Medicine®

OPEN

Integrated analysis identifies an immune-based prognostic signature for the mesenchymal identity in colorectal cancer

Bin Zhang, MD[a],[*], Liangbin Wang, MD[a], Zhixian Liu, MD[a], Bin Shao, MD[a], Wenliang Jiang, MD[a], Peng Shu, MD[b],[*]

Abstract

Background: Colorectal cancer (CRC) has been divided into 4 consensus molecular subtypes (CMSs), of which CMS4 has the mesenchymal identity and the highest relapse rate. Our goal is to develop a prognostic signature by integrating the immune system and mesenchymal modalities involved in CMS4.

Methods: The gene expression profiles collected from 5 public datasets were applied to this study, including 1280 samples totally. Network analysis was applied to integrate the mesenchymal modalities and immune signature to establish an immune-based prognostic signature for CRC (IPSCRC).

Results: We identified 6 immune genes as key factors of CMS4 and established the IPSCRC. The IPSCRC could significantly divide patients into high- and low- risk groups in terms of relapse-free survival (RFS) and overall survival (OS) and in discovery (RFS: $P < .0001$) and 4 independent validation sets (RFS range: $P = .01$ to $< .0001$; OS range: $P = .02–.0004$). After stage stratification, the IPSCRC could still distinguish poor prognosis patients in discovery (RFS: $P = .04$) and validation cohorts (RFS range: $P = .04–.007$) within stage II in terms of RFS. Further, in multivariate analysis, the IPSCRC remained an independent predictor of prognosis. Moreover, Macrophage M2 was significantly enriched in the high-risk group, while plasma cells enriched in the low-risk group.

Conclusion: We propose an immune-based signature identified by network analysis, which is a promising prognostic biomarker and help for the selection of CRC patients who might benefit from more rigorous therapies. Further prospective studies are warranted to test and validate its efficiency for clinical application.

Abbreviations: CMSs = consensus molecular subtypes, CRC = colorectal cancer, EMT = epithelial-mesenchymal transition, GEO = Gene Expression Omnibus, GEPs = gene expression profiles, GSEA = gene set enrichment analysis, IPSCRC = immune-based prognostic signature for CRC, IRGs = immune genes, MRA = master regulator analysis, OS = overall survival, RFS = relapse-free survival, TME = tumor microenvironment.

Keywords: colorectal cancer, immune-based prognostic signature, network analysis

1. Introduction

Colorectal cancer (CRC) is one of the most common types of cancer worldwide, has a high incidence in developed countries, and is the fourth leading cause of cancer-related death, with a mortality rate of about 608,000 per year.[1,2] By 2030, there will be 2.2 million new cases of CRC worldwide, resulting in 1.1 million deaths.[3] Despite new detection methods and improved treatment strategies for CRC, the 5-year survival rate is about 55%.[4] Surgical is still the priority treatment. Some patients may have local recurrence or distant metastasis after surgery. At the same time, patients with similar clinical evaluations showed different clinical progression.[5] The genetic heterogeneity of patients greatly affects intrinsic clinical diversity.[6]

In the past, researchers have identified many prognostic signature genes, as well as constructed multigene-based prognostic signatures that can divide CRC into different risk groups.[7–10] However, due to the genetic heterogeneity of CRC, the prediction effect of most markers is not as expected and cannot be widely used in clinical practice. Recently, CRC has been classified into 4 consensus molecular subtypes (CMSs) with different molecular characteristics and clinical features, of which CMS4 has

mesenchymal modalities.[11] Patients within CMS4 had the highest recurrence rate and the worst relapse-free survival (RFS). Therefore, the intrinsic characteristics of the more malignant CMS4 could potentially be used to assess the risk of recurrence in CRC patients, further guiding more accurate and targeted treatment.

There is accelerating evidence that tumor microenvironment (TME) plays an important role in the development, progression, and metastasis of cancer.[12,13] Immunotherapy for immune checkpoints is widely studied.[14,15] At the same time, previous studies have revealed that the immune system can be used to assess the prognosis of CRC.[16,17] Therefore, based on immune genes, it is possible to develop markers for the evaluation of CRC prognosis.

