Development of a metastasis-related genes prognostic model for colorectal cancer

Tong Li
Department of Laboratory Medicine, Zhongshan Hospital, Fudan University, Shanghai
Qian Yu
Department of Laboratory Medicine, Wusong Branch, Zhongshan Hospital, Fudan University, Shanghai
Te Liu
Shanghai Geriatric Institute of Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai
Wenjing Yang
Department of Laboratory Medicine, Zhongshan Hospital, Fudan University, Shanghai
Wei Chen
Department of Laboratory Medicine, Zhongshan Hospital, Fudan University, Shanghai
Anli Jin
Department of Laboratory Medicine, Zhongshan Hospital, Fudan University, Shanghai
Hao Wang
Department of Laboratory Medicine, Zhongshan Hospital, Fudan University, Shanghai
Lin Ding
Department of Laboratory Medicine, Zhongshan Hospital, Fudan University, Shanghai
Chunyan Zhang
Department of Laboratory Medicine, Xiamen Branch, Zhongshan Hospital, Fudan University, Xiamen
Baishen Pan
Department of Laboratory Medicine, Zhongshan Hospital, Fudan University, Shanghai
Beili Wang
Department of Laboratory Medicine, Zhongshan Hospital, Fudan University, Shanghai
Wei Guo (guo.wei@zs-hospital.sh.cn)
Department of Laboratory Medicine, Zhongshan Hospital, Fudan University, Shanghai

---

Research Article

Keywords: TCGA, GEO, Lasso regression analysis, prognostic signature, immune infiltration

Posted Date: April 19th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1517037/v1

Abstract

Background

The colorectal tumor is a malignant tumor, the most important cause of death from colorectal cancer is metastasis. The primary objective of this study was to build a prognosis model by analyzing the differential genes between metastatic and non-metastatic CRC; Create a corresponding nomogram for clinical risk assessment of prognostic indicators to facilitate interpretation.

Methods

In this study, the main data on colorectal tumors are obtained from TCGA data set. Then, we used univariate COX regression analysis & Lasso regression analysis to find and filter characteristic variables. We construct the Cox proportional hazards regression model and verified the stability of the model by GEO external data set, the nomograms based on MR-risk score were obtained for clinical use. Finally, we explored the underlying mechanism of MR-risk score and colorectal cancer prognosis by using immune infiltration and immune checkpoint analysis.

Results

In this study, the prognostic model of colorectal cancer patients was constructed based on the colorectal metastasis-related genes, and the correlation between MR-risk core and the prognosis of colorectal cancer patients was obtained. In TCGA database, AUC of training set was 0.72, and the AUC of the testing set was 0.76, at the same time, AUC of GEO external data set was 0.68, which proved that treatment and prognosis of patients could be effectively determined. Finally, we found that the immune cells infiltrated in the tissues of patients with high MR-risk score were mostly in immune static or inactivated state compared with those of patients with low MR-risk score.

Conclusions

In summary, MR-risk score has a direct correlation with the prognosis of colorectal cancer, It is useful for predicting prognosis and immunotherapy of colorectal cancer.

Background

When a malignant tumor of the digestive tract appears, it is likely to cause colorectal cancer, which poses a serious threat to human health. According to the Global Cancer Epidemic Statistics (GLOBOCAN 2020) published by the World Health Organization's International Agency for Research on Cancer (IARC), 1.93 million new cases of colorectal cancer and 0.94 million deaths worldwide in 2020 is located in the third and second place of all malignant tumors[1, 2]. Patients with colorectal cancer have approximately a 50% chance of survival within five years, and more patients had metastasis or even lost the opportunity of surgery at the time of diagnosis due to the early symptoms were not obvious, resulting in the 5-year

survival rate decreased to 12% and 30–50% of colorectal cancer patients have recurrence and metastasis after treatment, which seriously affects patients' prognosis and quality of life [3, 4].

Metastasis is the leading cause of death in patients with colorectal cancer. Liver metastasis and lung metastasis are the common metastasis modes in patients with colorectal cancer, but the potential molecular mechanism remains unclear[4–6]. At present, the rapid development of Immunotherapy and targeted therapy can improve the survival time of patients with colorectal cancer to some extent[7, 8]. Still, the curative effect of colorectal cancer patients with metastasis is not apparent[9, 10]. Therefore, early warning of colorectal cancer metastasis and finding reliable markers to develop new colorectal cancer prognosis prediction models are essential for finding disease progression and timely adjusting treatment strategies.

We assume that metastasis-related genes are related to the prognosis and tumor immunity of colorectal cancer. We conducted a series of studies to evaluate this hypothesis, it includes univariate Cox regression, Lasso regression, Cox proportional hazards regression, functional enrichment analysis and clinical correlation analysis. We constructed an innovative prognostic model for predicting tumor metastasis and clinical outcomes of patients, and the model also had a particular predictive effect on colorectal tumor metastasis. Our results provide new insights into the role of colorectal metastasis-related genes in the tumor immune microenvironment and provide a theoretical basis for predicting colorectal cancer prognosis.

Materials And Methods

Data collection :

The mRNA expression data in this study were downloaded from TCGA-COAD dataset (The Cancer Genome Atlas), including normal group (n=42), tumor group (n=479). The "Limma" R package was used to integrate and standardize the downloaded FPKM mRNA Level 3 data, and the transfer-related differentially expressed genes and their expression levels were analyzed. The screening conditions of transfer-related differentially expressed genes were ($|\log2FC| \geq 1$, p.value < 0.05). In addition, the Matrix File data of GSE38832[11] was downloaded from the NCBI Gene Expression Omnibus (GEO) public database, and the annotation platform was GPL570. The data of 122 COAD patients with complete expression profile and survival information were downloaded. The data from TCGA and GEO are not classified but publicly available. This provides great convenience for this research. During the research process, TCGA and GEO data access rules are observed, and a systematic research strategy is developed.

GO and KEGG functional enrichment analysis :

A variety of theories, such as gene ontology and genome encyclopedia, were applied to reasonably analyze the functional categories of metastasis-related differential genes, and it was found that Gene ontology (GO) and Kyoto Encyclopedia of Genes pathways (KEGG) with p and q values less than 0.05 were significantly enriched in several pathways.

Prognostic model construction and validation :

In the TCGA cohort, the differential genes related to tumor metastasis were selected. Firstly, the differential genes related to colorectal cancer metastasis were preliminarily screened by univariate factor COX regression. In the Cox proportional hazards regression model, 14 genes with Lasso regression model $p < 0.01$ were included, and the most representative differential genes in further strains represented the metastasis-associated genes cluster. The MR-risk score of patients was mainly calculated by the expression level of their genes and the corresponding regression coefficient. The formula established is that the MR-risk score = sum (expression of each gene × corresponding coefficient). The results of this scoring scheme can be divided into two groups, with the median risk as the score point, the high-risk group above, and the low-risk group below.

Survival analysis :

The difference in survival time between the two groups was evaluated by Kaplan-Meier method and analyzed by Log-rank statistical method. As for the actual prognostic effect of MR-Risk score in predicting patients, calibration analysis method was used to detect the actual prediction results, and the ROC curve was used to study the accuracy of this model. The function of 'Surv Cutpoint' packaged by 'Surv Miner' R and 'Surv Cutpoint' packaged by 'Surv Miner' R were used to further understand the survival condition of genes, and the best cut value was found to predict the characteristics of genes.

Immunocyte infiltration analysis :

The RNA-seq data of COAD patients in different groups were calculated by Cibersort algorithm, so as to obtain the relative existence ratio of 22 immunoinfiltrating cells, and then Spearman was used to analyze the correlation between gene expression and the proportion content of immune cells. If the difference p.value <0.05, it is considered to be a statistically significant study.TISIDB integrates data information from various databases (TCGA, UniProt, GO, DrugBank, etc.) and is a valuable resource for cancer immunology research and treatment. Data on the interaction between tumors and the immune system can be retrieved from the TISIDB website(http://cis.hku.hk/TISIDB/browse.php).

