C. The corresponding



**Figure 1.** Data collection and analysis in this study.



**Figure 2.** Single-cell RNA sequencing analysis of 24,732 cells from four colorectal adenomas and four colorectal carcinoma tissues. **(A)** Cells were clustered into six types by a UMAP dimensionality reduction algorithm, and each color represented the annotated phenotype of each cluster. **(B)** The ferroptosis-dependent genes scores (FerrScores) of the six cell types. **(C)** Comparison of FerrScores for each sample. **(D)** Comparison of FerrScores between adenoma and carcinoma groups. **(E)** Comparison of FerrScores for different cell types in adenoma and carcinoma samples. **(F)** Top five differential expressed genes (DEGs) with log<sub>2</sub>FC values. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .

AUCs for the 1-, 2-, and 3-year OS in the GEO cohort were 0.646, 0.665, and 0.631 respectively, as presented in Fig. 5D. To further assess the predictive power of the risk traits, we conducted a comprehensive analysis that involved comparing the risk score distribution, survival time and status, and expression of model genes between the high-risk and low-risk groups Fig. 5E–F. After conducting both univariate and multivariate Cox proportional hazards analyses, we found that age and the riskscore were identified as independent prognostic factors, according to the results presented in Supplementary Fig. 3.

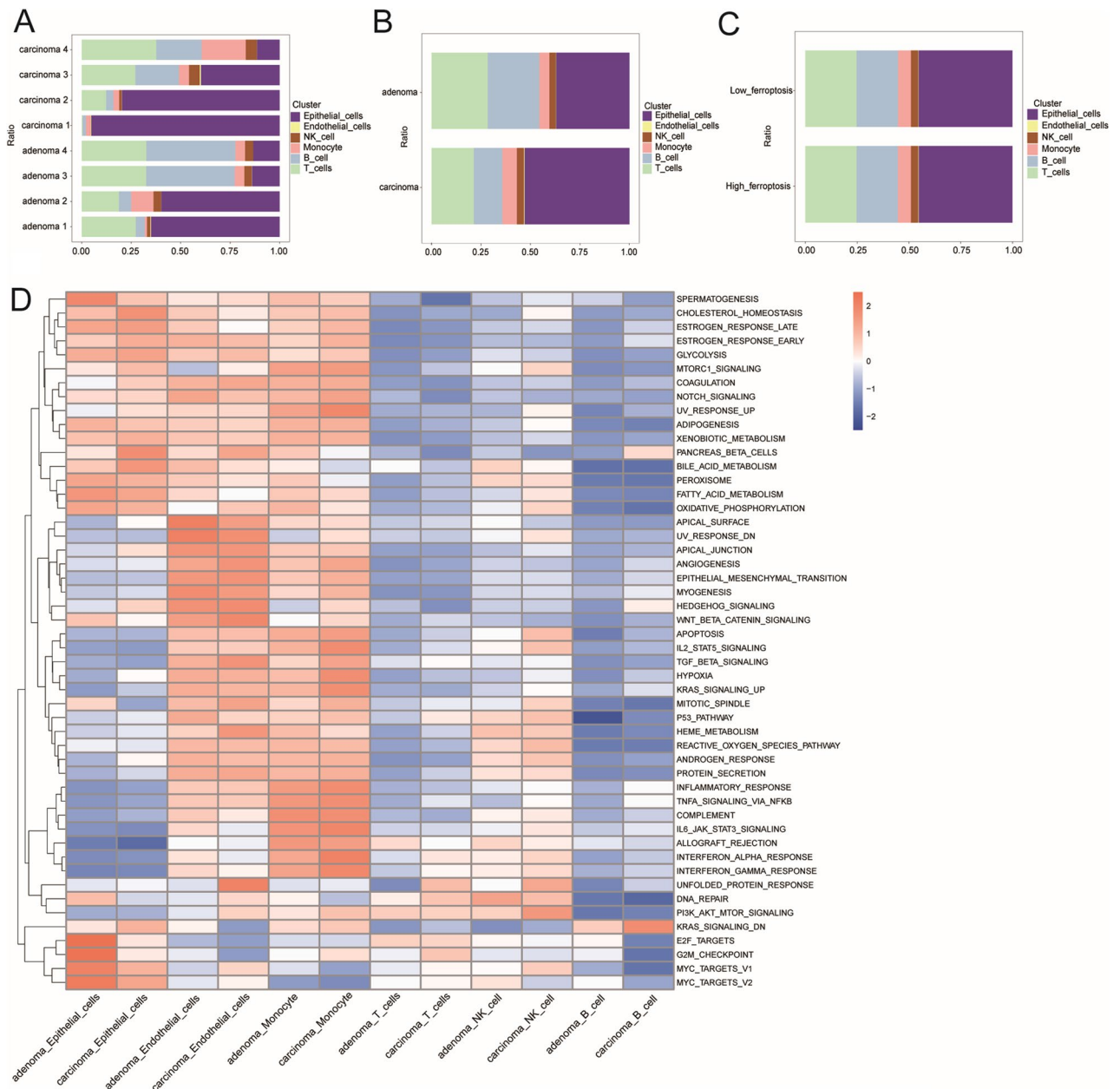
To investigate whether the signature genes have a prognostic significance, we divided them into high and low expression groups based on the optimal cutoff value. Lower expression of these genes displayed better OS, including HSPA1A, TIMP1, and RPS17, while PTMA and MANF's were not significantly associated with OS (Fig. 6A–E).

**Construction and Validation of nomogram.** To evaluate the potential clinical utility of a prognostic signature, a nomogram was constructed based on the clinical characteristics (age, gender, stage, T, N and M) of the patient (Fig. 7A). The nomogram may aid in determining a patient's risk with greater precision, which can ultimately inform and improve future treatment decisions. The calibration curves showed good consistency between the actual and predicted survival rates (Fig. 7B). To further assess the accuracy of the nomogram, prognostic ROC analysis was carried out, which demonstrated superior prognostic performance as compared to other clinical shapes and risk scores. The AUC values for 1-, 2-, and 3-year survival were 0.812, 0.813, and 0.825, respectively, as shown in Fig. 7C–E. These results indicate that the predictive signature has great potential as a biomarker for predicting the prognosis of CRC.

**Comparison of our gene signature with other COAD prognostic models.** In order to assess the superiority of our gene signature over other COAD prognostic models, we conducted a comparison of 10 models across the entire TCGA cohort<sup>24–33</sup>. The 1-year AUC values for all 10 prognostic models were observed, and our findings indicated that our model incorporating the 5 gene signatures had significantly superior predictive power compared to the other 10 prognostic models (Supplementary Fig. 4).