In this study, we applied network analysis to integrate mesenchymal modalities and immune signature genes from the ImmPort database underlying CMS4. The master regulator analysis showed that 6 immune genes were the key factors of CMS4. We pooled and analyzed 5 public cohorts containing 1280 CRC patients to develop and validate an immune gene-based prognostic signature for CRC (IPSCRC) using these 6 immune genes. Although immune prognostic markers for CRC have been reported,[18] no research has been done for risk stratification by integrating the characteristics of the mesenchymal subtype. The robustness and reliability of our model were proved by sufficient verification in 4 independent data sets. In addition, we conducted a comprehensive analysis to investigate the intrinsic biological and clinical relevance of IPSCRC. Our signature combines the mesenchymal modalities involved in CMS4 and would be used to screen CRC patients who may benefit from more rigorous treatment.

2. Materials and methods

2.1. Ethical approval

The researchers were authorized to conduct the study by the Ethics Committee of the Beilun People's Hospital, Ningbo, China. All procedures were implemented in accordance with the Declaration of Helsinki and relevant policies in China.

2.2. Patient series

The gene expression profiles (GEPs) from 5 independent datasets were comprehensively analyzed, containing 1280 cases. The complete lists of all GEPs are shown in the Supplemental Table 1, http://links.lww.com/MD/E352. These datasets involved patients from the GSE14333 (n=290),[19] the GSE17538 (n=232),[20] the GSE39582 (n=564),[21] the GSE33113 (n=90),[22] and the GSE37892 (n=104).[23] The expression data of all cohorts together with the corresponding clinical parameters were downloaded from Gene Expression Omnibus (GEO). The molecular subtyping information for all cohorts was retrieved from Guinney study.[11] The detailed clinical characteristics of the 5 datasets were described in Table 1. The design and workflow of this study were illustrated in Fig. 1A.

Table 1

Patients' characteristics in public data sets.

|	Training cohort	Validation cohorts			
	GSE14333 (n=290)	GSE17538 (n=232)	GSE39582 (n=564)	GSE33113 (n=90)	GSE37892 (n=104)
Age, y	66 (26–92)	65 (23–94)	67 (22–97)	70 (34–95)	68 (22–97)
Gender					
Female	126 (43%)	110 (47%)	256 (45%)	48 (53%)	61 (47%)
Male	164 (57%)	122 (53%)	309 (55%)	42 (47%)	69 (53%)
Location					
Left	161 (56%)		342 (61%)		72 (55%)
Right	128 (44%)		223 (39%)		57 (44%)
Unknown					1 (1%)
Stage					
I	44 (15%)	28 (12%)	32 (6%)		
II	94 (32%)	72 (31%)	264 (47%)	90 (100%)	73 (56%)
III	91 (31%)	76 (33%)	205 (36%)		57 (44%)
IV	61 (21%)	56 (24%)	60 (11%)		
Unknown			4 (1%)		
MSI					
MSI			75 (13%)	25 (28%)	
MSS			443 (78%)	65 (72%)	
Unknown			47 (8%)		
BRAF					
Wild			460 (81%)	73 (81%)	
Mut			51 (9%)	17 (19%)	
Unknown			54 (10%)		
Subtype					
CMS1	63 (22%)	43 (19%)	91 (16%)	20 (22%)	11 (8%)
CMS2	119 (41%)	72 (31%)	232 (41%)	34 (38%)	49 (38%)
CMS3	50 (17%)	36 (16%)	69 (12%)	10 (11%)	19 (15%)
CMS4	58 (20%)	47 (20%)	126 (22%)	21 (23%)	39 (30%)
Unknown		34 (15%)	47 (8%)	5 (6%)	12 (9%)

CMSs = consensus molecular subtypes.
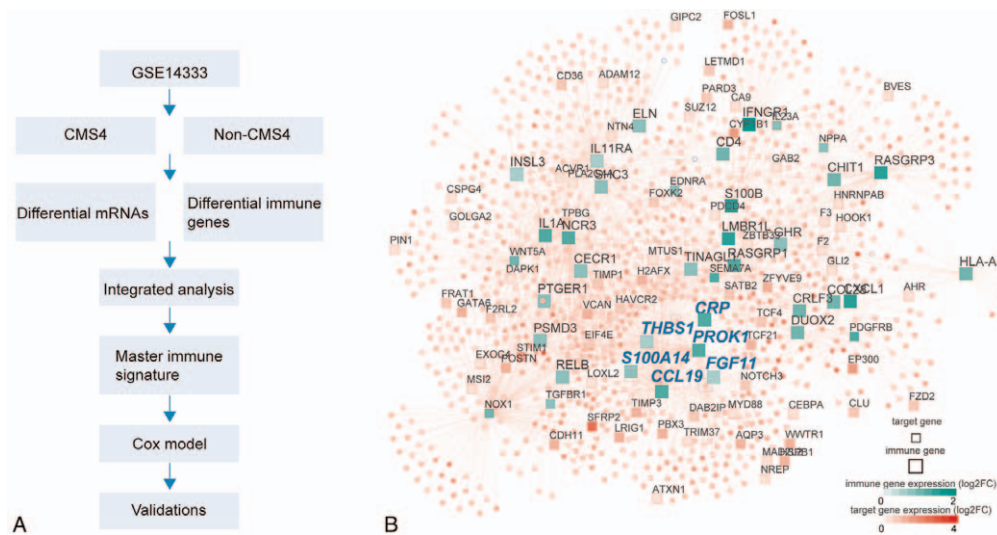
Figure 1. Network inference identifies 6 immune signature genes as key regulators of the mesenchymal subtype in CRC. Study design (A). The integrated network displays the relationships between the 6 immune signature genes and the regulated EMT signature genes (B). CRC = colorectal cancer; EMT = epithelial-mesenchymal transition.

