Received: 14 July 2022 | Revised: 27 January 2022 | Accepted: 31 October 2022

DOI: 10.1002/jcla.24800

RESEARCH ARTICLE

WILEY

Development of 14-gene signature prognostic model based on metastasis for colorectal cancer

Tong Li[1] | Qian Yu[1,2] | Te Liu[1,3] | Wenjing Yang[1] | Wei Chen[1] | Anli Jin[1] | Hao Wang[1] | Lin Ding[1] | Chunyan Zhang[1,4] | Baishen Pan[1] | Beili Wang[1,4] | Wei Guo[1,2,4,5]

[1]Department of Laboratory Medicine, Zhongshan Hospital, Fudan University, Shanghai, China

[2]Department of Laboratory Medicine, Wusong Branch, Zhongshan Hospital, Fudan University, Shanghai, China

[3]Shanghai Geriatric Institute of Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China

[4]Department of Laboratory Medicine, Xiamen Branch, Zhongshan Hospital, Fudan University, Xiamen, China

[5]Cancer Center, Shanghai Zhongshan Hospital, Fudan University, Shanghai, China

Correspondence
Wei Guo and Beili Wang, Department of Laboratory Medicine, Zhongshan Hospital, Fudan University, Shanghai, China.
Email: guo.wei@zs-hospital.sh.cn and wang.beili1@zs-hospital.sh.cn

Abstract

Background: Metastasis is the main cause of death of colorectal tumors, in our study a prognosis model was built by analyzing the differentially expressed genes between metastatic and non-metastatic colorectal cancer (CRC). We used this feature to predict CRC patient prognosis and explore the causes of colorectal tumor metastasis by characterizing the immune status alteration.

Methods: CRC patient data were obtained from TCGA and GEO databases. We constructed a risk prognostic model by using Cox regression and the least absolute shrinkage and selection operator (LASSO) based on CRC metastasis-related genes. We also obtained a nomogram to predict the prognosis of CRC patients. Finally, we explored the underlying mechanism of these metastasis-related genes and CRC prognosis using immune infiltration analysis and experimental verification.

Results: According to our prognostic model, in TCGA, the area under the curve (AUC) values of the training and test sets were 0.72 and 0.76, respectively, and 0.68 for the GEO external data set. This suggested that the treatment and prognosis of patients could be effectively determined. At the same time, we found that the B and T cells in both tissues and peripheral blood of high MR-risk score patients were mostly in immune static or inactivated states compared with those of low MR-risk score patients.

Conclusions: MR-risk score has a direct correlation with CRC patient prognosis. It is useful for predicting the prognosis and patient immune status for these patients.

KEYWORDS
GEO, immune infiltration, lasso regression analysis, prognostic signature, TCGA

1 | BACKGROUND

Colorectal cancer (CRC) is a malignant tumor of the digestive tract that poses a serious threat to human health.[1] According to the Global Cancer Epidemic Statistics (GLOBOCAN 2020), 1.93 million new CRC cases and 0.94 million CRC-related deaths worldwide in 2020 places this disease in third and second place, respectively, among all malignant tumor types.[2,3] CRC patients have an approximately 50%

Tong Li and Qian Yu contributed equally to this work.

chance of survival within 5 years, and more patients have metastasis or lost the opportunity of surgery at the time of diagnosis. Because the early symptoms are often not obvious, CRC patients' 5-year survival rate decreased to 12%. Furthermore, 30%–50% of patients have recurrence and metastasis after treatment, which seriously affects their prognosis and quality of life.[4,5]

Metastasis is the leading cause of death in CRC patients. Liver and lung metastases are common in these individuals, but the specific molecular mechanisms controlling this remain unclear.[6,7] Currently, the rapid development of immunotherapy and targeted therapy can improve the survival time of CRC patients to some extent.[8-10] Yet, a curative effect for patients with metastasis is not apparent.[11,12] Therefore, early indications of CRC metastasis and finding reliable markers to develop new prognosis prediction models are essential for understanding disease progression and effectively adjusting treatment strategies.

Colorectal tumor metastasis has been widely discussed and studied from a bioinformatics perspective. In similar articles on colorectal tumor metastasis, models were often constructed by comparing differentially expressed genes in metastatic colorectal tumor tissues and adjacent tissues.[13,14] Others also studied CRC metastasis mechanisms by using the widely recognized metastasis genes as gene sets.[15] However, because of this selective screening, specific artificial selection bias may be present in the model. In our study, we generated a prognosis model by directly comparing and analyzing the differential genes between metastatic and non-metastatic CRC tumors, focusing on the internal change process and gene expression characteristics of metastatic tumors relative to non-metastatic tumors. We also explored the causes of metastasis through immunological analysis. Our results may provide new insights into the role of CRC metastasis-related genes in the tumor immune microenvironment and offer a theoretical basis for predicting CRC prognosis.

2 | MATERIALS AND METHODS

2.1 | Data collection and data processing

The mRNA expression data of CRC tumor group in this study were downloaded from TCGA-COAD dataset (The Cancer Genome Atlas-colorectal adenocarcinoma). The tumor group samples were divided into two groups according to the occurrence of metastasis (64 cases of colorectal metastasis samples and 338 non-metastatic samples were finally included in the analysis), and the metastasis prognosis model was constructed by comparing the differences between the two groups. In addition, the validation data set GSE38832 (with 122 COAD patients)[16] was downloaded from the NCBI Gene Expression Omnibus (GEO) public database. The complete clinical profile and survival information of these two data sets were downloaded at the same time.

2.2 | Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment analysis

The "Limma" R package was used to integrate and standardize differentially expressed genes (DEGs). DEGs were functionally annotated by using clusterProfiler (R3.6) to comprehensively perform GO and KEGG analyses. GO and KEGG enriched pathways with both p- and q-Values less than 0.05 were considered significant categories.

2.3 | Prognostic model construction and validation

In the TCGA cohort, the DEGs related to tumor metastasis were selected by comparing genes in the tumor group with and without metastases. Firstly, the DEGs were preliminarily screened by univariate COX regression analysis ($p < 0.05$). Secondly, least absolute shrinkage and selection operator (LASSO) regression analysis was used to create the CRC prognostic risk score model ($p < 0.01$). Finally, multivariate Cox regression analysis was used to construct risk model. The risk score was calculated by using the following formula: MR-risk score = sum (expression of each gene * corresponding coefficient).

After calculating the risk score of all samples, all subjects were divided into high-risk and low-risk groups according to the median risk score. To confirm the value of the model in predicting patients' overall survival (OS) between the two risk groups, patients were assessed by Kaplan–Meier (KM) survival curves and compared by log-rank tests. We also used receiver operating characteristic (ROC) curves to study the efficacy of the model prediction. Finally, we used univariate and multivariate analyses to explore whether the MR-risk score model is an independent prognostic factor for COAD patients.

2.4 | Establishment of the nomogram

The nomogram was constructed by using data from the TCGA cohort with R package "rms" including age, gender, tumor stage, and risk score. The validity and prognostic value of the nomogram were evaluated by using calibration curves and ROC curve analysis.

2.5 | Immunocyte infiltration analysis

The RNA sequencing (RNA-seq) data of TCGA -COAD patients in the high-risk and low-risk groups were calculated by CIBERSORT (http://cibersort.stanford.edu/) [17] to obtain the relative existence ratios of 22 immunoinfiltrating cell types. Spearman algorithm was used to analyze the correlations between gene expression and the proportion of immune cell infiltration. Statistical significance was set at $p < 0.05$.

2.6 | Tissue and blood samples

All samples were obtained from Zhongshan Hospital Fudan University. Fresh tissues and blood samples were acquired from three non-metastatic and three metastatic CRC patients. All CRC clinicopathological diagnosis information were collected from pathology diagnosis reports. This study was approved by the research ethics committee of Zhongshan Hospital Fudan University, and the participants provided informed consent for using their samples in this study.