Statistical analysis

The survival curve generated by Kaplan-Meier method not only adopts the scientific Log-rank method as the comparison method, but also adopts the multivariate Cox proportional risk analysis model. In addition, all statistical analyses were carried out with R language of 3.6 version, and all statistical tests were also conducted with bilateral tests.

Flowchart:

Flow chat of the study screening process is shown in Figure.1.

Results

1. Cancer Genome Atlas (TCGA) analysis revealed metastasis-related differentially expressed genes

We downloaded raw mRNA expression data (FPKM) of processed COAD from the TCGA database and obtained the metastasis through clinical indicator (M = 1: metastasis; M = 0: no metastasis). Differential expression analysis was performed between colon cancer metastasis patients and colon cancer non-metastasis patients through the 'Limma' package. From the research results, we can see that a total of 553 genes that might be related to metastasis were differentially expressed ($|\log2FC| \geq 1$ and $p < 0.05$). Among them, 331 genes were up-regulated and 222 genes were down-regulated(Fig. 2.A). Then 552 differentially expressed genes were screened by gene expression value (genes whose expression value was not 0 or the average expression value was more significant than 0.3 in 50% samples). GO and KEGG enrichment analysis showed that These differential genes are enriched in GO pathways such as antigen processing presentation (antigen processing presentation), tumor necrosis factor response (response to tumor necrosis factor), and MHC protein complex binding (MHC protein complex binding)(Fig. 2.B). In the process of KEGG enrichment, there are Antigen processing and presentation (antigen processing and presentation), Primary immunodeficiency (primary immune deficiency), TNF signaling pathways, and a large number of genes are enriched in metabolic-related pathways (Fig. 2.C).

Moreover, the basic characteristics of the Training and Test Sets are listed in Table 1.

Table 1
The Characteristics of Patients in the Training and Test Sets

	TCGA testing	TCGA training	p
n	80	318	
futime (mean (SD))	780.88 (890.88)	710.01 (685.21)	0.439
fustat = 1 (%)	12 (15.0)	58 (18.2)	0.606
age (mean (SD))	68.90 (11.89)	66.74 (12.58)	0.166
gender = MALE (%)	35 (43.8)	176 (55.3)	0.083
stage (%)			0.433
Stage I	12 (15.0)	61 (19.2)	
Stage II	3 (3.8)	23 (7.2)	
Stage IIA	29 (36.2)	95 (29.9)	
Stage IIB	1 (1.2)	8 (2.5)	
Stage IIC	0 (0.0)	1 (0.3)	
Stage III	1 (1.2)	13 (4.1)	
Stage IIIA	1 (1.2)	5 (1.6)	
Stage IIIB	9 (11.2)	42 (13.2)	
Stage IIIC	11 (13.8)	18 (5.7)	
Stage IV	10 (12.5)	35 (11.0)	
Stage IVA	2 (2.5)	13 (4.1)	
Stage IVB	0 (0.0)	2 (0.6)	
unknow	1 (1.2)	2 (0.6)	
T (%)			0.367
T1	2 (2.5)	7 (2.2)	
T2	11 (13.8)	61 (19.2)	
T3	62 (77.5)	209 (65.7)	
T4	3 (3.8)	24 (7.5)	
T4a	2 (2.5)	11 (3.5)	
T4b	0 (0.0)	6 (1.9)	

	TCGA testing	TCGA training	p
M (%)			0.723
M0	68 (85.0)	268 (84.3)	
M1	11 (13.8)	39 (12.3)	
M1a	1 (1.2)	8 (2.5)	
M1b	0 (0.0)	3 (0.9)	
N (%)			0.143
N0	46 (57.5)	198 (62.3)	
N1	13 (16.2)	50 (15.7)	
N1a	2 (2.5)	10 (3.1)	
N1b	0 (0.0)	11 (3.5)	
N1c	0 (0.0)	2 (0.6)	
N2	18 (22.5)	36 (11.3)	
N2a	0 (0.0)	5 (1.6)	
N2b	1 (1.2)	6 (1.9)	

2. Construction of the prognosis model for metastatic colorectal cancer patients

In order to make the study more valuable and to explore the key genes of differential gene concentration related to metastasis of colorectal cancer, the relevant data of TCGA-COAD patients were integrated in this study and Cox univariate regression and Lasso regression feature selection algorithm were used to screen feature genes related to colon cancer metastasis (Fig. 3.A-C). The results showed that 26 prognostic genes were screened by Cox univariate regression, which was GAL, UCHL1, TRIP10, SERPINE1, SNAI1, BCL10, GSR, PHF2, DNAJB2, LRRC8A, CST6, JAG2, ASAH1, C4orf19, MOGS, GDI1, SNCG, ASRGL1, LEPROTL1, FDFT1, CNOT7, TSC22D3, TNK2, RNASET2, CPT2, PGM2 (Fig. 3.A). After the screening of characteristic variables by Lasso regression, 14 selected metastasis related genes were finally determined (Fig. 3.B-C). TCGA patients were randomly assigned, and the ratio of training set to validation set was 4 to 1 to construct and verify the Cox proportional hazards regression model. The optimal MR-risk score for each sample was obtained by regression model for subsequent analysis. (MR-Risk Score = ASRGL1 × (-0.200486209) + GSR × (-0.107200303) + ASAH1 × (-0.086956773) + BCL10 × (-0.065018786) + SNAI1 × 0.001969402 + TRIP10 × 0.02686487 + TSC22D3 × 0.063610852 + LRRC8A × 0.072751154 + PHF2 × 0.075148411 + SERPINE1 × 0.077260841 + RNASET2 × 0.110933675 + DNAJB2 × 0.13931055 + UCHL1 × 0.205894794 + GAL × 0.209549353)(Fig. 3.D). COX coefficient is shown in Table 2.

Table 2
COX coefficient

gene	coef	hr	low.ci	upp.ci
ASRGL1	-0.20049	0.559465	0.370334	0.845184
GSR	-0.1072	0.602317	0.445522	0.814294
ASAH1	-0.08696	0.650391	0.487754	0.867258
BCL10	-0.06502	0.586685	0.433268	0.794424
SNAI1	0.001969	1.424898	1.179443	1.721434
TRIP10	0.026865	1.548546	1.243982	1.927677
TSC22D3	0.063611	1.253047	1.06009	1.481127
LRRC8A	0.072751	1.298438	1.099505	1.533362
PHF2	0.075148	1.418221	1.140282	1.763906
SERPINE1	0.077261	1.222669	1.100739	1.358106
RNASET2	0.110934	1.268938	1.062059	1.516116
DNAJB2	0.139311	1.42694	1.141169	1.784273
UCHL1	0.205895	1.380657	1.195374	1.594658
GAL	0.209549	1.525341	1.270088	1.831894

3. Kaplan-Meier curve and time-varying ROC analysis were performed based on the risk values of training and test data sets

According to the median MR-Risk score (training set risk score median: -0.0374618011019423. The median MR-risk score in the test set was 0.0223075396080433). In the TCGA-COAD dataset, patients were divided into two groups, namely high-risk and low-risk groups based on median MR-risk score in the training (Fig. 4.A) and test set (Fig. 4.D), and studied by Kaplan-Meier curve. We can see that the low risk group has a significantly higher OS than the high risk group (Fig. 4.B & E). The results of ROC curve show that the C-index index of training set and test set is 0.72 and 0.76 (Fig. 4.C & F), respectively, suggesting that the model has good verification efficiency.