**Functions by KEGG and GO.** The results of the KEGG enrichment analysis showed that the high-risk group expressed traits mainly related to the regulation of cell adhesion, such as extracellular matrix (ECM) receptor interaction and focal adhesion (Fig. 8A). In the low-risk group, a number of metabolic pathways were associated, including aldarate, ascorbate, and butanoate metabolism (Fig. 8B). The KEGG enrichment results have been presented in Supplementary Table 1.

In the high-risk group, GO enrichment analysis revealed that many immune-related pathways were enriched, including adaptive immune responses and B-cell-mediated immunity (Fig. 8C). However, the low-risk group

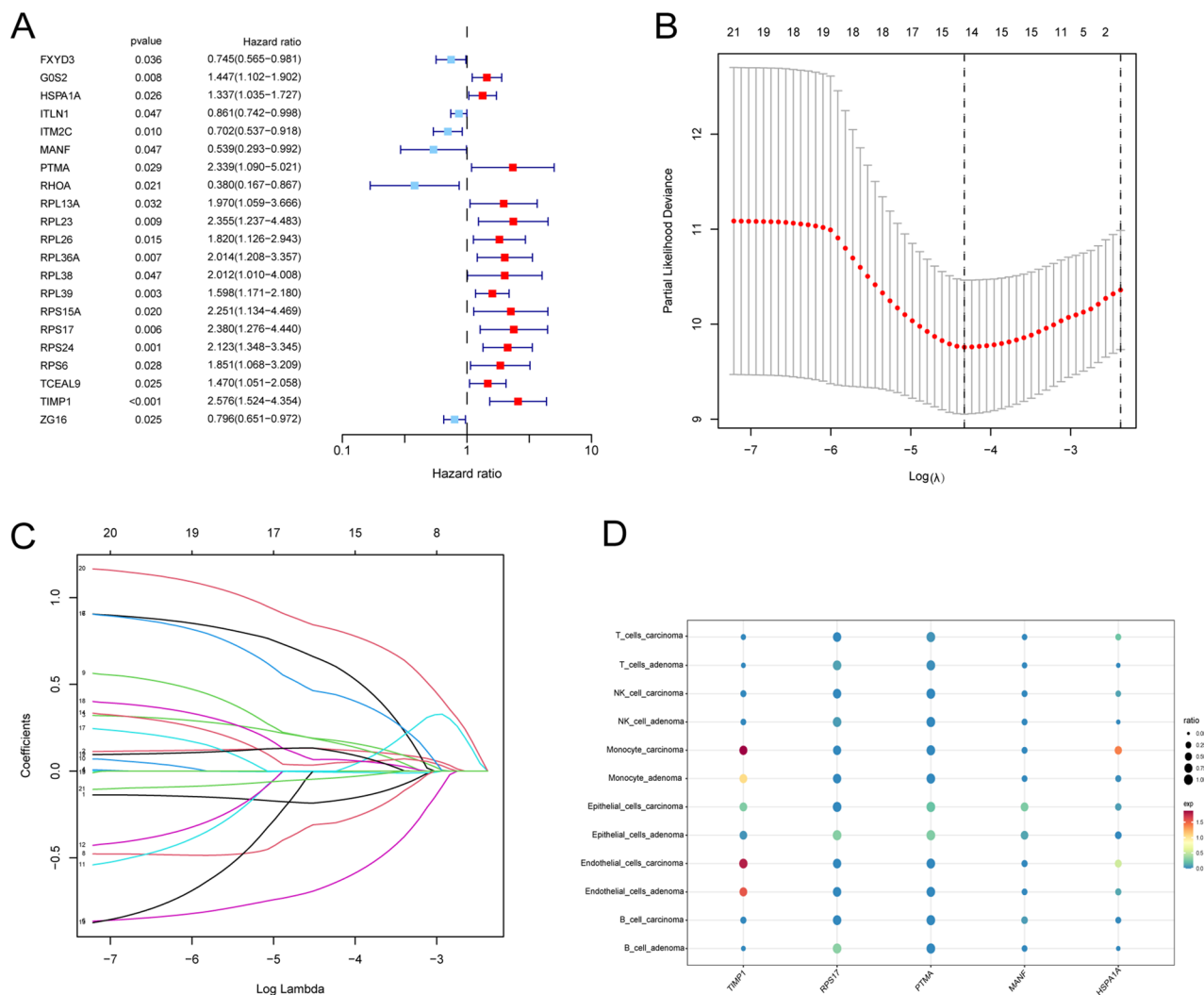


**Figure 3.** Cell ratios of different samples and GSEA analysis. **(A)** Cell ratios of each sample. **(B)** Cell ratios of different cell types in adenoma and carcinoma samples. **(C)** Cell ratios of the high-ferroptosis and low-ferroptosis cell groups. **(D)** Results of GSEA of different cell types in adenoma and carcinoma.

was primarily associated with mRNA splicing traits (Fig. 8D). The GO enrichment results have been presented in Supplementary Table 2.

**Differences in immune infiltration between the two risk groups.** To investigate the immune infiltration levels of different risk groups, we compiled a list of immune infiltration scores for each sample in the TCGA dataset, which is presented in Supplementary Table 3. According to the ssGSEA results, the high-risk group showed more immune infiltration and activated immune function (Fig. 9A–B). Compared to the low-risk group, the immune, stromal, and ESTIMATE scores of the high-risk group were significantly higher (Fig. 9C). In addition, immune checkpoint genes in the high-risk group were significantly upregulated (Fig. 9D). (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ ).

**qRT-PCR assay in colorectal cell models.** As shown in Fig. 6A–E, High expression of HSPA1A, TIMP1, and RPS17 were positively associated with poor prognosis in CRC, while PTMA and MANF's were not significantly associated with prognosis. As shown in Fig. 10A–E, the mRNA expression levels of HSPA1A, TIMP1, RPS17, PTMA, and MANF in CRC cells were significantly elevated compared with normal intestinal epithelial cells. Among them, HSPA1A, TIMP1, and RPS17 were consistent with the clinical prognostic model we con-



**Figure 4.** Construction of a prognostic signature. (A) Univariate Cox analysis of the TCGA cohort. (B–C) Using least absolute shrinkage and selection operator (LASSO) regression, a signature was constructed based on the optimum  $\lambda$ . (D) Bubble plot visualizing the expression of signature genes in cell types between adenoma and carcinoma samples.