2.7 | Detection of MR-genes by qPCR

Total RNA was isolated from colorectal tumor tissues and adjacent normal tissues from three non-metastatic and three metastatic CRC patients using TRizol reagent (Invitrogen). RNA samples were reverse transcribed into cDNA by using GoScript™ Reverse Transcription Mix (Promega A2800). PCR primers used here are listed in supplemental information Table 3. RNA was analyzed using regular reverse transcription PCR (RT-PCR) with GoTaq® qPCR Master Mix (Promega A6002). All experimental procedures were performed according to the manufacturer's instructions (Table 3: RT-PCR primer sequences).

2.8 | Analysis of lymphocyte subpopulations

To examine changes in the immune status of patients with metastatic CRC relative to patients with non-metastatic CRC, T-cell and B-cell lymphocyte subpopulations were randomly assessed in three non-metastatic primary and three metastatic CRC patients. Various fluorescein-labeled monoclonal antibodies were added to each patient's peripheral blood to bind the corresponding antigens on leukocytes.

Each patient's peripheral blood was stained with fluorescent-labeled antibodies against the these cell surface markers: CD19+/CD27+/IgD- (Switched Memory B Cells), CD19+/CD27+/IgD+ (Marginal Zone B Cells), CD19+/CD27-/IgD+ (Naïve B Cells), CD3+/CD8+/CD27+/CD45RA- (Central memory CD8+ T Cells), CD3+/CD8+/CD27+/CD45RA+ (Naïve CD8+ T Cells), CD3+/CD8+/CD27-/CD45RA+ (Effector memory CD45RA re-expressing CD8+ T Cells), CD3+/CD8+/CD27-/CD45RA- (Effector memory CD8+ T Cells), CD3+/CD4+/CD27+/CD45RA- (Central memory CD4+ T Cells), CD3+/CD4+/CD27+/CD45RA+ (Naïve CD4+ T Cells), CD3+/CD4+/CD27-/CD45RA+ (Effector memory CD45RA re-expressing CD4+ T Cells), and CD3+/CD4+/CD27-/CD45RA- (Effector memory CD4+ T Cells). Storage events were gated on populations of interest (Supplemental Fig.S1).

The stained peripheral blood was hemolyzed with erythrocyte lysate (ABSIN abs9101) for 10 min at room temperature, washed with PBS, and then analyzed by flow cytometry. Flow cytometry data were collected on FACSAria II (Becton Dickinson) or Navios instrument (Beckman Coulter). Statistical analysis was performed using FlowJo v10 software (FlowJo, LLC http://www.flowjo.com/solutions/flowjo/).

2.9 | Flowchart

A flowchart outlining the study screening process is shown in Figure 1.

3 | RESULTS

3.1 | TCGA analysis revealed metastasis-related DEGs in CRC

We downloaded raw mRNA expression data (FPKM) in COAD from the TCGA database and obtained the metastasis status through the clinical indicator (M = 1: metastasis; M = 0: no metastasis). Differential expression analysis was performed between CRC patients with or without metastasis using the "Limma" package. The results suggest that a total of 552 DEGs were screened by differential gene expression value (Figure 2A). GO enrichment analyses showed that these DEGs are enriched in immune related 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) (Figure 2B). For KEGG enrichment, there are Antigen processing and presentation (antigen processing and presentation), Primary immunodeficiency (primary immune deficiency), and TNF signaling pathways. The results suggest that there may be a large difference in immune function between metastatic and non-metastatic colorectal tumors. Furthermore, many genes are also enriched in metabolic-related pathways (Figure 2C).

3.2 | Construction of a prognosis model for metastatic CRC patients

To explore the key DEGs related to CRC metastasis, the relevant data of TCGA-COAD patients were integrated in this study. Cox regression and the least absolute shrinkage and selection operator (LASSO) were used to screen feature genes related to CRC metastasis (Figure 3A–C). The results suggest that 26 prognostic genes were screened by Cox univariate regression: 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, and PGM2 (Figure 3A). After screening of characteristic variables by Lasso regression, 14 selected metastasis-related genes were obtained (Figure 3B,C). By using multivariate Cox regression analysis the regression model were constructed and the MR-risk score for each sample was obtained for subsequent analysis (MR-Risk

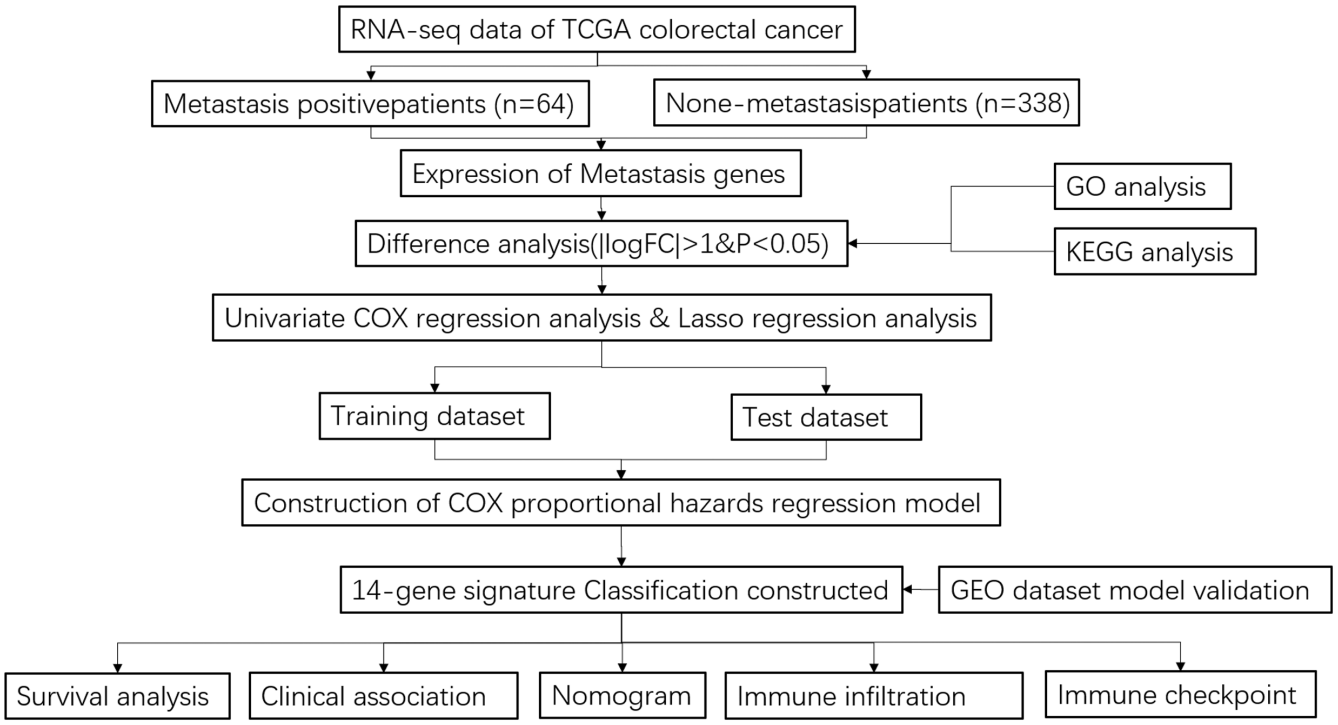

FIGURE 1 Workflow of the study.