2.3. Expression data preprocessing

GEPs were downloaded from GEO by "GEOquery" (R package, version 1.0.7)[24] and normalized with the robust multiarray analysis (RMA). For each cohort, the GEPs were collapsed from probe IDs to genes symbols, if multiple probe IDs correspond to the same genes symbol, the one with the highest mean value was kept as the representative of the corresponding gene.[25]

2.4. Integrated network analysis

Immune genes (IRGs) were downloaded from the ImmPort database.[26] IRGs measured by all cohorts were kept. Network analysis was applied to integrate mesenchymal modalities and immune genes underlying CMS4. Together, we used the GSE14333 dataset as the training cohort. Fourty five immune genes (log2FC > 0.75, $P < .05$) and 1319 target genes (log2FC > 0.5, $P < .05$) differentially expressed by comparing CMS4 with the other 3 subtypes (CMS1, CMS2, and CMS3). Integrated network analysis was performed by the "RTN" (R package, version 2.10.0).[27] Master regulator analysis (MRA) was done to examine significantly overrepresented epithelial-mesenchymal transition (EMT) genes[28] within each immune gene's regulon. Six immune genes of top significance were kept as the key factors of CMS4.

2.5. Development of the immune-based prognostic signature for CRC (IPSCRC)

Six immune genes are differentially up-expressed in the poorest survival subtype and are the master regulatory factors of the mesenchymal subtype-specific genes (including EMT genes). The Cox proportional-hazards model was applied to test their association with relapse-free survival. Based on these 6 immune genes, we develop a Cox-model named the immune-based prognostic signature for CRC (IPSCRC) as follows: risk score = $(0.2304 \times PROK1) + (0.2989 \times THBS1) + (0.3787 \times FGF11) + (0.2587 \times CRP) + (-0.1192 \times S100A14) + (-0.3350 \times CCL19)$.

2.6. Validation of the IPSCRC

The IPSCRC score was further evaluated in the 4 independent validation cohorts in terms of RFS and OS by the log-rank test, respectively. The IPSCRC then was evaluated with other clinical parameters in the uni- and multivariate Cox analysis. In the multivariable Cox regression, sex, tumor stages, tumor locations, microsatellite instability (MSI) status, and B-Raf proto-oncogene serine/threonine kinase (BRAF) mutation status were included as covariates.

2.7. Profiling of immune cells infiltration

To analyze the immunobiological characteristics of high- and low-risk groups, we used CIBERSORT,[29] to characterize immune cells' abundance of tumor tissue GEPs. Based on a set of reference immune cell GEPs, CIBERSORT used support vector regression[29] to deconvolute tumor tissue gene expression profile with each type of immune cell enrichment. More specifically, standardized gene expression profiles were submitted to the CIBERSORT Web portal (http://cibersort.stanford.edu/) with 1000 permutations. For each sample, CIBERSORT quantified the relative proportions of 22 infiltrated immune cell types.

2.8. Gene set enrichment analysis (GSEA)

GSEA[30] was conducted using "fgsea" (Bioconductor package, version 1.12.0) with 1000 permutations. Gene sets were retrieved from the Molecular Signature Database (MSigDB hallmark and kegg, version 7).[30] A P-value below .05 was used to choose significant gene sets.