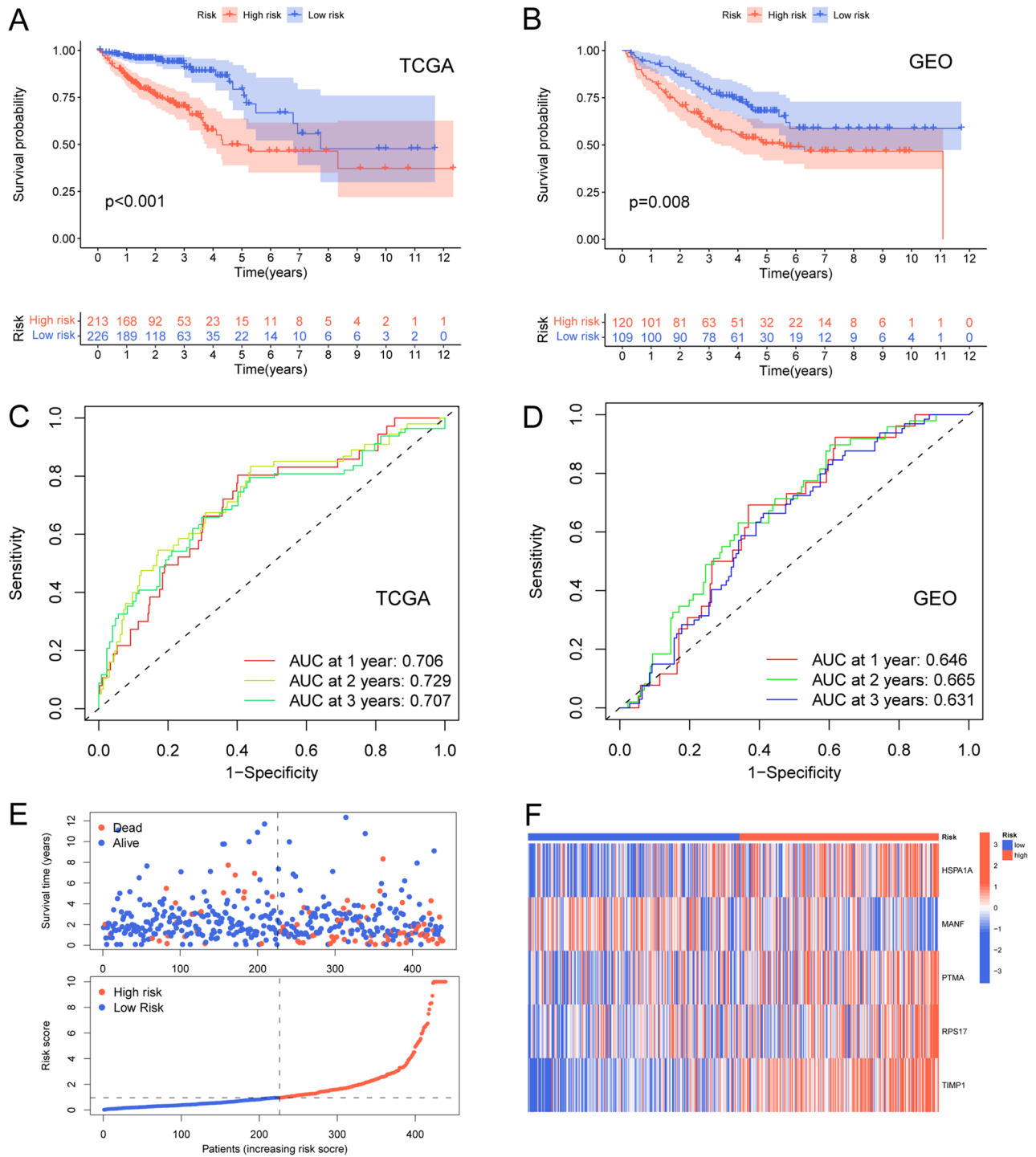
No	Gene name	Coef
1	T1MP1	1.17800493596346
2	PTMA	1.07281342991973
3	RPS17	0.820155083563555
4	HSPA1A	0.274668948995861
5	MANF	-1.21739069354923

**Table 3.** The genes involved in the signature and their coefficients.

structed, and PTMA and MANF could not be used as prognostic signature genes, which could be related to our cell specificity or insufficient sample size ( $*p < 0.05$ ,  $**p < 0.01$ ,  $***p < 0.001$ ,  $****p < 0.0001$ ).

## Discussion

With continuous in-depth research on ferroptosis, extensive experimental research data have shown that ferroptosis plays a vital role in tumor development. For instance, the inhibition of STAT3-ferroptosis inhibits tumor growth and alleviates chemotherapy resistance in gastric cancer<sup>34</sup> Moreover, studies have suggested that cisplatin was an inducer for both ferroptosis and apoptosis in lung cancer cells<sup>35</sup>. However, few studies have examined ferroptosis in patients with CRC. The possible regulatory genes and markers involved in ferroptosis have not