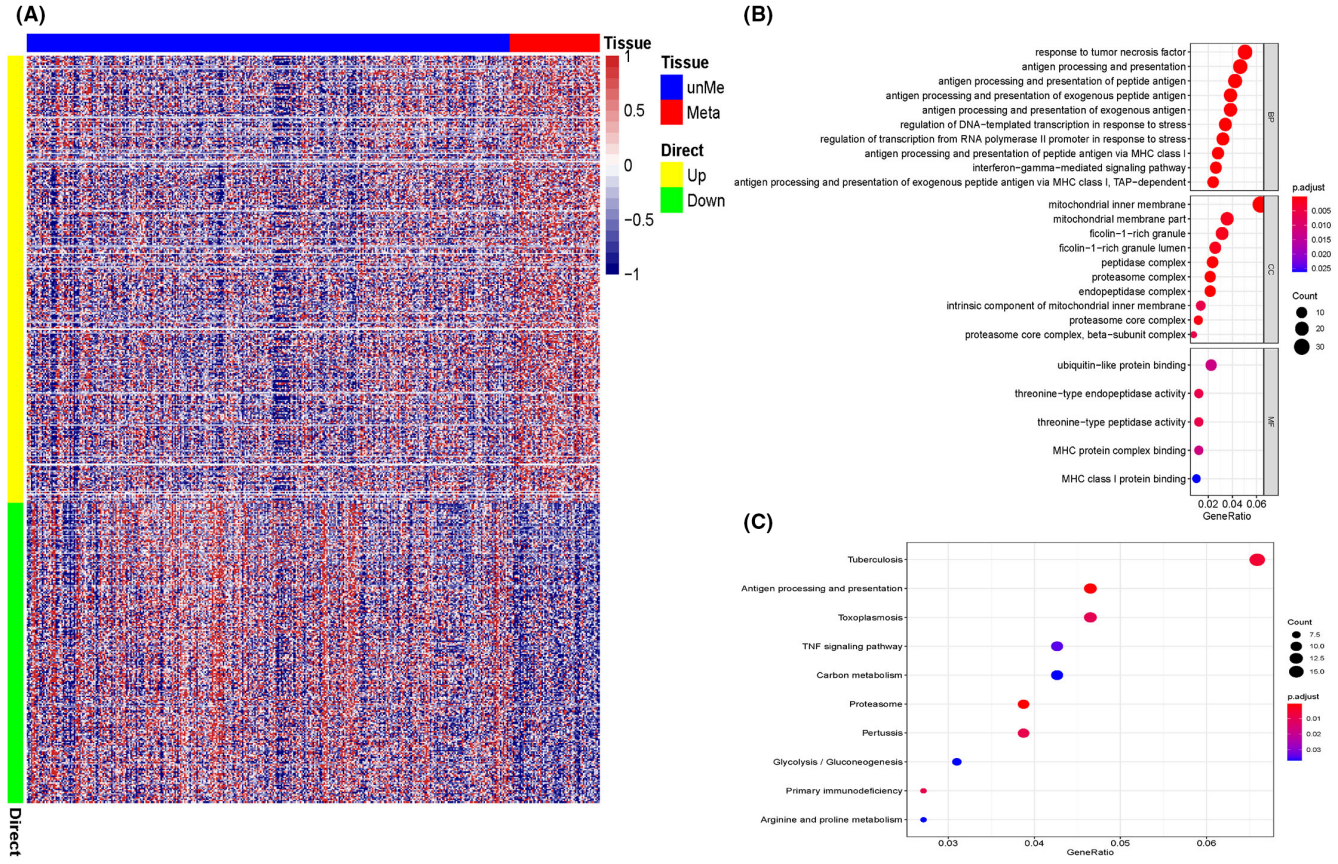

FIGURE 2 The Cancer Genome Atlas (TCGA) analysis revealed metastasis-related differentially expressed genes (DEGs) in colorectal cancer (CRC). (A) Heat map of DEGs in TCGA analysis. unMe: CRC group without metastasis; Meta: CRC group with metastasis. (B) Gene ontology (GO) pathway analysis results showed that GO-related biological functions (BF), cellular components (CC), and molecular functions (MF) were the most abundant. (C) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis showed that biological pathways were the most abundant.

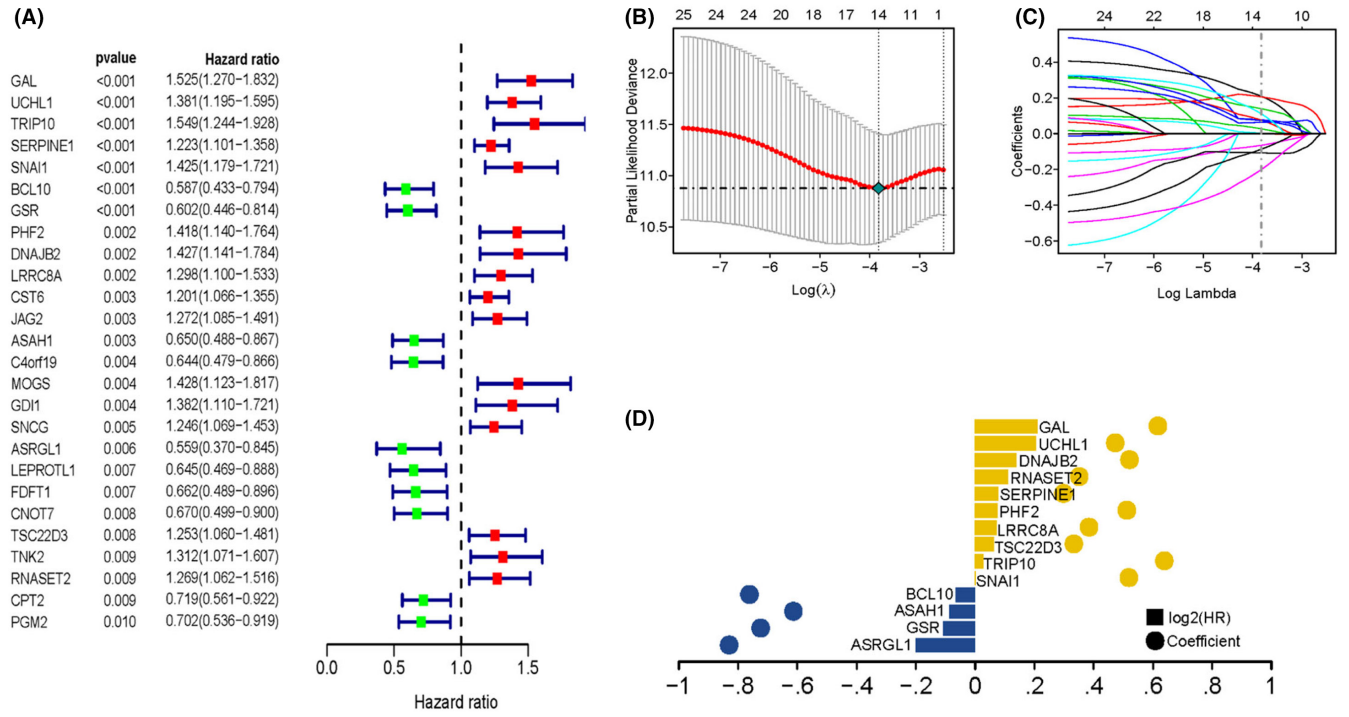

FIGURE 3 Construction of the prognosis model for metastatic colorectal cancer (CRC) patients. (A) Cox univariate regression analysis of metastasis-associated differentially expressed genes (DEGs) in CRC; ($P < 0.01$). (B) Lasso regression analysis using the maximum criterion of 10-fold cross validation. (C) The Lasso coefficient spectrum of DEGs was screened; Lasso, minimum absolute contraction, and selection operators. (D) The model coefficient diagram shows the hazard ratio (HR) values and correlation coefficients of 14 constituent genes in the prognostic model.

TABLE 1 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

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) (Figure 3D). The COX coefficient is shown in Table 1, and the Gene official full name is shown in Table 4.