4. External validation of the performance of the MR-risk score model in GEO dataset

In order to ensure the stability of the prediction model, the RNA-Seq data (GSE38832) of COAD patients with processed survival data was retrieved from the GEO database, an external data set, and based on this model, the clinical classification of COAD patients was predicted. In the GSE38832 dataset, patients were divided into two groups, namely high-risk and low-risk groups based on median MR-risk score

(Fig. 5.A & B). Kaplan-Meier method was used to evaluate survival differences between the two groups. The results showed that the high risk group in the GEO external validation set had a significantly lower OS than the low risk group(Fig. 5.C).To verify the accuracy of the model, we used the external data set to analyze the ROC curve of the model, and the results showed that the model had a strong predictive effect on the prognosis of patients(GSE38832 C-index = 0.68) (Fig. 5.D). At the same time, the expression levels of 14 genes in the prediction model were detected in the GEO external validation set, and it was found that the expression levels of 14 genes were highly consistent in the GSE38832 data set (Fig. 5. E) and the TCGA-COAD dataset (Fig. 5. F).

5. 14 metastasis-related prognostic features associated with overall survival in patients with colorectal cancer

In order to prove the application value of this prediction model in clinical practice, We first used univariate and multivariate Cox risk regression models to evaluate the impact of MR-risk score on patients' survival(Fig. 6.A-B). The results showed that MR-risk score was an independent prognostic factor in patients with COAD. Clinically, we often classify the progression and malignant degree of tumors by staging classification so as to achieve the purpose of different surgical and drug treatments for tumors with further advancement and malignant degrees. According to the above, we divide each patient's MR-risk score value evaluated by the prediction model into different clinical commonly used staging classification groups and display the results of each clinical index group in the form of a box diagram (Fig. 6.C) and find that the distribution of MR-risk score value in stage, T, M, N and other clinical ($p < 0.05$) by rank-sum test (Kruskal. test). It shows that the MR-risk score obtained by our modeling analysis has good applicability for the grouping of samples, and with the increase of MR-risk score value, it indicates that the clinical tumor staging classification of patients shows a trend of deterioration and the possibility of tumor metastasis increases.In order to play the practical role of the prognostic model, we classified the samples into high risk group and low risk group based on the median of MR-risk score during the study.The results of regression analysis are shown in the form of a line chart(Fig. 6.D). In order to observe whether the prediction probability of the model is close to the empirical probability, we build the calibration plot by using the predicted five-year survival and seven-year survival (OS) of colon cancer(Fig. 6.E). We also calculated the proportion of cancer metastasis in high versus low-risk score subgroups, results shows that in high-risk subgroup patients have a higher probability of cancer metastasis(Fig. 6.F).

6. The immune infiltration performance of the model in the high-risk and low-risk Groups

The microenvironment of tumor formation is generally composed of a variety of complex components such as fibrocells and immune cells related to tumor, the extracellular matrix, a variety of growth factors that promote cell growth, cancer cells themselves and related inflammatory factors, and unique physical and chemical components that promote the formation of cancer cells. Tumor microenvironment has a certain or significant influence on the judgment of tumor and patient survival as well as clinical treatment. Studying the association between MR-Risk score and tumor immunoinvasion can further

optimize the mechanism effect of MR-Risk score in predicting potential colon cancer molecules.The results showed that the MR-risk score was positively correlated with Macrophages M0, Treg, and B cell naive, and negatively correlated with T cell CD4 memory resting, Eosinophils, Neutrophils and T cell follicular helper (Fig. 7.A-C). The correlation coefficient scatter plot of T cells gamma delta were not shown because the cell infiltration rate is too low to show the linear correlation between high and low MR-risk score. In order to evaluate the possible sensitivity of MR-risk score to immunotherapy, we examined the correlation between MR-risk score and the expression level of immune checkpoint in tumor tissues(Fig. 7.D). Interestingly, we found a negative correlation between MR-risk score and gene expression at most immune checkpoints in cancer tissues.

Discussion

Globally, colorectal cancer is the largest type of cancer, with 1.09 million new cases and 551,000 deaths in 2018[12]. At present, metastasis is still the leading cause of death in colorectal cancer patients. Therefore, an in-depth understanding of the molecular biology principles of metastasis is essential for the early detection of colorectal cancer metastasis, the optimization of metastasis monitoring, and timely prevention. Current diagnostic imaging tools, such as enhanced CT, positron emission tomography, and magnetic resonance imaging, can detect metastatic lesions of colorectal cancer[13]. However, these methods have limited value because they cannot effectively identify early metastatic lesions. Considering these clinical challenges, it is necessary to develop metastasis-specific molecular markers and prediction models that can help predict the prognosis of colorectal metastasis.

Our study innovatively constructed a prognostic gene risk model based on 14 CRC metastasis-related genes by analyzing the differentially expressed genes between metastatic CRC and non-metastatic CRC and using Lasso cox regression. This conclusion is verified internally by the TCGA-COAD dataset and also by GEO external dataset. In addition, we used CIBERSORT and TCDB to further analyze and explore the mechanism of metastatic CRC from the perspective of the immune microenvironment.

Through enrichment analysis of GO and KEGG signaling pathways[14, 15], we found that differentially expressed genes associated with metastatic CRC affect tumor immune-related Antigen processing and presentation (antigen processing and presentation), MHC protein complex binding (MHC protein complex binding), Primary immunodeficiency (primary immunodeficiency) and TNF signaling pathway. Antigen processing is a process in which cells collect antigens and degrade them into peptides[16]. Antigen presentation is when the processed antigen peptides are displayed on the cell surface to facilitate T cell-specific recognition[17]. Antigen presentation is an essential immune process, which is essential for triggering T cell immune response[18]. Extracellular antigen processing must be mediated by MHC-II, which is recognized by CD4 + T cells (helper T cells) and only expressed on the surface of antigen-presenting cells such as macrophages, dendritic cells, and B cells[19, 20].TNF is considered to be mainly produced by macrophages, and its main role is to regulate immune cells. As an endogenous pyrogen, TNF can induce fever, apoptosis, cachexia, inflammation, and inhibit tumorigenesis. It is used as an immune stimulant for the treatment of certain cancers[21]. TNF can bind to two receptors, TNFR1 (type 1

TNF receptor) and TNFR2 (type 2 TNF receptor)[22]. After exposure to their ligands, TNF receptors also form trimers leading to conformational changes in receptors, thereby activating downstream MAPK signaling pathways while promoting translocation of NF-κB and JNK to the nucleus and activating downstream transcription factors, thereby mediating a large number of transcriptions involving cell survival and proliferation, inflammation and anti-apoptotic proteins[23, 24].

With the gradual deepening of research on tumor microenvironment and tumor immunity, new technologies and means will be applied to the diagnosis and treatment of metastatic CRC[25]. According to the above information, we analyzed the infiltrating immune cells in the tumor tissues of high-risk score and low-risk score, and the results showed that compared with the high prognosis score group with the low model score, B cells, and macrophages, the two antigen-presenting cells, were both immature and unable to drive the antigen presentation usually. Memory CD4 + T can rapidly respond to secondary antigen stimulation, release cytokines interferon y (IFN-y), interleukin 4 (IL-4), IL5, and 1L-2 in a short time, and split rapidly[26, 27]. In the lymphocytes infiltrated in tumor tissues of patients with metastasis, we found that CD4 + memory cells were resting and could not effectively mediate the role of re-immune response[28].

In summary, we found the characteristics of immune cell infiltration in metastatic colorectal cancer tissues: antigen-presenting cells were immature and undifferentiated. Auxiliary memory cells in a resting state cannot exert secondary immune defense. Therefore, we have reason to speculate further in this ineffective antigen presentation and immune conditions, the function of CD8 + T cells will be affected by driving specific killing, but the results did not show that there is a difference in the proportion of CD8 + T cells infiltrated in tumor tissues between high-risk score and low-risk score. Therefore, we further consider whether the function of immune cells is inhibited, and the immune checkpoint plays an vital role in the normal physiological function of immune cells[29]. We found that many gene expression of immune checkpoints on tumor tissue cells in high-risk score group were inhibited by immune checkpoint analysis. This means that the immune checkpoint markers in tumor tissues of patients with high risk are reduced, and the response of patients to immunotherapy may not be obvious, which may be one of the possible reasons for poor prognosis of patients.