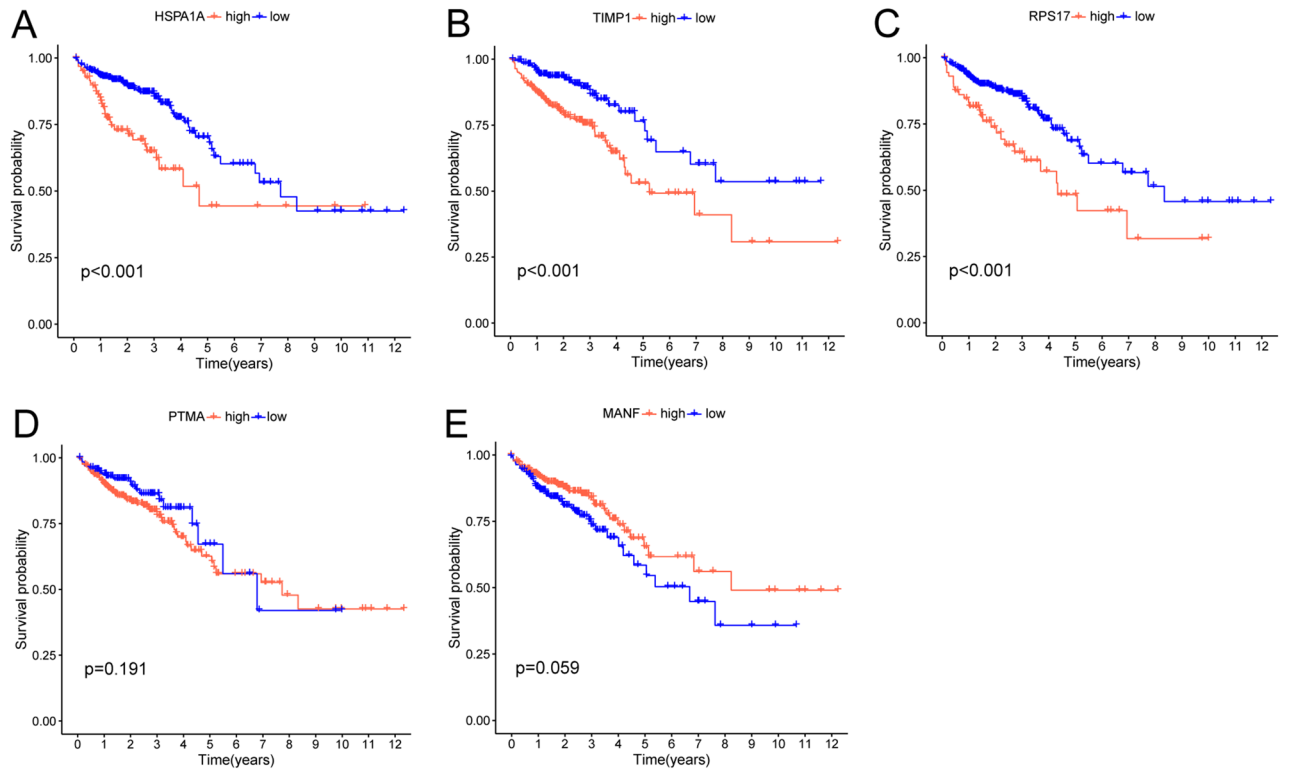


**Figure 5.** Validation of the prognostic signature. (A–B) Overall Survival (OS) for TCGA cohort and GEO cohort. (C–D) ROC curves of signatures in the TCGA cohort and GEO cohort. (E–F) Correlation of risk-scores and survival statuses of CRC patients.

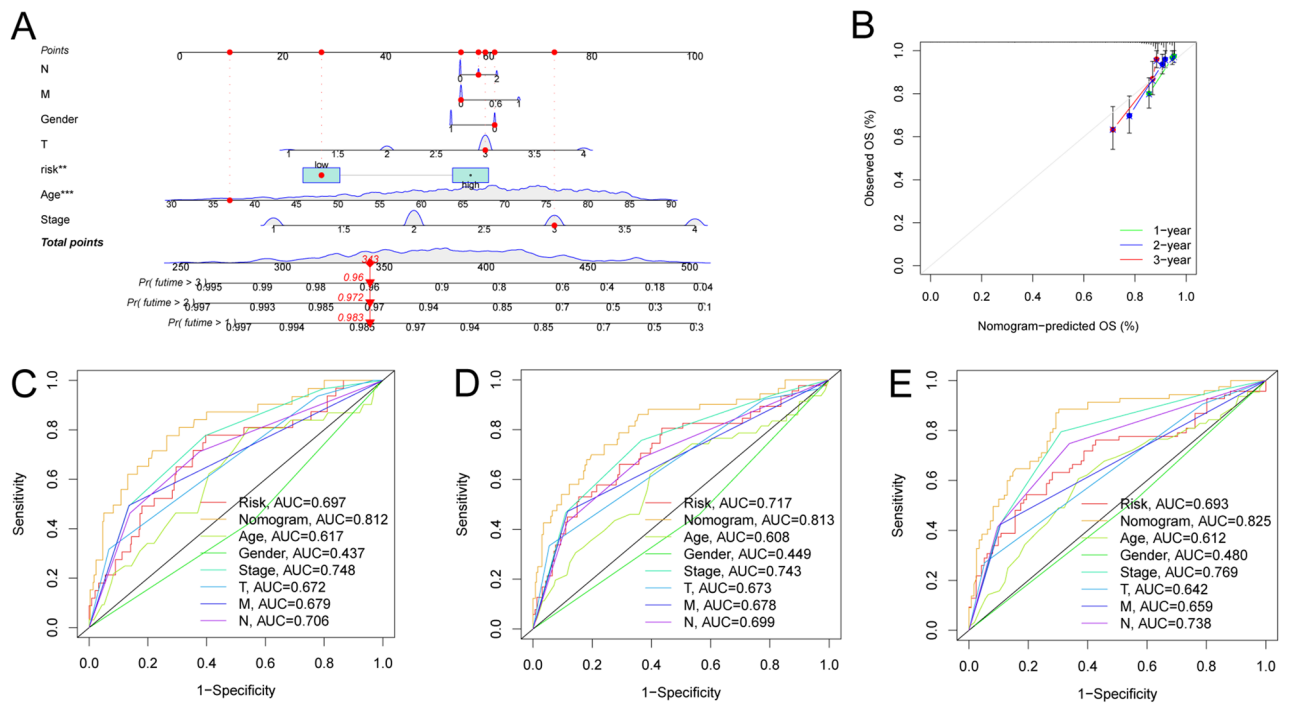
yet been fully elucidated. Therefore, a comprehensive analysis of FDRGs can contribute to the advancement of this field.

In this study, we calculated the FerrScore of different CRC-related cell types and defined low- and high-ferroptosis cell groups based on each cell type-specific median FerrScore. We also identified FDRGs in the two groups. Notably, using the TCGA database, we developed an FDRG-based risk signature to predict CRC prognosis. Five genes (HSPA1A, MANF, PTMA, RPS17, and TIMP1) were proven to have significant prognostic value. We also used qRT-PCR to test our predicted gene model and finally found that HSPA1A, TIMP1, and RPS17 were highly expressed in CRC cells and positively correlated with poor prognosis. These genes aid in distinguishing between the two risk groups. The high-risk group had a worse prognosis. Our ROC and calibration curves showed that