3.3 | KM survival and time-varying ROC analyses were performed for the risk values of the training and test data sets

To confirm model efficacy, survival analysis was performed in the training cohort using the Kaplan–Meier method. In the TCGA-COAD dataset, patients were divided into high-risk and low-risk groups

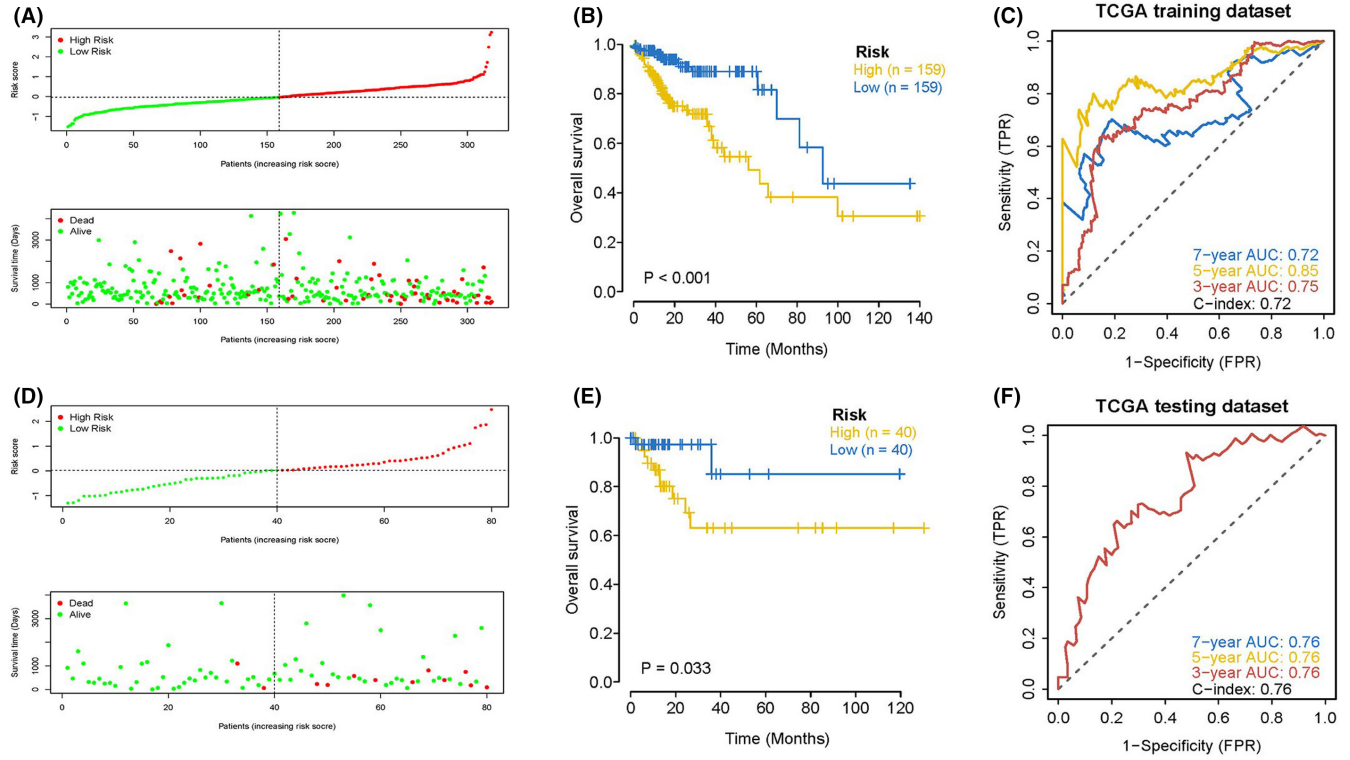

FIGURE 4 Kaplan–Meier (KM) survival and time-varying receiver operating characteristic (ROC) curve analyses were performed based on the risk values of the training and test data sets. Patients were divided into high-risk and low-risk groups based on the median MR-risk score in the training (A) and test sets (D). KM curves and time-dependent ROC analyses for overall survival (OS) prediction based on the risk score in the training (B, C) and test sets (E, F), respectively. High, high-Risk score; Low, low Risk score.

based on the median MR-risk score in the training (Figure 4A) and test sets (Figure 4D). The data show that the low-risk group has a significantly higher OS than the high-risk group (Figure 4B/E). The ROC curve analysis shows that the C-index of the training and test sets are 0.72 and 0.76, respectively (Figure 4C/F), suggesting that the model has good verification efficiency.

The basic characteristics of the training and test sets are listed in Table 2.

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

To ensure the stability of the prediction model, the RNA-Seq data (GSE38832) of COAD patients with processed survival data were retrieved from the GEO database. From this model, COAD patient clinical classifications were predicted. In the GSE38832 dataset, patients were divided into high-risk and low-risk groups based on the median MR-risk score (Figure 5A,B). The KM method was used to evaluate survival differences between the two groups. The results show that the high-risk group in the GEO external validation set had a significantly lower OS than the low-risk group (Figure 5C). To verify the accuracy of this model, we used the external data set to analyze the ROC curve of the model. The results suggest that the model had a strong predictive effect for patient prognosis (GSE38832 C-index = 0.68) (Figure 5D). Finally we also find that 14 genes were highly consistent between the GSE38832 (Figure 5E) and TCGA-COAD datasets (Figure 5F).

3.5 | Fourteen metastasis-related prognostic features associated with OS in CRC patients

To demonstrate 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 patient survival rates (Figure 6A,B). The results show that MR-risk score was an independent prognostic factor in COAD patients. Clinically, the progression and malignant degree of tumors are often categorized by staging classification to determine the appropriate surgical and drug treatment method. According to this, we divided each patient's MR-risk score value from the prediction model into the commonly used staging classification groups. The results of each clinical index group are displayed in a box diagram (Figure 6C), which shows the distribution of MR-risk score value in stage, T, M, N, and other clinical parameters ($p < 0.05$) by the rank-sum test (Kruskal test). The MR-risk score obtained with our modeling analysis has good applicability for sample grouping.

TABLE 2 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 (%)			
Stage I	12 (15.0)	61 (19.2)	0.433
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 (%)			
T1	2 (2.5)	7 (2.2)	0.367
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)	
M (%)			
M0	68 (85.0)	268 (84.3)	0.723
M1	11 (13.8)	39 (12.3)	
M1a	1 (1.2)	8 (2.5)	
M1b	0 (0.0)	3 (0.9)	
N (%)			
N0	46 (57.5)	198 (62.3)	0.143
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)	

Additionally, increasing MR-risk score values indicate that the clinical tumor staging classification tends to deteriorate with a higher possibility of tumor metastasis. We then classified the samples into high-risk and low-risk groups based on the median of MR-risk score. The regression analysis results are shown as a line chart (Figure 6D). To determine whether the prediction probability of the model is close to the empirical probability, we generated a calibration plot by using the predicted five-year and seven-year OS rates of CRC (Figure 6E). We also calculated the proportion of cancer metastasis in the high- and low-risk score subgroups, with the results suggesting that the high-risk patients have a higher probability of metastasis (Figure 6F).

TABLE 3 RT-PCR primer sequences

ASRGL1	**Gene ID: 80150**	**F**	**CCATCTCCAAGGATCGGAAGG**
		R	GACGACAGCTCCCTCTACG
GSR	Gene ID: 2936	F	CACTTGCGTGAATGTTGGATG
		R	TGGGATCACTCGTGAAGGCT
ASAH1	Gene ID: 427	F	AGATGTCATGTGGATAGGGTTCC
		R	GGGGCCAATATCTTGGTCTTG
BCL10	Gene ID: 8915	F	GTGAAGAAGGACGCCTTAGAAA
		R	TCAACAAGGGTGTCCAGACCT
SNAI1	Gene ID: 6615	F	TCGGAAGCCTAACTACAGCGA
		R	AGATGAGCATTGGCAGCGAG
TRIP10	Gene ID: 9322	F	GAAAGAACGCACCGAAGTGGA
		R	TGGAGAATCTGTACGAAGGACTG
TSC22D3	Gene ID: 1831	F	AACACCGAAATGTATCAGACCC
		R	TGTCCAGCTTAACGGAAACCA
LRRC8A	Gene ID: 56262	F	CCTGCCTTGTAAGTGGGTCAC
		R	CACAGCGTCCACGTAGTTGTA
PHF2	Gene ID: 5253	F	CTCCCCTACGACGTTACCC
		R	CAGTGGTATATGTCGATGTCGG
SERPINE1	Gene ID: 5054	F	ACCGCAACGTGGTTTTCTCA
		R	TTGAATCCCATAGCTGCTTGAAT
RNASET2	Gene ID: 8635	F	GCGAGAAAATTCAAAACGACTGT
		R	CCTTCACTTTTATCGGGCCATAG
DNAJB2	Gene ID: 3300	F	ATGGCATCCTACTACGAGATCC
		R	GAGAGCCTTGCGCCGATAC
UCHL1	Gene ID: 7345	F	AATGTCGGGTAGATGACAAGGT
		R	GGCATTCGTCCATCAAGTTCATA
GAL	Gene ID: 51083	F	CCGGCCAAGGAAAAACGAG
		R	GAGGCCATTCTTGTCGCTGA
GAPDH	Gene ID: 2597	F	GGAGCGAGATCCCTCCAAAAT
		R	GGCTGTTGTCATACTTCTCATGG