Therefore, we infer that the change of immune microenvironment in primary tumor tissue before colorectal cancer metastasis leads to the fact that antigen-presenting cells are immature and cannot perform standard antigen presentation. Memory CD4 + T cells in a static state cannot effectively carry out secondary immunization. At the same time, the expression level of immune checkpoints in the corresponding multiple immune response processes can be inhibited, leading to immune disorders in tumor tissues, immunosuppressive therapy unable to resist tumors, and ultimately leading to tumor removal of immune barriers and metastasis. Therefore, early prediction of the prognosis of metastatic colorectal tumors can further prevent and treat metastatic colorectal tumors in advance to improve the quality of life of patients.

Conclusions

Our study is the first time to build a model by analyzing the differential genes between metastatic and non-metastatic CRC patients. Previous studies mainly focus on modeling by analyzing the intersection of differential genes between colorectal cancer and adjacent tissues and metastasis-related gene sets. Before modeling, selective screening was made to remove other non-metastatic genes that can represent the characteristics of metastatic CRC. The loss of these factors is unreasonable for the universality and internal mechanism of the model, which makes the model itself have a specific choice deviation. Our model has been verified in addition to the internal verification of the TCGA data set and the external data set of GEO. The model has strong robustness. At the same time, we create the corresponding nomogram for clinical risk assessment of prognostic indicators to facilitate interpretation.

Abbreviations

CRC Colorectal cancer

COAD Colon adenocarcinoma

TCGA The Cancer Genome Atlas

GEO Gene Expression Omnibus

MR-risk score metastasis-related risk score

ROC receiver operating characteristic

Declarations

Ethics approval and consent to participate

The studies not involving human participants, written informed consent for participation was not required for this study. All methods were performed in accordance with the relevant guidelines.

Consent for publication

Not applicable.

Availability of data and materials

Data sets in current research data analysis come from public data sets : The Cancer Genome Atlas (TCGA) https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga and Gene Expression Omnibus(GEO) http://www.ncbi.nlm.nih.gov/geo (GSE38832)[11].

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by National Natural Science Foundation of China (81772263), National Natural Science Foundation of China (81972000), National Natural Science Foundation of China Youth Fund (81902139),National Natural Science Foundation of China Youth Fund (82000275), Specialized Fund for the clinical researches of Zhongshan Hospital affiliated Fudan University (2018ZSLC05), Shanghai outstanding medical freshmen talents (Yi Yuan Xin Xing 2019ZSYXQN27), The constructing project of clinical key disciplines in Shanghai (shslczdzk03302), The Key medical and health projects of Xiamen (YDZX20193502000002), Shanghai Medical Key Specialty (ZK2019B28).

Author's contributions

Qian Yu,Tong Li and Te Liu, Wenjing Yang designed the experiment;Wei Guo, Beili Wang and Chunyan Zhang, Baishen Pan gave administrative support; Wei Chen, Anli Jin and Hao Wang, Lin Ding collected and summarized data;Tong Li and Qian Yu analyzed data sorted results and wrote the manuscript; All authors reviewed the manuscript.

Acknowledgements

Not applicable.

References

1. Abualrous ET, Sticht J and Freund C. Major histocompatibility complex (MHC) class I and class II proteins: impact of polymorphism on antigen presentation. Curr Opin Immunol, 2021. 70: p. 95–104.
2. Bipat S, van Leeuwen MS, Ijzermans JN, et al. Evidence-base guideline on management of colorectal liver metastases in the Netherlands. Neth J Med, 2007. 65(1): p. 5–14.
3. Brahmer JR, Drake CG, Wollner I, et al. Phase I study of single-agent anti-programmed death-1 (MDX-1106) in refractory solid tumors: safety, clinical activity, pharmacodynamics, and immunologic correlates. J Clin Oncol, 2010. 28(19): p. 3167–75.
4. 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): p. 394–424.
5. Chadwick W, Magnus T, Martin B, et al. Targeting TNF-alpha receptors for neurotherapeutics. Trends Neurosci, 2008. 31(10): p. 504–11.
6. Chen G and Goeddel DV. TNF-R1 signaling: a beautiful pathway. Science, 2002. 296(5573): p. 1634–5.
7. Deng M, Gui X, Kim J, et al. LILRB4 signalling in leukaemia cells mediates T cell suppression and tumour infiltration. Nature, 2018. 562(7728): p. 605–609.
8. Dewson G and Silke J. The Walrus and the Carpenter: Complex Regulation of Tumor Immunity in Colorectal Cancer. Cell, 2018. 174(1): p. 14–16.

9. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. CA Cancer J Clin, 2016. 66(2): p. 115–32.
10. Ganesh K, Stadler ZK, Cercek A, et al. Immunotherapy in colorectal cancer: rationale, challenges and potential. Nat Rev Gastroenterol Hepatol, 2019. 16(6): p. 361–375.
11. Tripathi MK, Deane NG, Zhu J, et al. Nuclear factor of activated T-cell activity is associated with metastatic capacity in colon cancer. Cancer research, 2014. 74(23): p. 6947–6957.
12. Gene Ontology C. The Gene Ontology project in 2008. *Nucleic Acids Res*, 2008. 36(Database issue): p. D440-4.
13. Kanehisa M, Goto S, Hattori M, et al. From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res, 2006. 34(Database issue): p. D354-7.
14. Lederman S, Yellin MJ, Krichevsky A, et al. Identification of a novel surface protein on activated CD4 + T cells that induces contact-dependent B cell differentiation (help). J Exp Med, 1992. 175(4): p. 1091–101.
15. Mellman I and Steinman RM. Dendritic cells: specialized and regulated antigen processing machines. Cell, 2001. 106(3): p. 255–8.
16. Miller KD, Fidler-Benaoudia M, Keegan TH, et al. Cancer statistics for adolescents and young adults, 2020. CA Cancer J Clin, 2020. 70(6): p. 443–459.
17. Miller KD, Nogueira L, Mariotto AB, et al. Cancer treatment and survivorship statistics, 2019. CA Cancer J Clin, 2019. 69(5): p. 363–385.
18. Olszewski MB, Groot AJ, Dastych J, et al. TNF trafficking to human mast cell granules: mature chain-dependent endocytosis. J Immunol, 2007. 178(9): p. 5701–9.
19. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. Nat Rev Cancer, 2012. 12(4): p. 252–64.
20. Purcell AW, Croft NP and Tscharke DC. Immunology by numbers: quantitation of antigen presentation completes the quantitative milieu of systems immunology! Curr Opin Immunol, 2016. 40: p. 88–95.
21. Schmitt M and Greten FR. The inflammatory pathogenesis of colorectal cancer. Nat Rev Immunol, 2021. 21(10): p. 653–667.
22. Scully C, Georgakopoulou EA and Hassona Y. The Immune System: Basis of so much Health and Disease: 4. Immunocytes. Dent Update, 2017. 44(5): p. 436–8, 441–2.
23. Shin H and Iwasaki A. Tissue-resident memory T cells. Immunol Rev, 2013. 255(1): p. 165–81.
24. Siegel RL, Miller KD, Goding Sauer A, et al. Colorectal cancer statistics, 2020. CA Cancer J Clin, 2020. 70(3): p. 145–164.
25. Stern LJ and Santambrogio L. The melting pot of the MHC II peptidome. Curr Opin Immunol, 2016. 40: p. 70–7.
26. Van Cutsem E, Cervantes A, Nordlinger B, et al. Metastatic colorectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol, 2014. 25 Suppl 3: p. iii1-9.