2.9. Statistical analysis

Continuous variables were compared using the Wilcoxon signed-rank test or Kruskal-Wallis rank-sum test. Kaplan-Meier analysis was performed using the log-rank test using "survival" (R package, version 2.41.3). Uni- and multivariable analyses were conducted by the Cox proportional hazards model. For all tests, a
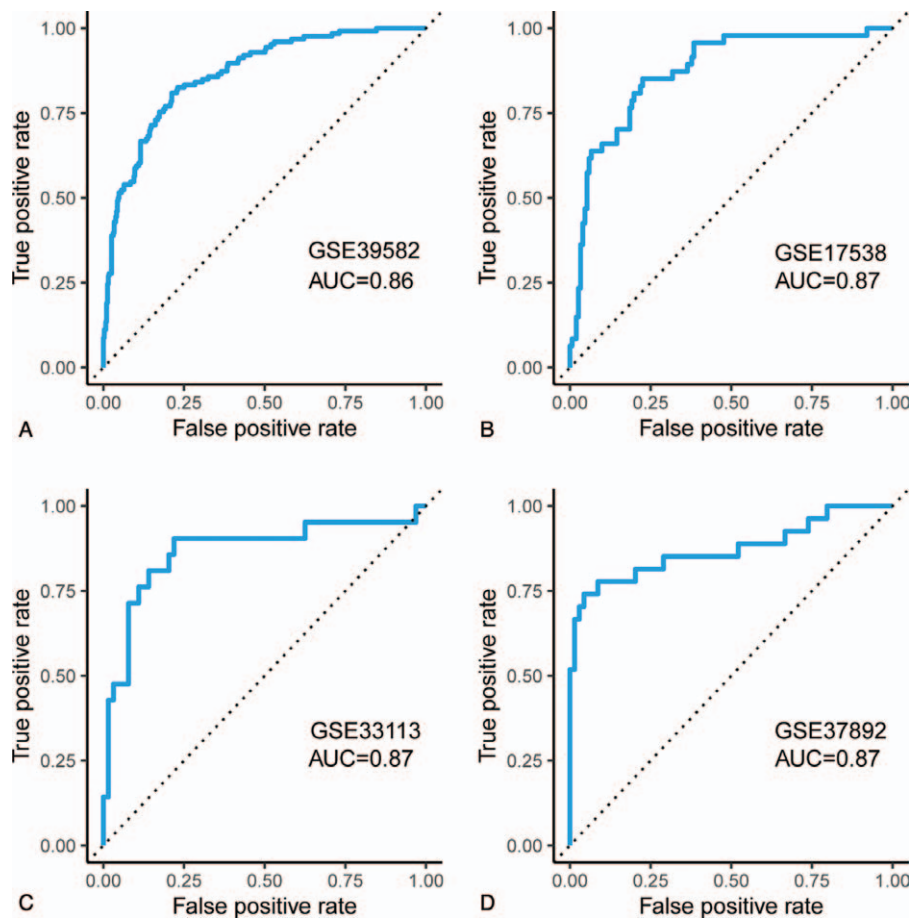
Figure 2. ROC curves of the IPSCRC. ROC curves for the detection efficiency of CMS4 in GSE39582 (A), GSE17538 (B), GSE33113 (C), and GSE37892 (D). ROC = receiver operating characteristic curve, IPSCRC = immune-based prognostic signature for CRC.

P-value below .05 was used to choose significant gene sets. Statistical significance is presented as following *P < .05, **P < .01, ***P < .001. All the statistical tests were conducted using R (version 3.6.1).

3. Results

3.1. Integrative analysis reveals 6 immune genes as master regulators for CMS4 of CRC

CRC is a molecularly heterogeneous disease. In recent studies,[11] 4 consensus molecular subtypes (CMSs) have been identified, among which CMS4 has the highest recurrence rate and the worst relapse-free survival (RFS) (Supplemental Fig. 1, http://links.lww. com/MD/E347). To investigate the immune system role underlying CMS4, a total of 1280 patients with CRC from 5 independent public cohorts were included (Table 1). We applied network analysis to integrate mesenchymal modalities and immune genes in the GSE14333 cohort (Fig. 1A). The networks consist of immune genes that are significantly up expressed in CMS4 compared with the other 3 subtypes and were found to regulate most of the mesenchymal specific target genes (Fig. 1B). Master regulator analysis (MRA) identified 6 immune genes (PROK1, THBS1, FGF11, CRP, S100A14, CCL19) as the key factors of CMS4 (Supplemental Table 2, http://links.lww.com/MD/E353).

These 6 immune genes are significantly up-expressed in CMS4 in the training cohort and validation dataset containing molecular subtyping information (Supplemental Fig. 2, http://links.lww. com/MD/E348). Results from the univariable Cox proportional analysis demonstrated strong prognostic values of the 6 immune genes for RFS (Supplemental Fig. 3, http://links.lww.com/MD/ E349). Therefore, these 6 immune genes are key factors of the mesenchymal modalities and can be potentially applied for risk assessment of CRC patients.