3.6 | The immune infiltration performance of the model in the high-risk and low-risk groups

The tumor microenvironment is generally composed of a variety of complex components, such as fibroblasts, immune cells, extracellular matrix, and a variety of growth factors that can promote cell growth. The microenvironment also contains cancer cells themselves, as well as related inflammatory factors that can promote tumorigenesis. The makeup of the tumor microenvironment has a significant influence on patient prognosis and clinical treatment decisions. Studying the association between the MR-Risk score and tumor immunoinvasion can further optimize the mechanistic effect of MR-Risk score on CRC prediction. The results show that the MR-risk score was positively correlated with M0 macrophages, Tregs, and naïve B cells, while negatively correlated with CD4 memory resting T cells, eosinophils, neutrophils, and follicular helper T cells (Figure 7A–C). The correlation coefficient scatter plot of gamma delta T cells was not shown because the cell infiltration rate is too low to show a linear correlation between the two MR-risk score groups. To evaluate the possible sensitivity of the MR-risk score for immunotherapy, we examined the correlation between MR-risk score and immune checkpoint expression levels in tumor tissues (Figure 7D). Interestingly, we found a negative correlation between MR-risk score and gene expression for most immune checkpoints in cancer tissues. This suggests that when a patient has a higher risk value, fewer immune checkpoint molecules are expressed on the patient's tumor cells. Therefore, for patients with tumor metastasis, the use of novel immune checkpoint inhibitors may not have a significant survival benefit.

3.7 | Validation of metastasis-related gene expression and immune status in clinical samples

Next, we clinically selected three metastatic and three non-metastatic CRC patients to detect 14 metastasis-related genes in cancer tissues and adjacent tissue samples. For each gene expression value in the tumor tissue of each patient, the expression value of the patient's

TABLE 4 Gene information

Gene	Official Full Name	URL
GAL	galanin and GMAP prepropeptide	https://www.ncbi.nlm.nih.gov/gene/51083
UCHL1	ubiquitin C-terminal hydrolase L1	https://www.ncbi.nlm.nih.gov/gene/7345
TRIP10	thyroid hormone receptor interactor 10	https://www.ncbi.nlm.nih.gov/gene/9322
SERPINE1	serpin family E member 1	https://www.ncbi.nlm.nih.gov/gene/5054
SNAI1	snail family transcriptional repressor 1	https://www.ncbi.nlm.nih.gov/gene/6615
BCL10	BCL10 immune signaling adaptor	https://www.ncbi.nlm.nih.gov/gene/8915
GSR	glutathione-disulfide reductase	https://www.ncbi.nlm.nih.gov/gene/2936
PHF2	PHD finger protein 2	https://www.ncbi.nlm.nih.gov/gene/5253
DNAJB2	DnaJ heat shock protein family (Hsp40) member B2	https://www.ncbi.nlm.nih.gov/gene/3300
LRRC8A	leucine rich repeat containing 8 VRAC subunit A	https://www.ncbi.nlm.nih.gov/gene/56262
CST6	cystatin E/M	https://www.ncbi.nlm.nih.gov/gene/1474
JAG2	jagged canonical Notch ligand 2	https://www.ncbi.nlm.nih.gov/gene/3714
ASAH1	N-acylsphingosine amidohydrolase 1	https://www.ncbi.nlm.nih.gov/gene/427
C4orf19	chromosome 4 open reading frame 19	https://www.ncbi.nlm.nih.gov/gene/55286
MOGS	mannosyl-oligosaccharide glucosidase	https://www.ncbi.nlm.nih.gov/gene/7841
GDI1	GDP dissociation inhibitor 1	https://www.ncbi.nlm.nih.gov/gene/2664
SNCG	synuclein gamma	https://www.ncbi.nlm.nih.gov/gene/6623
ASRGL1	asparaginase and isoaspartyl peptidase 1	https://www.ncbi.nlm.nih.gov/gene/80150
LEPROTL1	leptin receptor overlapping transcript like 1	https://www.ncbi.nlm.nih.gov/gene/23484
FDFT1	farnesyl-diphosphate farnesyltransferase 1	https://www.ncbi.nlm.nih.gov/gene/2222
CNOT7	CCR4-NOT transcription complex subunit 7	https://www.ncbi.nlm.nih.gov/gene/29883
TSC22D3	TSC22 domain family member 3	https://www.ncbi.nlm.nih.gov/gene/1831
TNK2	tyrosine kinase non receptor 2	https://www.ncbi.nlm.nih.gov/gene/10188
RNASET2	ribonuclease T2	https://www.ncbi.nlm.nih.gov/gene/8635
CPT2	carnitine palmitoyltransferase 2	https://www.ncbi.nlm.nih.gov/gene/1376
PGM2	phosphoglucomutase 2	https://www.ncbi.nlm.nih.gov/gene/55276

adjacent tissue was used as a reference. The results show that genes positively correlated with risk values in the formula were significantly expressed in patients with metastatic CRC compared with those without metastasis (Figure 8A). To verify any changes in the immune microenvironment of metastatic CRC patients compared with non-metastatic CRC patients, we evaluated the changes of immune status in B and T cells using peripheral blood lymphocyte fine classification in the patients above (Figure 8B). Proportion of T cell and B cell subsets were assessed by flow cytometry (Supplemental Table.S1). The results show a similar trend to the previous analysis of immune infiltration. In the peripheral blood of metastatic CRC patients with high-risk scores, compared with non-metastatic CRC patients, the functions of B and T cells were significantly weakened. This manifested in an increased number of naïve immune cells (Naïve B cells and Naïve CD4+T cells) without killing and memory functions.

4 | DISCUSSION

Globally, CRC is one of the most common types of cancer, with 1.9 million new cases and 0.9 million deaths in 2020.[18] At present, metastasis remains the leading cause of death in CRC patients. Current diagnostic imaging tools, such as enhanced computed tomography (CT), positron emission tomography (PET), and magnetic resonance imaging (MRI) scans, can detect metastatic CRC lesions.[19] 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 CRC metastasis.

Our study constructed an innovative prognostic gene risk model based on 14 CRC metastasis-related genes by analyzing the DEGs between metastatic CRC and non-metastatic CRC cases and using Lasso regression. This conclusion was verified internally using the TCGA-COAD and GEO external datasets. After analyzing the ROC curve, the model is clearly superior to other clinical factors for predicting the survival rates of CRC patients.