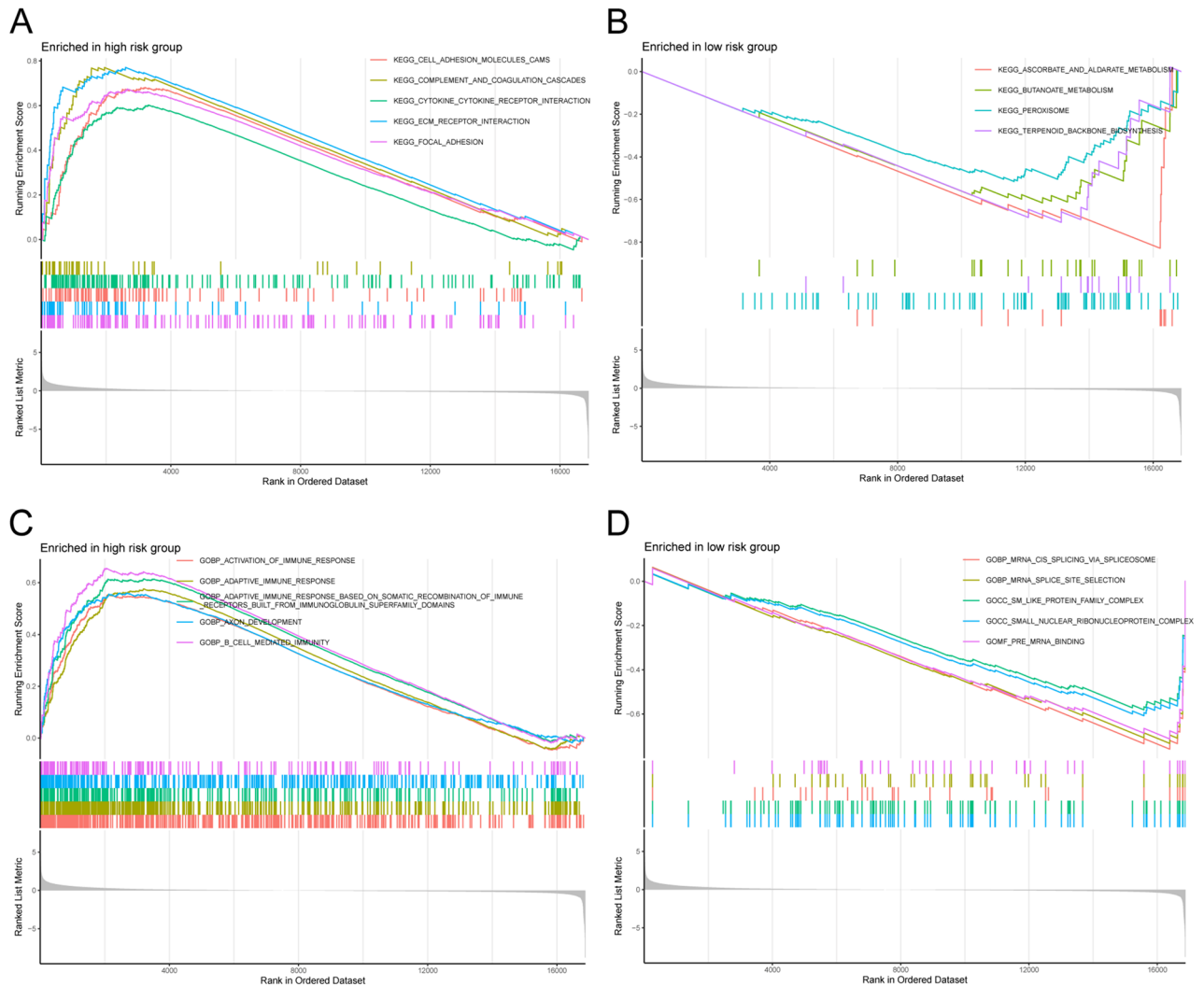




**Figure 6.** OS differences of different expression levels of genes in the signature (A–E). Lower expression of these genes displayed better OS, including HSPA1A, TIMP1, and RPS17, while PTMA and MANF’s were not significantly associated with OS.



**Figure 7.** Nomogram construction. (A) A nomogram was constructed to facilitate prognosis prediction. Age, TNM, stage and risk-score variables were factored in the nomogram, and the total points indicated the survival probabilities of patients. (B) The calibration curve indicated that the predicted OSs by the nomogram were relatively consistent with the actual observed OSs. (C–E) The AUC values for diverse clinical factors, risk scores, and nomogram scores were determined via ROC curves at intervals of 1, 2, and 3 years.