27. van den Broek T, Borghans JAM and van Wijk F. The full spectrum of human naive T cells. Nat Rev Immunol, 2018. 18(6): p. 363–373.
28. Wajant H, Pfizenmaier K and Scheurich P. Tumor necrosis factor signaling. Cell Death Differ, 2003. 10(1): p. 45–65.
29. Yoshino T, Arnold D, Taniguchi H, et al. Pan-Asian adapted ESMO consensus guidelines for the management of patients with metastatic colorectal cancer: a JSMO-ESMO initiative endorsed by CSCO, KACO, MOS, SSO and TOS. Ann Oncol, 2018. 29(1): p. 44–70.

Figures

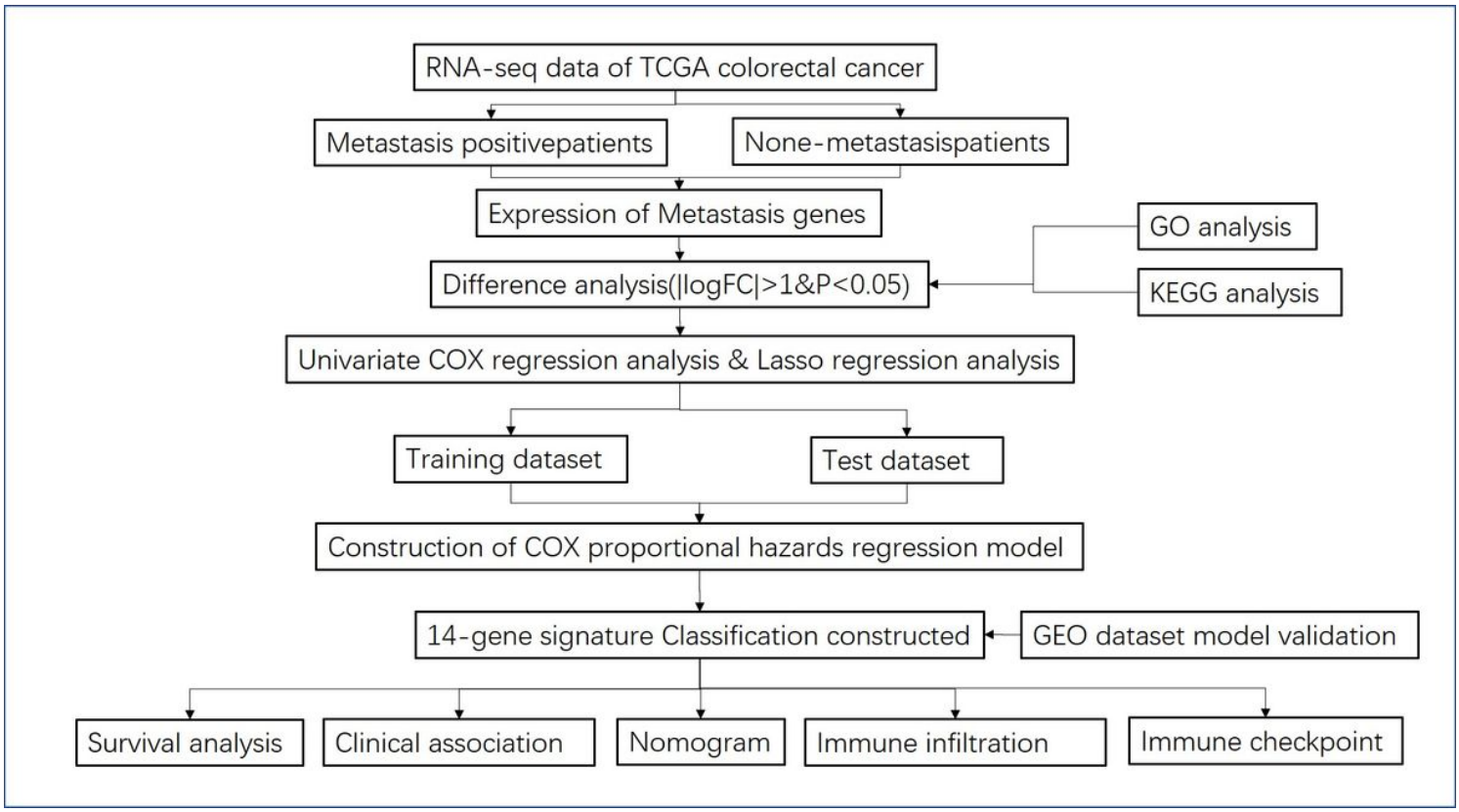

Figure 1

Workflow of this study.

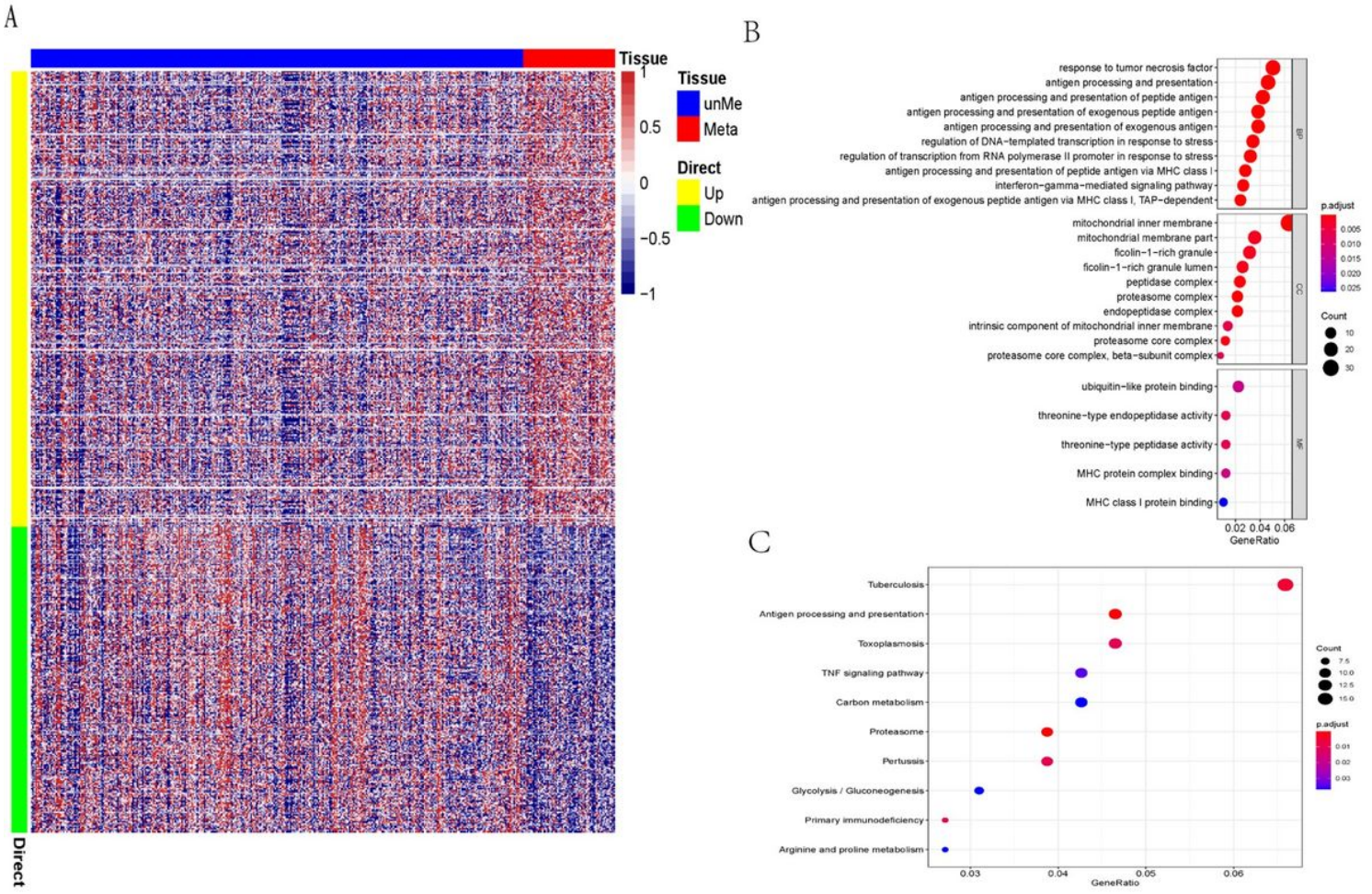

Figure 2

Cancer Genome Atlas (TCGA) analysis revealed metastasis-related differentially expressed genes.