Through GO and KEGG enrichment analyses of signaling pathways,[20,21] we found that DEGs 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

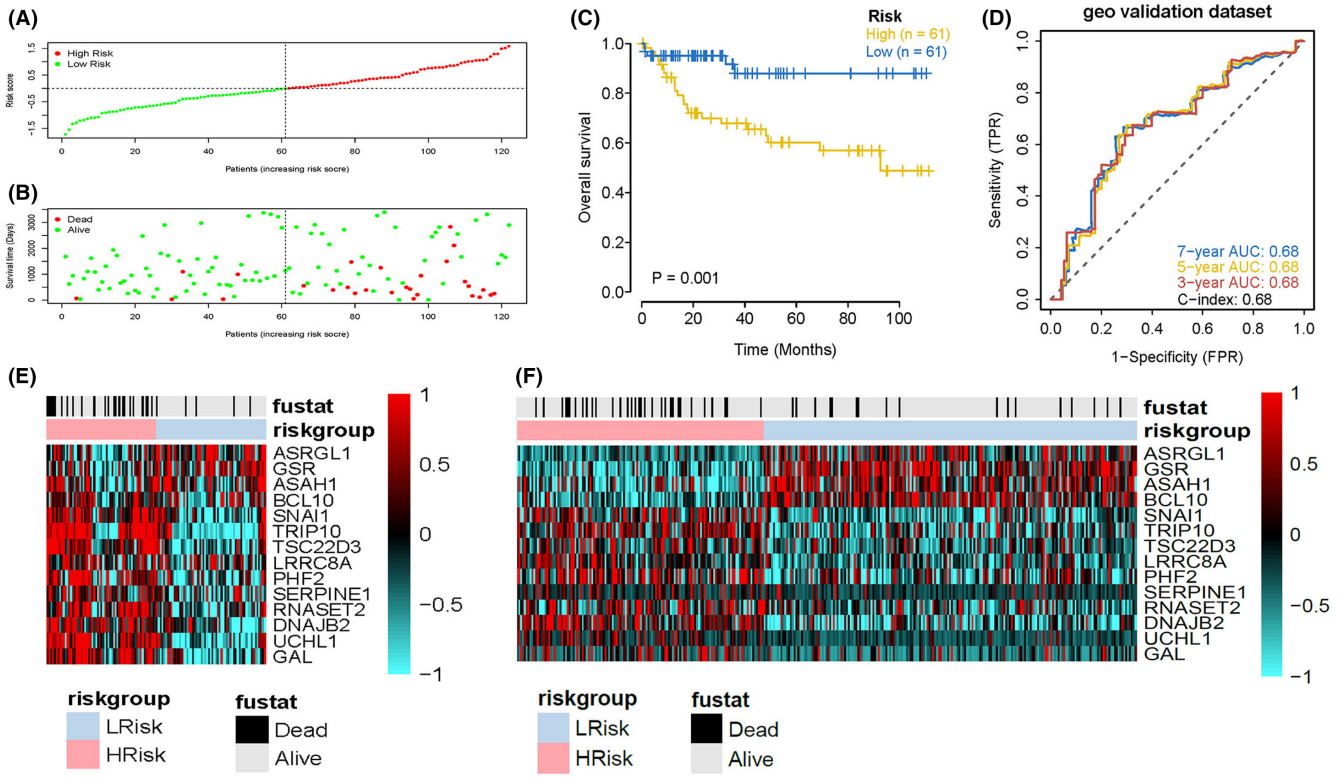

FIGURE 5 External validation of the performance of the MR-risk score model in the Gene Expression Omnibus (GEO) dataset. (A) Distribution of risk scores for the GEO dataset using the 14 metastasis-related genes in the prediction model. (B) The survival status of colorectal cancer (CRC) patients in the GEO dataset belonging to the high- and low-MR-risk score groups. (C) Kaplan–Meier (KM) survival curves of patients in the two risk score groups. (D) Time-dependent receiver operating characteristic (ROC) curve for predicting survival time with area under the curve (AUC) values in the two risk score groups. Heat map of MR-risk score model for expression levels of 14 genes in the GEO (E) and TCGA-COAD datasets (F).

(primary immunodeficiency), and TNF signaling pathway. Antigen extraction and processing is an essential immune process for triggering the T cell-mediated immune response.[22] The main role of TNF is to regulate immune cells.[23] The various functions of the DEGs in metastatic colorectal tumors cause significant changes to the specific recognition of tumors by T cells and production of antibodies by B cells.

With the gradually increasing research on the tumor microenvironment and tumor immunity, new technologies and methods will be applied to the diagnosis and treatment of metastatic CRC.[24] According to this, we analyzed the infiltrating immune cells in the tumor tissues of high-risk and low-risk score patients. The results showed that when comparing the high and low model score groups, B cells and macrophages, the two antigen-presenting cells, were both immature and unable to drive the antigen presentation as usual. Memory CD4+ T cells can rapidly respond to secondary antigen stimulation, quickly release the cytokines interferon γ (IFN-γ), interleukin 4 (IL-4), IL-5, and IL-2, and divide rapidly.[25] Among the infiltrated lymphocytes in the tumor tissues of patients with metastasis, we found that CD4+ memory T cells were resting and could not effectively mediate an immune response. Combining the above prognostic analysis and GSEA analysis results, it is clear that patients with metastatic colorectal tumors have a high MR-risk score. The poor prognosis is potentially caused by the loss of normal functions of T and B cells.

In summary, we characterized immune cell infiltration in metastatic CRC tissues: antigen-presenting cells were immature and undifferentiated and auxiliary memory cells were in a resting state and could not exert secondary immune defense functions. Therefore, we further consider whether immune checkpoint inhibitors may play vital roles in such a situation. We found that the expression of immune checkpoint genes in tumor cells of the high-risk score group were inhibited. This suggests that the immune checkpoint markers in tumor tissues of high-risk patients are reduced. The response of these patients to immunotherapy may not be obvious, which may be a possible reason for their poor prognosis.

To verify the reliability of our model, we then used the in situ carcinoma and adjacent tissues of metastatic or non-metastatic CRC patients to examine the expression of 14 metastasis-related genes. The tumor tissues of metastatic CRC patients had high expression levels of several genes in this panel. Additionally, to verify the changes in immune cell function in these patients, peripheral blood samples were subjected to a fine analysis of T and B cells. Consistent with our previous analysis, patients with metastatic colorectal