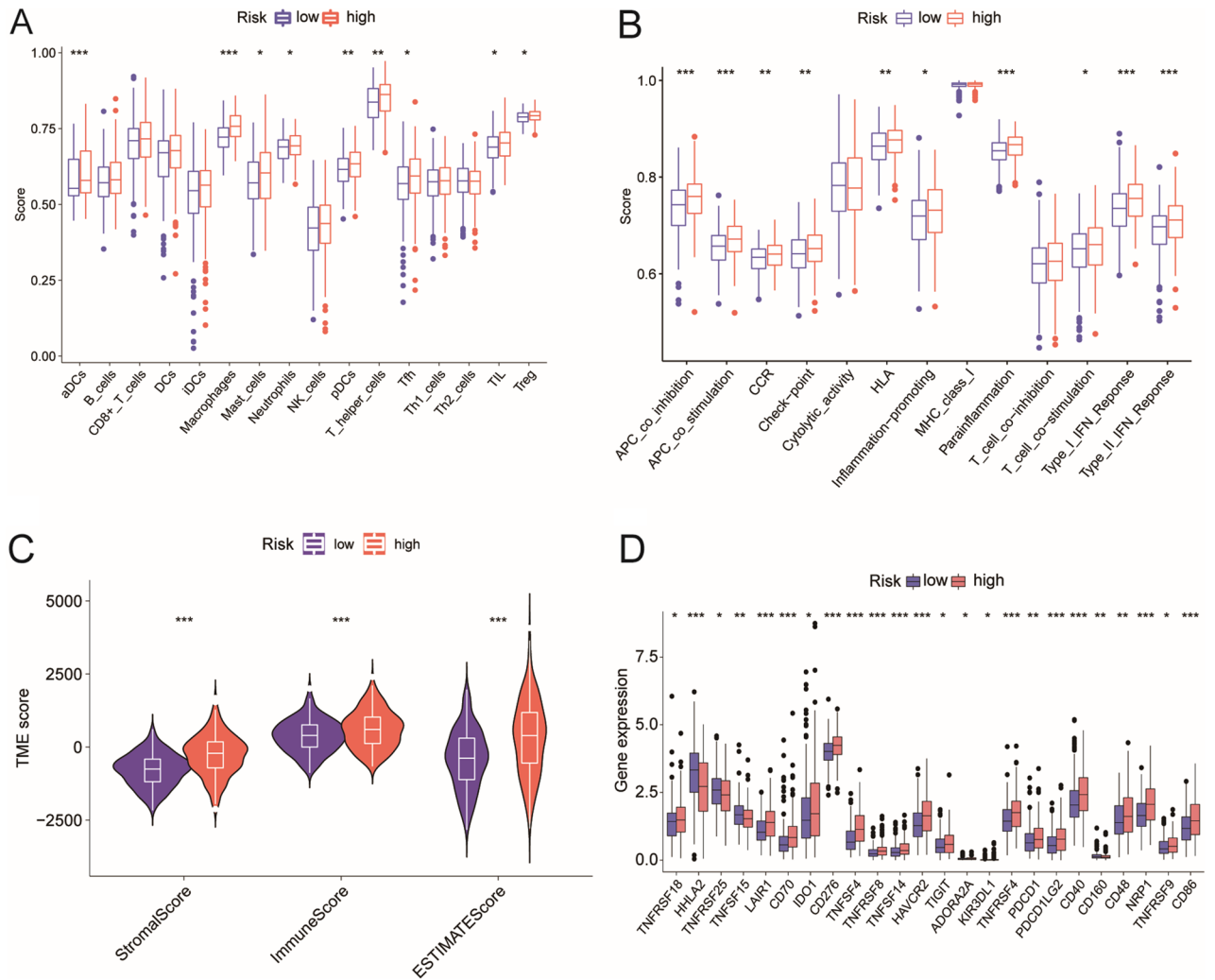


**Figure 8.** Gene set enrichment analysis. (A–B) The differentially pathways were significantly enriched in the risk groups by Kyoto Encyclopedia of Genes and Genomes (KEGG). (C–D) The differentially pathways were significantly enriched in the risk groups by Gene Ontology (GO).

our method accurately predicted 1-, 2-, and 3-year survival in CRC cases. To further investigate the mechanism of this method, KEGG and GO enrichment analyses were performed. As shown by KEGG analysis, the high-risk group possessed mutational traits mainly related to the regulation of cell adhesion, as demonstrated by ECM-receptor interactions and focal adhesions. Several studies have demonstrated that ECM-receptor interactions contribute to invasion and metastasis in CRC<sup>36–38</sup>. Our GO analysis showed that the high-risk group exhibited mutational traits mainly related to immune responses and B cell-mediated immunity. Moreover, GSEA showed that epithelial cells, monocytes, and endothelial cells were positively correlated with related pathways involved in NOTCH signaling, PI3K-AKT-MTOR signaling, and WNT-BETA-CATENIN signaling. Evidence suggests that these pathways play essential roles in the development of cancer.<sup>39–41</sup>. Therefore, we speculated that ferroptosis may have contributed to the occurrence and development of CRC through these signaling pathways.

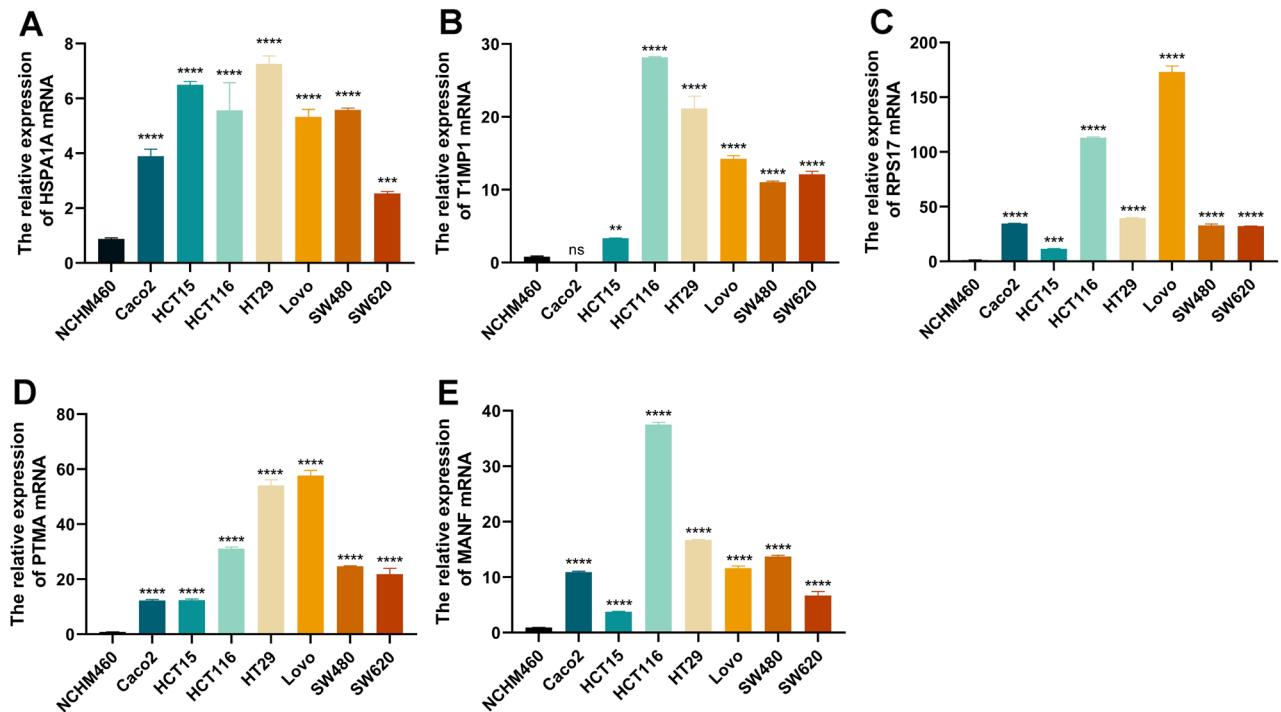
Ferroptosis has been linked to tumor immune infiltration. Wang et al. found that CD8<sup>+</sup> T cells and fatty acids orchestrate tumor ferroptosis and immunity via ACSL4<sup>42</sup>. However, ferroptosis and tumor immune infiltration in CRC have not been extensively studied. We conducted ssGSEA to investigate the immune statuses of the different groups after enriching many immune-related functions in our GO study. Our results indicated that the high-risk group had increased immune cell infiltration and immune function activation. Additionally, the high-risk group had high stromal scores, immune scores, and estimated cores. Li et al.<sup>43</sup> demonstrated that high immune and stromal scores are associated with poor prognosis, which is consistent with our results. Studies have shown that immune checkpoint genes are strongly associated with immunotherapy outcomes<sup>44</sup>. Our study revealed a general increase in the expression of immune checkpoint genes in high-risk populations. Moreover, high expression levels of these genes in high-risk patients can cause immune failure, leading to the upregulation of inhibitory checkpoint genes. Consequently, high-risk populations exhibit elevated levels of inhibitory checkpoint genes, which hold potential value as targets for immunotherapy.

The majority of genes in this signature contributed positively to the risk score, indicating that they are oncogenes. Tissue inhibitor of matrix metalloproteinase 1 (TIMP1) plays a vital role in carcinogenesis. Shou et al.