(A) Heat map of differentially expressed genes in TCGA analysis. unMe: Colorectal cancer group without metastasis; Meta: Colorectal cancer group with metastasis. **(B)** GO pathway analysis results showed that GO related biological functions (BF), cellular components (CC) and molecular functions (MF) were the most abundant. **(C)** KEGG pathway analysis showed that biological pathways were the most abundant.

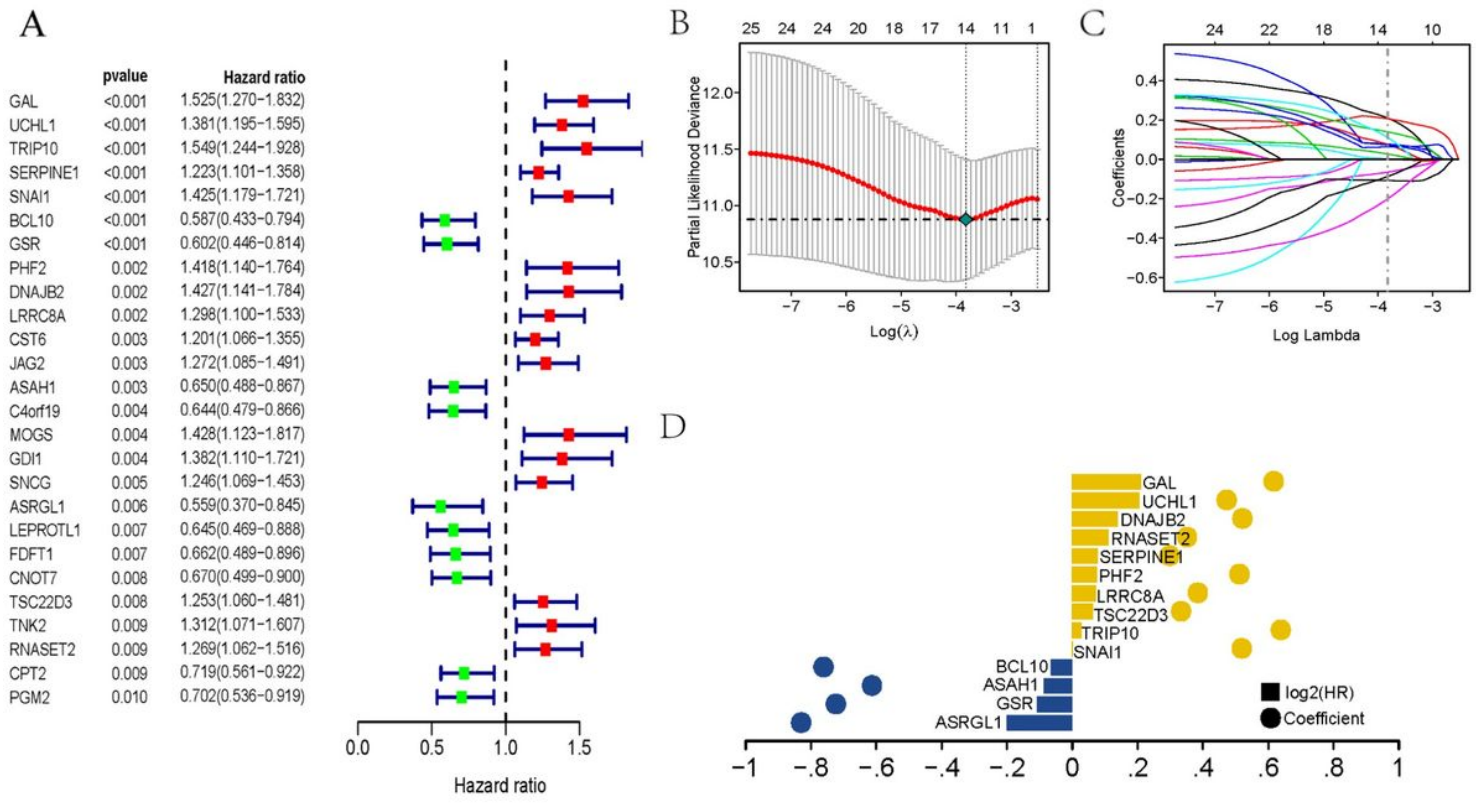

Figure 3

Construction of the prognosis model for metastatic colorectal cancer patients.

(A) Cox univariate regression analysis of metastasis-associated differentially expressed genes. (P<0.01). **(B)**Lasso regression analysis used the maximum criterion of 10-fold cross validation. **(C)** The Lasso coefficient spectrum of differentially expressed genes was screened. Lasso, minimum absolute contraction and selection operators. **(D)** The model coefficient diagram shows the HR values and correlation coefficients of 14 constituent genes in the prognostic model.