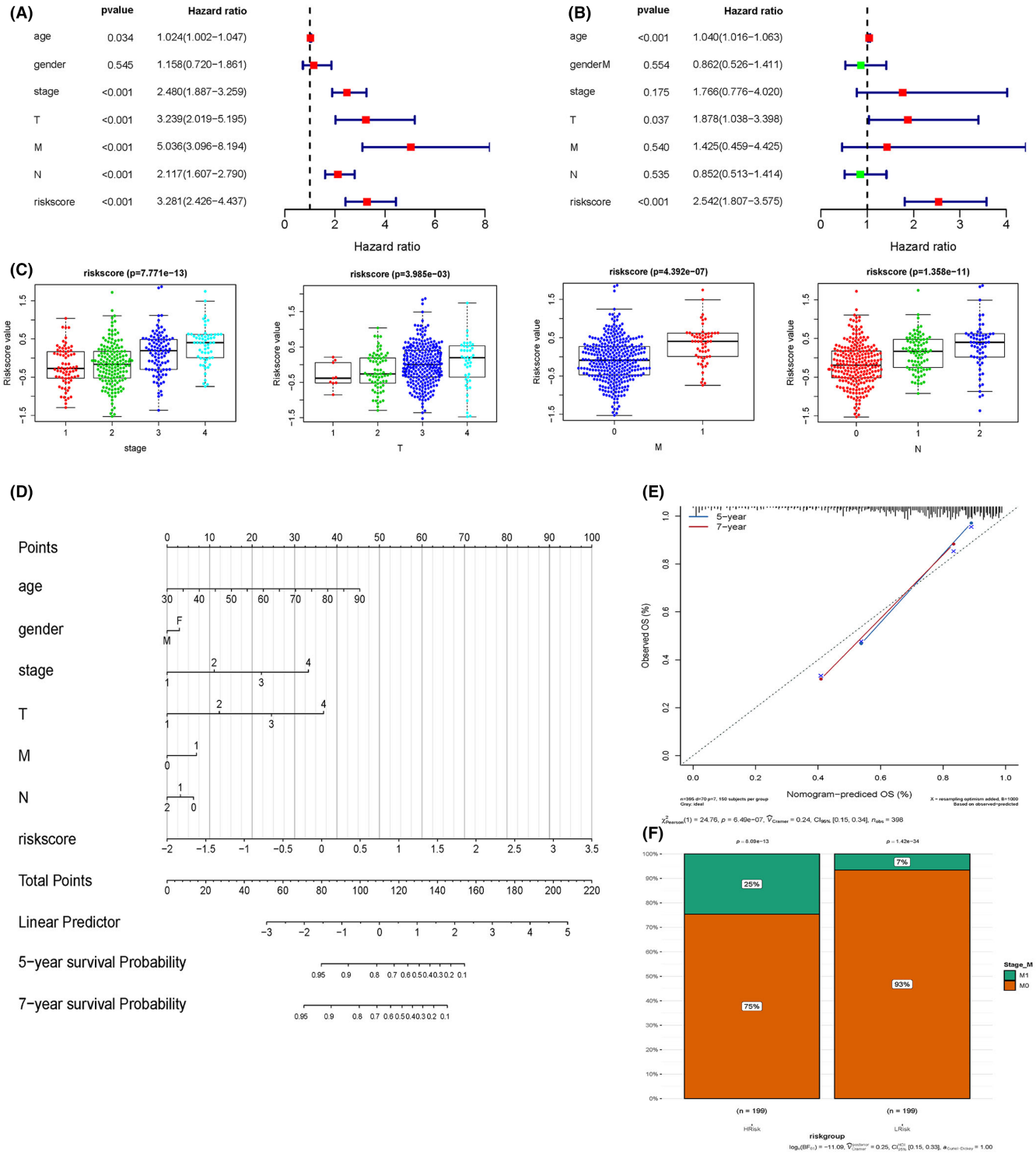

FIGURE 6 Fourteen metastasis-related prognostic features associated with overall survival (OS) in colorectal cancer (CRC) patients. (A) Univariate Cox regression analysis showed that the clinicopathological data parameters in The Cancer Genome Atlas (TCGA) cohort were associated with OS in CRC patients. (B) Multivariate Cox regression analysis showed that the clinicopathological data parameters were associated with OS 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 COAD 5 – and 7-year survival rates combined with cancer metastatic gene profiles. (E) Calibration plot of the prognostic model. The y-axis is the actual OS rate and the x-axis is the predicted OS rate. (F) The proportion of cancer metastasis in the high- and low-risk score subgroups.

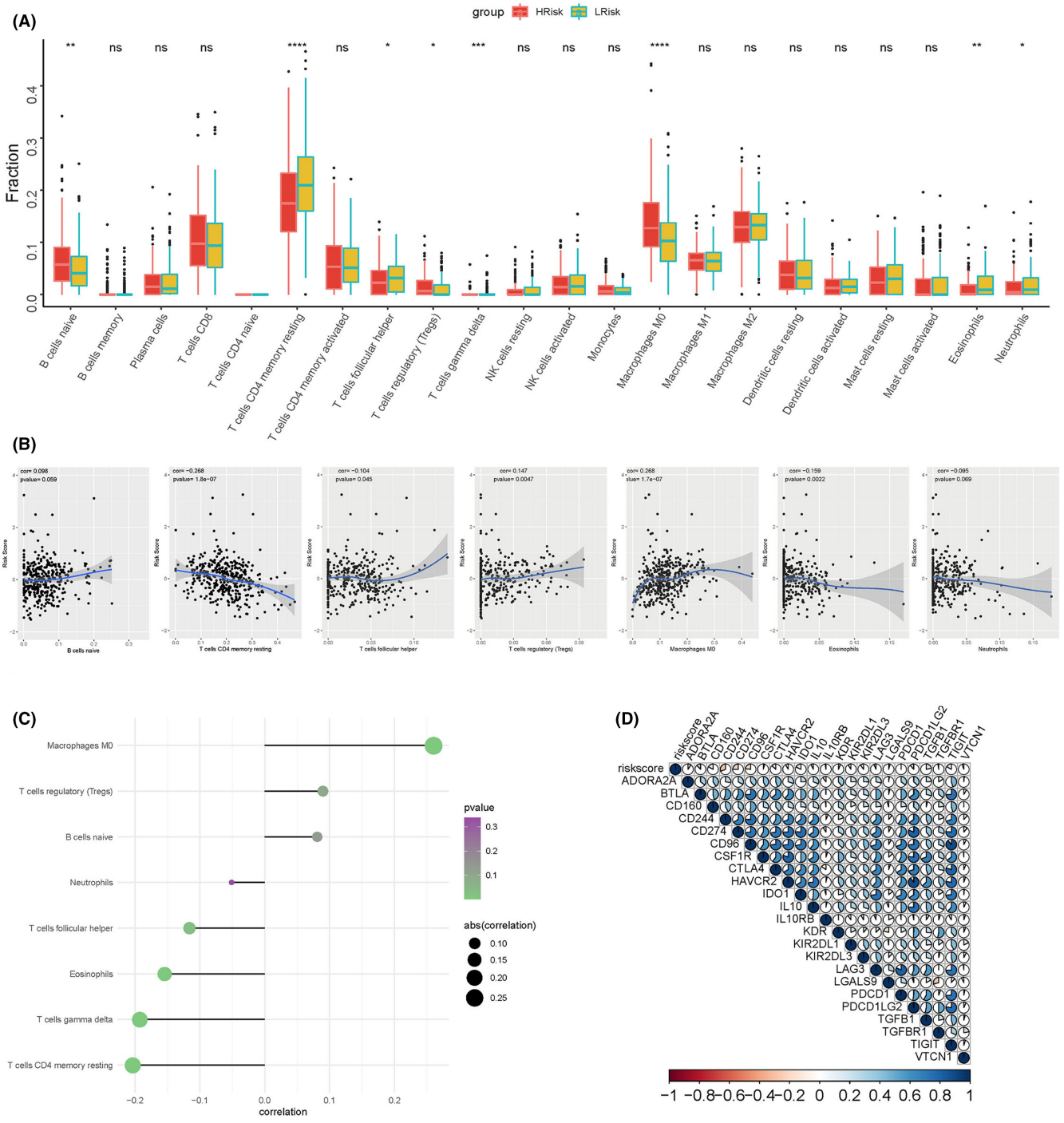

FIGURE 7 The immune infiltration performance of the model in the high-risk and low-risk groups. (A) Comparison of infiltration rates of 22 immune cell types in tumor tissues of the high- and low-risk groups. (B) The correlation coefficient scatter plot showed the correlation between tumor immune cell infiltration and the MR-Risk score. (C) Immunocyte infiltration in tumor tissues with high and low prognosis scores. (D) Correlation between MR-risk score and immune checkpoint gene expression levels in tumor tissues.

tumors had many B and T cells in the naïve state without normal immune functions. This partly explains why colorectal tumors tend to metastasize, but whether these alterations to immune cells are mediated by changes to the above 14 metastasis-related genes requires further investigation.