**Figure 9.** Immune infiltration differences between risk groups. (A–B) ssGSEA scores of immune cells and immune functions in the risk groups. (C) Comparison of StromalScore, ImmuneScore, and ESTIMATEScore between high- and low-risk groups. (D) The difference in immune checkpoint gene expression between the risk groups. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .

have shown that TIMP1 overexpression is associated with poor prognosis in renal cell carcinoma<sup>45</sup>. In colon cancer, TIMP1 induces cell proliferation and invasion through the FAK/Akt signaling pathway<sup>46</sup>. The human ribosomal protein RPS17 plays an important role in several diseases. The mutations in the ribosomal protein S17 (RPS17) gene are closely associated with hereditary bone marrow failure syndrome<sup>47</sup>, while it has also been shown that RPS17 may affect CRC prognosis by controlling amino acid metabolism<sup>48</sup>, which is consistent with our study. The heat shock protein family A member 1A (HSPA1A) belongs to the heat shock protein 70 family. Studies have shown that HSPA1A decreases survival rates in cancers and is associated with cell proliferation and tumor grade<sup>49</sup>. Chen et al.<sup>50</sup> confirmed that LIM and SH3 protein 1 (LASP1) and HSPA1A are both upregulated in head and neck squamous cell carcinoma, and directly bind to one another. However, few studies have explored the roles of these genes in CRC, which is an aspect where our research may provide some insights. MANF has been shown to improve colonic injury and negatively regulate macrophage polarization in colitis, indicating its potential as a therapeutic target for colonic inflammation<sup>51</sup>. Additionally, PTMA has been found to promote the malignant phenotype and participate in the progression of CRC, suggesting its involvement in cancer development<sup>52</sup>. The two studies that have shed light on the role of MANF and PTMA in colorectal cancer (CRC) development. Although the expression level of both high and low MANF and PTMA did not have a direct association with prognostic outcomes in predictive models, our study identified the high expression of five genes in CRC cells. This underscores the crucial and pressing need for additional research to further validate the predictive model and uncover more genes that could potentially serve as clinically valuable prognostic markers for CRC. Understanding the functions and interactions of the various factors involved in CRC development can help identify diagnostic and prognostic biomarkers, as well as potential targets for therapeutic intervention. As such, continued research and exploration in this field is crucial for developing effective strategies for the prevention, diagnosis, and treatment of CRC.



**Figure 10.** qRT-PCR analysis of predicted gene models for mRNA levels in CRC cells. (A–E) mRNA levels of HSPA1A, TIMP1, RPS17, PTMA, MANF in normal colonic epithelial cells NCHM460 versus CRC Caco2, HCT15, HCT116, HT29, Lovo, SW480, SW620. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .

Overall, we identified FDRGs and developed a transcriptome-based prognosis prediction method for CRC. However, owing to the inherent complexity of tumors, other omics and experimental studies are needed to further verify the roles of these biomarkers.

In conclusion: We defined a low ferroptosis and a high ferroptosis cell group based on each cell type-specific median FerrScore. We also identified FDRGs from the two groups. The prognostic signature developed from these FDRGs demonstrated a high predictive ability for CRC outcomes.

### Data availability

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

Received: 2 November 2022; Accepted: 25 July 2023

Published online: 04 August 2023

### References

- Sung, H. *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* **71**(3), 209–249 (2021).
- Puccini, A. & Lenz, H. J. Colorectal cancer in 2017: Practice-changing updates in the adjuvant and metastatic setting. *Nat. Rev. Clin. Oncol.* **15**(2), 77–78 (2018).
- Islam, S. *et al.* Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* **21**(7), 1160–1167 (2011).
- Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Meth.* **6**(5), 377–382 (2009).
- Li, H. *et al.* Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat. Genet.* **49**(5), 708–718 (2017).
- Poonpanichakul, T. *et al.* Capturing tumour heterogeneity in pre- and post-chemotherapy colorectal cancer ascites-derived cells using single-cell RNA-sequencing. *Biosci. Rep.* **41**(12), BSR20212093 (2021).
- Dixon, S. J. *et al.* Ferroptosis: An iron-dependent form of nonapoptotic cell death. *Cell* **149**(5), 1060–1072 (2012).
- Chen, P. *et al.* Combinative treatment of beta-elemene and cetuximab is sensitive to KRAS mutant colorectal cancer cells by inducing ferroptosis and inhibiting epithelial-mesenchymal transformation. *Theranostics* **10**(11), 5107–5119 (2020).
- Lu, Y. *et al.* KLF2 inhibits cancer cell migration and invasion by regulating ferroptosis through GPX4 in clear cell renal cell carcinoma. *Cancer Lett.* **522**, 1–13 (2021).
- Wang, Y. *et al.* Iron metabolism in cancer. *Int. J. Mol. Sci.* **20**(1), 95 (2018).
- Wang, W. *et al.* CD8(+) T cells regulate tumour ferroptosis during cancer immunotherapy. *Nature* **569**(7755), 270–274 (2019).
- Zheng, X. *et al.* Single-cell transcriptomic profiling unravels the adenoma-initiation role of protein tyrosine kinases during colorectal tumorigenesis. *Sign. Transduct. Target Ther.* **7**(1), 60 (2022).
- McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:180203426 (2018).
- Aran, D., Hu, Z. & Butte, A. J. xCell: Digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**(1), 220 (2017).