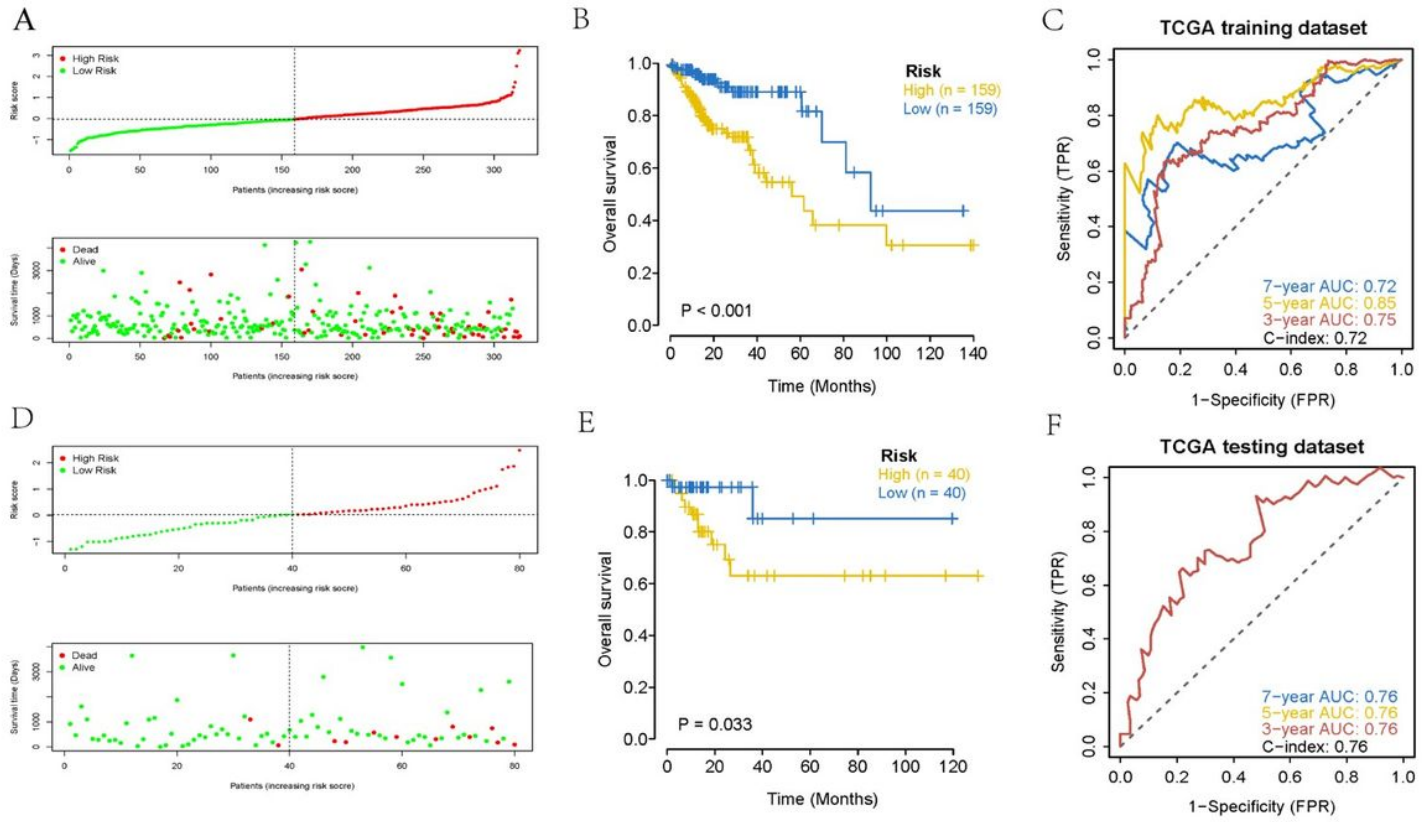

Figure 4

Kaplan-Meier curve and time-varying ROC analysis were performed based on the risk values of training and test data sets.

Patients were divided into high-risk and low-risk groups based on median MR-risk score in the training **(A)** and test set **(D)**. Kaplan-Meier curves and time-dependent ROC analyses for OS prediction based on the Risk score in the training **(B, C)** and test set **(E, F)**, respectively. High, high Risk score; Low, low Risk score; ROC, receiver operating characteristic.

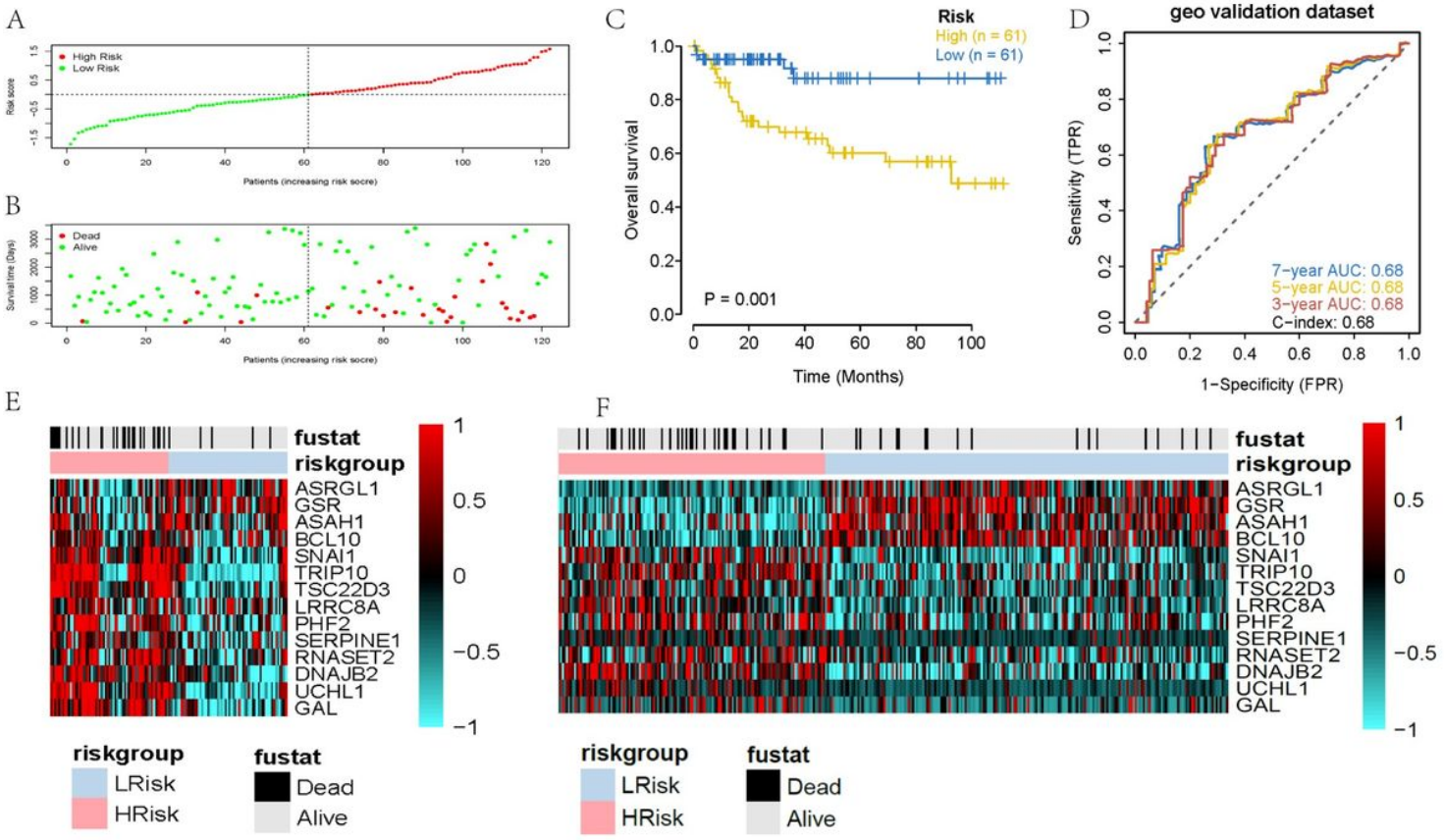

Figure 5

External validation of the performance of the MR-risk score model in GEO dataset.

(A) Distribution of risk scores for the GEO dataset based on the 14 metastasis-related genes in the prediction model. **(B)** The survival status of CRC patients in the GEO dataset belonging to the high- and low-MR-risk score groups. **(C)** Kaplan-Meier survival curves of patients in the two Risk score groups. **(D)** Time-dependent ROC curve for predicting survival time with AUC value in the two Risk score groups. Heat map of MR-risk score model-based 14 gene expression in GEO dataset **(E)** and TCGA-COAD dataset**(F)**.