5 | CONCLUSIONS

We successfully established a CRC prognostic risk score model by analyzing genes specifically expressed in metastatic colorectal tumors. This model can be used to predict the survival rates of CRC

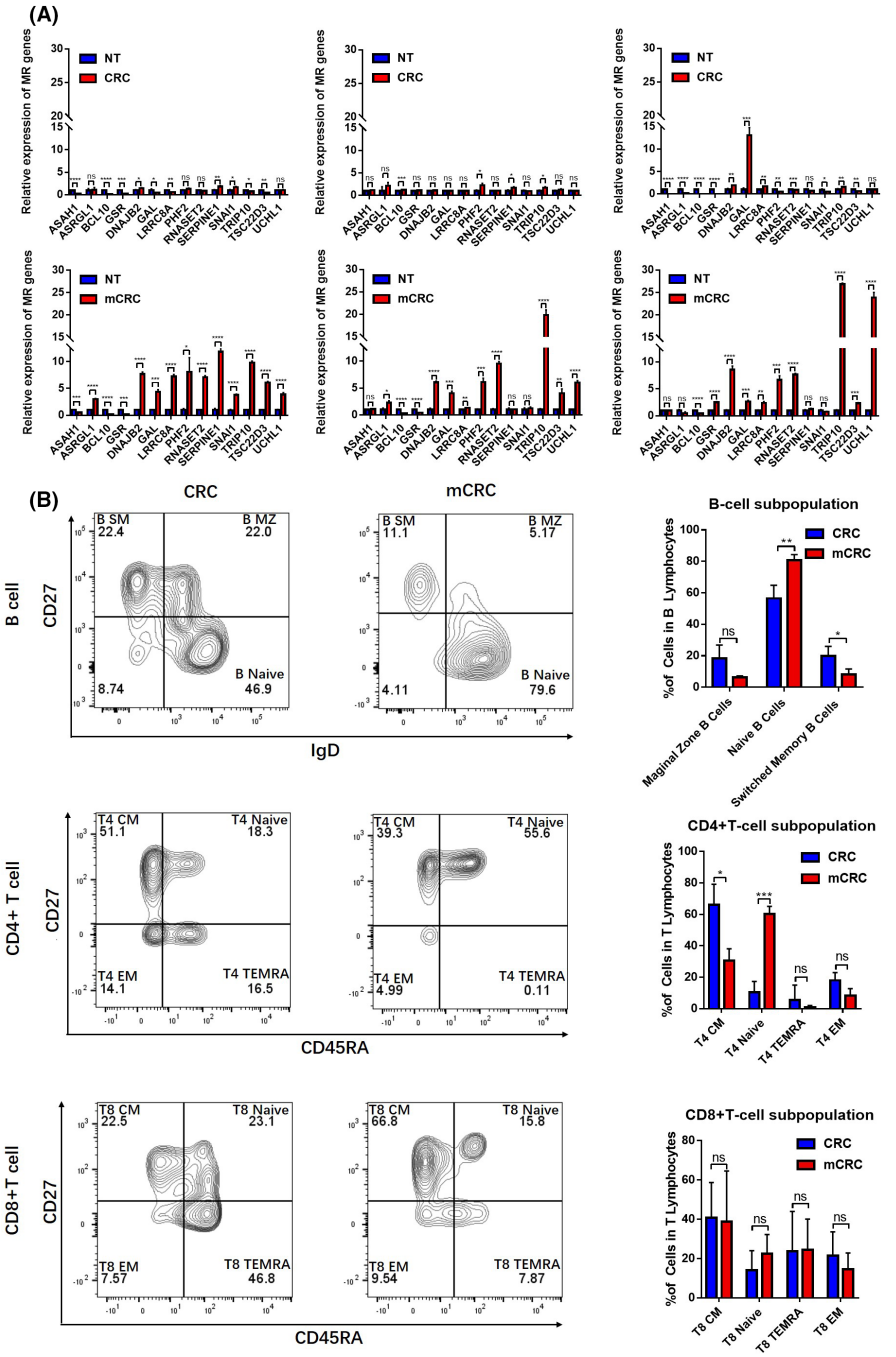

FIGURE 8 Validation of metastasis-related gene expression and immune status in clinical samples. (A) Expression levels of 14 metastasis-related genes in non-metastatic and metastatic colorectal cancer (CRC) tissues. (B) Fine subsets of B lymphocytes and T lymphocytes in patients with or without metastatic CRC.

patients. We also preliminarily examined the immunology-based mechanisms controlling colorectal tumor metastasis. Poor prognosis may be affected by the loss of T and B cell functions and a maintained naïve state. These data may help guide CRC clinical practice and individualized treatment.

FUNDING STATEMENT

This study was supported by grants from the National Natural Science Foundation of China grants (82172348, 81972000, 81902139), the Constructing Project of Clinical Key Disciplines in Shanghai (shslczdzk03302), the Projects from Excellent backbone of Zhongshan Hospital, Fudan University (2021ZSGG08), the Key Medical and Health Projects of Xiamen (YDZX20193502000002), and the Shanghai Medical Key Specialty (ZK2019B28), the Shanghai "Rising Stars of Medical Talent" Youth Development Program (Outstanding Youth Medical Talents), the Specialized Fund for the Clinical Researches of Zhongshan Hospital, Fudan University (2020ZSLC54).

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

DATA AVAILABILITY STATEMENT

Some data analyzed in the current study were obtained from publicly available 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).[15]

REFERENCES

1. Biller LH, Schrag D. Diagnosis and treatment of metastatic colorectal cancer: a review. *JAMA*. 2021;325(7):669-685.
2. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-249.
3. 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):5-14.
4. 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):3167-3175.
5. 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.
6. Zhou H, Liu Z, Wang Y, et al. Colorectal liver metastasis: molecular mechanism and interventional therapy. *Signal Transduct Target Ther*. 2022;7(1):70.
7. Malki A, ElRuz RA, Gupta I, et al. Molecular mechanisms of colon cancer progression and metastasis: recent insights and advancements. *Int J Mol Sci*. 2020;22(1):130.
8. Deng M, Gui X, Kim J, et al. LILRB4 signalling in leukaemia cells mediates T cell suppression and tumour infiltration. *Nature*. 2018;562(7728):605-609.
9. Dewson G, Silke J. The walrus and the carpenter: complex regulation of tumor immunity in colorectal cancer. *Cell*. 2018;174(1):14-16.
10. Johdi NA, Sukor NF. Colorectal cancer immunotherapy: options and strategies. *Front Immunol*. 2020;11:1624.
11. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. *CA Cancer J Clin*. 2016;66(2):115-132.
12. Ganesh K, Stadler ZK, Cercek A, et al. Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nat Rev Gastroenterol Hepatol*. 2019;16(6):361-375.
13. Gao X, Yang J. Identification of genes related to clinicopathological characteristics and prognosis of patients with colorectal cancer. *DNA Cell Biol*. 2020;39(4):690-699.
14. Gong B, Kao Y, Zhang C, et al. Identification of hub genes related to carcinogenesis and prognosis in colorectal cancer based on integrated bioinformatics. *Mediators Inflamm*. 2020;2020:5934821.
15. Tang Q, Hu X, Guo Q, et al. Discovery and validation of a novel metastasis-related lncRNA prognostic signature for colorectal cancer. *Front Genet*. 2022;13:704988.
16. Tripathi MK, Deane NG, Zhu J, et al. Nuclear factor of activated T-cell activity is associated with metastatic capacity in colon cancer. *Cancer Res*. 2014;74(23):6947-6957.
17. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453-457.
18. Xi Y, Xu P. Global colorectal cancer burden in 2020 and projections to 2040. *Transl Oncol*. 2021;14(10):101174.
19. 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):1-9.
20. Kanehisa M, Goto S, Hattori M, et al. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res*. 2006;34(Database issue):D354-D357.
21. Gene OC. The Gene ontology project in 2008. *Nucleic Acids Res*. 2008;36(Database issue):D440-D444.
22. 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):1091-1101.
23. Olszewski MB, Groot AJ, Dastych J, et al. TNF trafficking to human mast cell granules: mature chain-dependent endocytosis. *J Immunol*. 2007;178(9):5701-5709.
24. Schmitt M, Greten FR. The inflammatory pathogenesis of colorectal cancer. *Nat Rev Immunol*. 2021;21(10):653-667.
25. Zhu J, Paul WE. CD4 T cells: fates, functions, and faults. *Blood*. 2008;112(5):1557-1569.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Li T, Yu Q, Liu T, et al. Development of 14-gene signature prognostic model based on metastasis for colorectal cancer. *J Clin Lab Anal*. 2023;37:e24800. doi:10.1002/jcla.24800