15. Hanzelmann, S., Castelo, R. & Guinney, J. GSEA: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* **14**, 7 (2013).
16. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc.: Ser. B (Methodol.)* **58**(1), 267–288 (1996).
17. Yu, G. *et al.* clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**(5), 284–287 (2012).
18. Galon, J. *et al.* The immune score as a new possible approach for the classification of cancer. *J. Transl. Med.* **10**, 1 (2012).
19. Shimada, K. *et al.* Frequent genetic alterations in immune checkpoint-related genes in intravascular large B-cell lymphoma. *Blood* **137**(11), 1491–1502 (2021).
20. Jiang, F. *et al.* An immune checkpoint-related gene signature for predicting survival of pediatric acute myeloid leukemia. *J. Oncol.* **2021**, 5550116 (2021).
21. Song, D. *et al.* A model of seven immune checkpoint-related genes predicting overall survival for head and neck squamous cell carcinoma. *Eur. Arch. Otorhinolaryngol.* **278**(9), 3467–3477 (2021).
22. Tian, M. *et al.* A novel immune checkpoint-related seven-gene signature for predicting prognosis and immunotherapy response in melanoma. *Int. Immunopharmacol.* **87**, 106821 (2020).
23. Mei, Y. *et al.* Single-cell analyses reveal suppressive tumor microenvironment of human colorectal cancer. *Clin. Transl. Med.* **11**(6), e422 (2021).
24. Ren, J. *et al.* Development and validation of a metabolic gene signature for predicting overall survival in patients with colon cancer. *Clin. Exp. Med.* **20**(4), 535–544 (2020).
25. Luo, D. *et al.* Development and validation of a novel epigenetic signature for predicting prognosis in colon cancer. *J. Cell. Physiol.* **235**(11), 8714–8723 (2020).
26. Zhou, Y., Guo, Y. & Wang, Y. Identification and validation of a seven-gene prognostic marker in colon cancer based on single-cell transcriptome analysis. *IET Syst. Biol.* **16**(2), 72–83 (2022).
27. Jiang, C. *et al.* In silico development and clinical validation of novel 8 gene signature based on lipid metabolism related genes in colon adenocarcinoma. *Pharmacol. Res.* **169**, 105644 (2021).
28. Zuo, D. *et al.* Construction and validation of a metabolic risk model predicting prognosis of colon cancer. *Sci. Rep.* **11**(1), 6837 (2021).
29. Fang, Z. *et al.* Identification of a prognostic gene signature of colon cancer using integrated bioinformatics analysis. *World J. Surg. Oncol.* **19**(1), 13 (2021).
30. Cheng, X. *et al.* A novel enterocyte-related 4-gene signature for predicting prognosis in colon adenocarcinoma. *Front. Immunol.* **13**, 1052182 (2022).
31. Huang, W. *et al.* Prognostic costimulatory molecule-related signature risk model correlates with immunotherapy response in colon cancer. *Sci. Rep.* **13**(1), 789 (2023).
32. Qiu, C. *et al.* Identification of molecular subtypes and a prognostic signature based on inflammation-related genes in colon adenocarcinoma. *Front. Immunol.* **12**, 769685 (2021).
33. Zheng, L., Yang, Y. & Cui, X. Establishing and validating an aging-related prognostic four-gene signature in colon adenocarcinoma. *Biomed. Res. Int.* **2021**, 4682589 (2021).
34. Ouyang, S. *et al.* Inhibition of STAT3-ferroptosis negative regulatory axis suppresses tumor growth and alleviates chemoresistance in gastric cancer. *Redox. Biol.* **52**, 102317 (2022).
35. Guo, J. *et al.* Ferroptosis: A novel anti-tumor action for cisplatin. *Cancer Res Treat.* **50**(2), 445–460 (2018).
36. Nersisyan, S. *et al.* ECM-receptor regulatory network and its prognostic role in colorectal cancer. *Front. Genet.* **12**, 782699 (2021).
37. Crotti, S. *et al.* Extracellular matrix and colorectal cancer: How surrounding microenvironment affects cancer cell behavior?. *J. Cell Physiol.* **232**(5), 967–975 (2017).
38. Maltseva, D. V. & Rodin, S. A. Laminins in metastatic cancer. *Mol. Biol. (Mosk.)* **52**(3), 411–434 (2018).
39. Xie, Q. *et al.* Tspan5 promotes epithelial-mesenchymal transition and tumour metastasis of hepatocellular carcinoma by activating Notch signalling. *Mol. Oncol.* **15**(11), 3184–3202 (2021).
40. Zhu, H., Zhao, N. & Jiang, M. Isovotexin attenuates tumor growth in human colon cancer cells through the modulation of apoptosis and epithelial-mesenchymal transition via PI3K/Akt/mTOR signaling pathway. *Biochem. Cell Biol.* **99**(6), 741–749 (2021).
41. Gorka, J. *et al.* MCP1 inhibits Wnt/beta-catenin signaling pathway activity and modulates epithelial-mesenchymal transition during clear cell renal cell carcinoma progression by targeting miRNAs. *Oncogene* **40**(50), 6720–6735 (2021).
42. Liao, P. *et al.* CD8(+)T cells and fatty acids orchestrate tumor ferroptosis and immunity via ACSL4. *Cancer Cell* **40**(4), 365–378.e6 (2022).
43. Li, X. Y. *et al.* A novel model based on necroptosis-related genes for predicting prognosis of patients with prostate adenocarcinoma. *Front. Bioeng. Biotechnol.* **9**, 814813 (2021).
44. Oshi, M. *et al.* A novel 4-gene score to predict survival, distant metastasis and response to neoadjuvant therapy in breast cancer. *Cancers (Basel)* **12**(5), 1148 (2020).
45. Shou, Y. *et al.* TIMP1 indicates poor prognosis of renal cell carcinoma and accelerates tumorigenesis via EMT signaling pathway. *Front. Genet.* **13**, 648134 (2022).
46. Ma, B. *et al.* TIMP1 promotes cell proliferation and invasion capability of right-sided colon cancers via the FAK/Akt signaling pathway. *Cancer Sci.* **113**(12), 4244–4257 (2022).
47. Farrar, J. E. *et al.* Ribosomal protein gene deletions in diamond-blackfan anemia. *Blood* **118**(26), 6943–6951 (2011).
48. Ren, Y. *et al.* A prognostic model for colon adenocarcinoma patients based on ten amino acid metabolism related genes. *Front. Public Health.* **10**, 916364 (2022).
49. Daugaard, M., Jaattela, M. & Rohde, M. Hsp70-2 is required for tumor cell growth and survival. *Cell Cycle* **4**(7), 877–880 (2005).
50. Chen, Q. *et al.* LASP1 promotes proliferation, metastasis, invasion in head and neck squamous cell carcinoma and through direct interaction with HSPA1A. *J. Cell. Mol. Med.* **24**(2), 1626–1639 (2020).
51. Yang, L. *et al.* MANF ameliorates DSS-induced mouse colitis via restricting Ly6C(hi)CX3CR1(int) macrophage transformation and suppressing CHOP-BATF2 signaling pathway. *Acta Pharmacol. Sin.* **44**(6), 1175–1190 (2023).
52. Yang, L. *et al.* Circular RNA hsa\_circ\_0004277 contributes to malignant phenotype of colorectal cancer by sponging miR-512-5p to upregulate the expression of PTMA. *J. Cell. Physiol.* <https://doi.org/10.1002/jcp.29484> (2020).

## Acknowledgements

The authors gratefully acknowledge each editor and reviewer for their efforts on this study.

## Author contributions

X.X. and X.Z. designed the study and wrote the manuscript. Q.L. and Y.Q. performed the analysis. Y.L. collected the dataset. W.T. reviewed and revised the manuscript.

## Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-39412-y>.

**Correspondence** and requests for materials should be addressed to W.T.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023