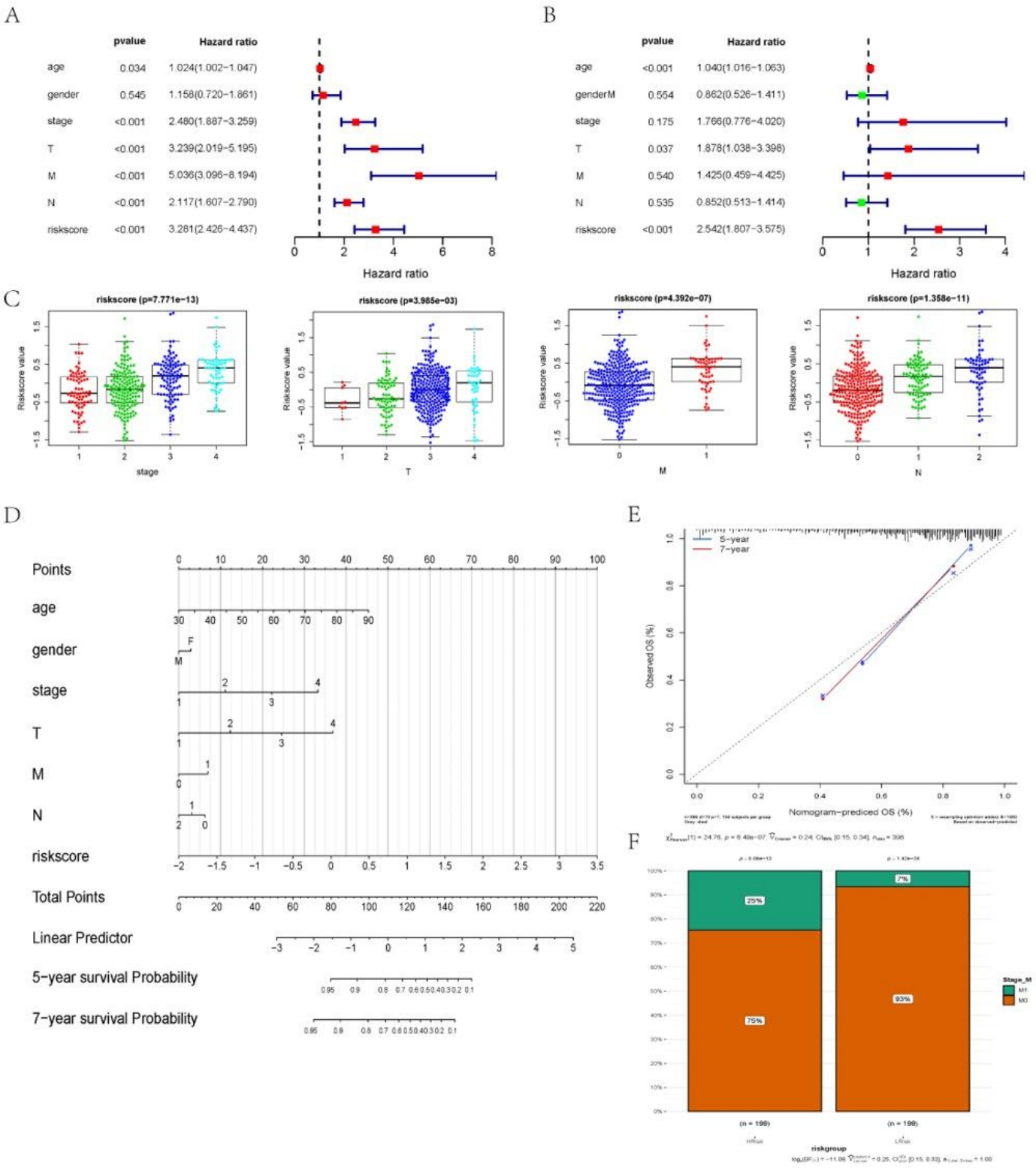

Figure 6

14 metastasis-related prognostic features associated with overall survival in patients with colorectal cancer.

(A) Univariate Cox regression analysis showed that clinicopathological data parameters in the TCGA cohort were associated with overall survival in CRC patients. **(B)** Multivariate Cox regression analysis

showed that clinicopathological data parameters were associated with overall survival in CRC patients in the TCGA cohort. **(C)** Correlation analysis between MR-Risk score and pathological staging time and TNM staging time of CRC patients. **(D)** A Nomogram of COAD5 and 7 years survival combined with cancer metastatic gene profiles. **(E)** Calibration plot of the prognostic model. The Y axis is the actual overall survival rate and the X axis is the predicted overall survival rate. **(F)** The proportion of cancer metastasis in high versus low-risk score subgroups.

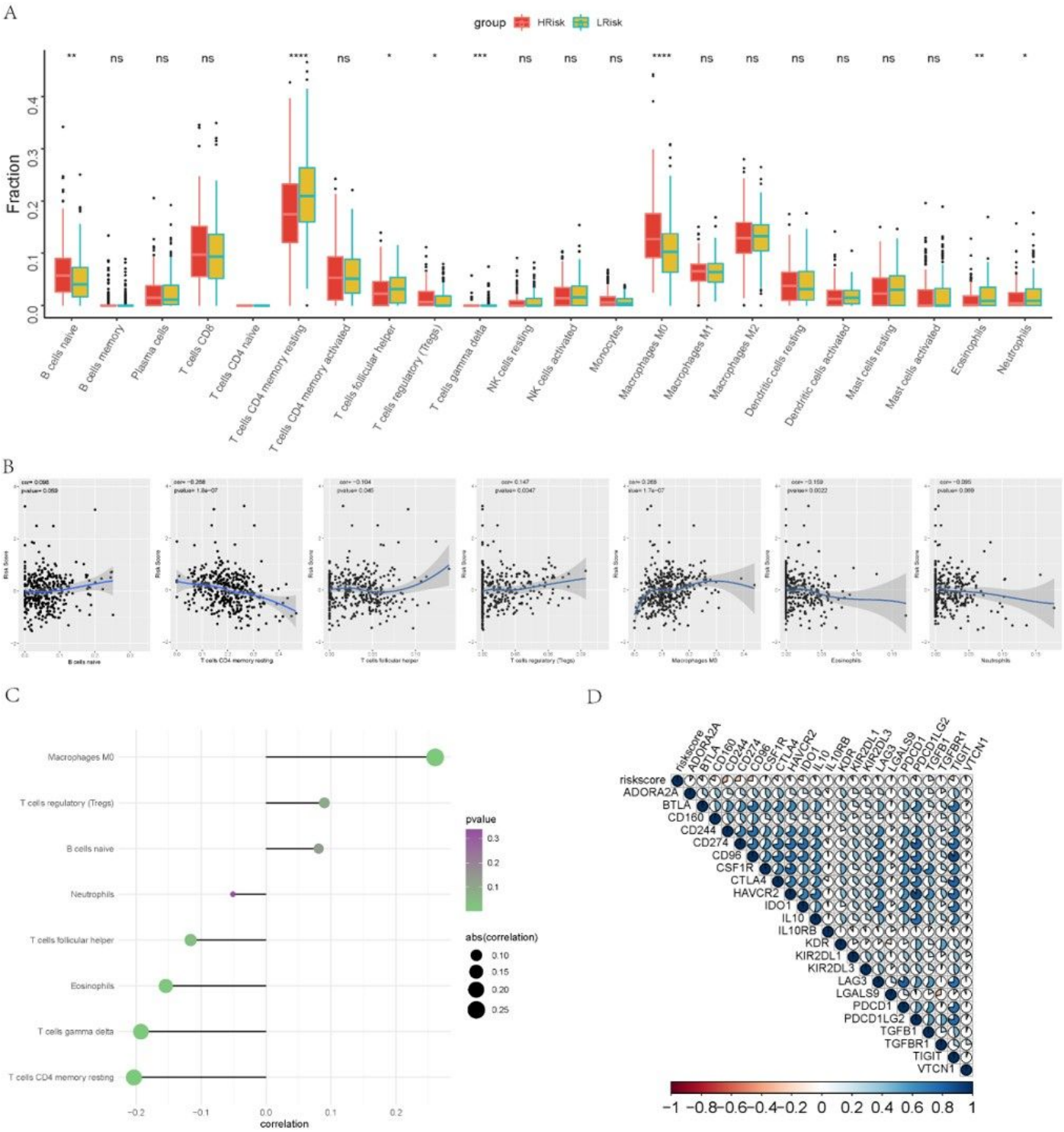

Figure 7

The immune infiltration performance of the model in the high-risk and low-risk Groups.

(A) Comparison of infiltration rate of 22 kinds of immune cells in tumor tissues of high and low risk groups. **(B)** The correlation coefficient scatter plot showed the correlation between tumor immune cell infiltration and the expression of MR-Risk score.**(C)** Immunocyte infiltration in tumor tissues with high and low prognosis scores. **(D)** Correlation between MR-risk score and gene expression of immune checkpoints in tumor tissues.