3.2. Development of the immune-based prognostic signature for CRC (IPSCRC)

Using the GSE14333 cohort as the training set, we defined the IPSCRC using Lasso Cox proportional hazards regression of these 6 immune genes: risk score = $(0.2304 \times PROK1) + (0.2989 \times THBS1) + (0.3787 \times FGF11) + (0.2587 \times CRP) + (-0.1192 \times S100A14) + (-0.3350 \times CCL19)$. Risk scores were calculated in all the training and validation cohorts (Supplemental Table 3, http://links.lww.com/MD/E354). For detecting CMS4, the IPSCRC achieved high AUC values for all validation cohorts (Fig. 2). The upper-quartile risk value was set as the cut-off to separate patients into high- and low-risk groups across all datasets. In the training set, the high-risk group displayed a worse RFS (hazard ratio [HR]: 3.65, 95% confidence interval
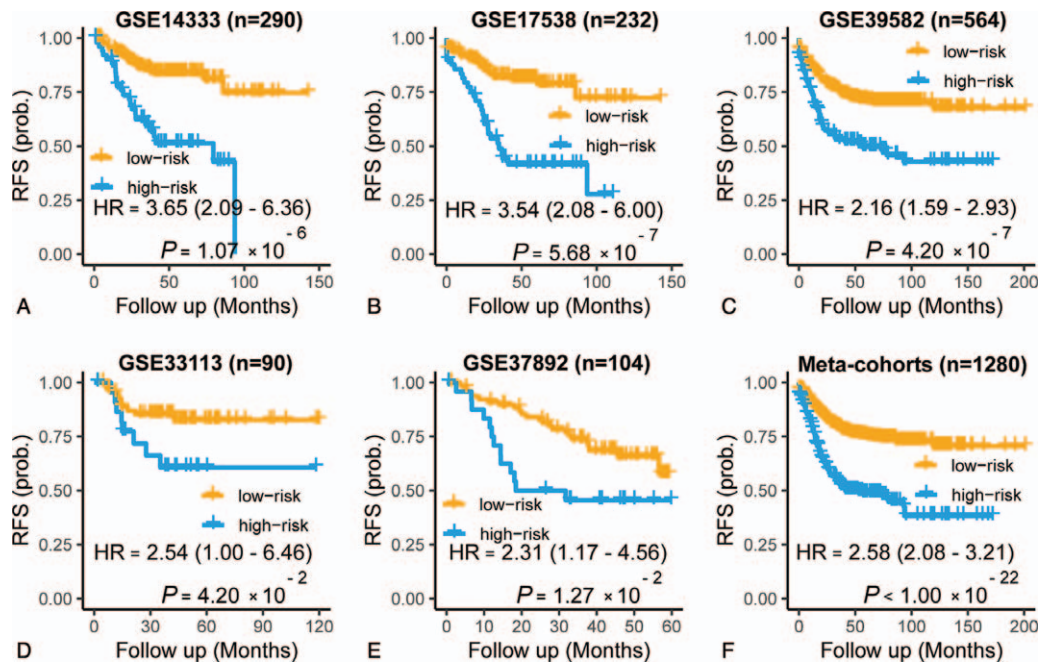
Figure 3. Kaplan-Meier plots showing differences in relapse-free survival among different risk groups within GSE14333 (A), GSE17538 (B), GSE39582 (C), GSE33113 (D), GSE37892 (E), and (F) the combination of all cohorts. *P*-values are based on log-rank tests.

[CI]: 2.09–6.36; $P = 1.07 \times 10^{-6}$) (Fig. 3A and Supplemental Table 3, http://links.lww.com/MD/E354). When considering patients with stage II CRC only, the IPSCRC remained prognostic in terms of RFS for the training set (HR: 2.92, 95% CI: 1.01–8.43; $P = 3.74 \times 10^{-2}$) (Fig. 4A and Supplemental Table 3, http://links.lww.com/MD/E354).

3.3. Validation of the IPSCRC

To verify the prognostic power of the IPSCRC, we calculated the survival difference within the 2 risk groups in 4 validation cohorts. As expected, the IPSCRC significantly stratified patients into high- and low-risk groups in terms of RFS (HR range: 2.16 [95% CI: 1.59–2.93; $P = 4.20 \times 10^{-7}$] to 3.54 [95% CI: 2.08–
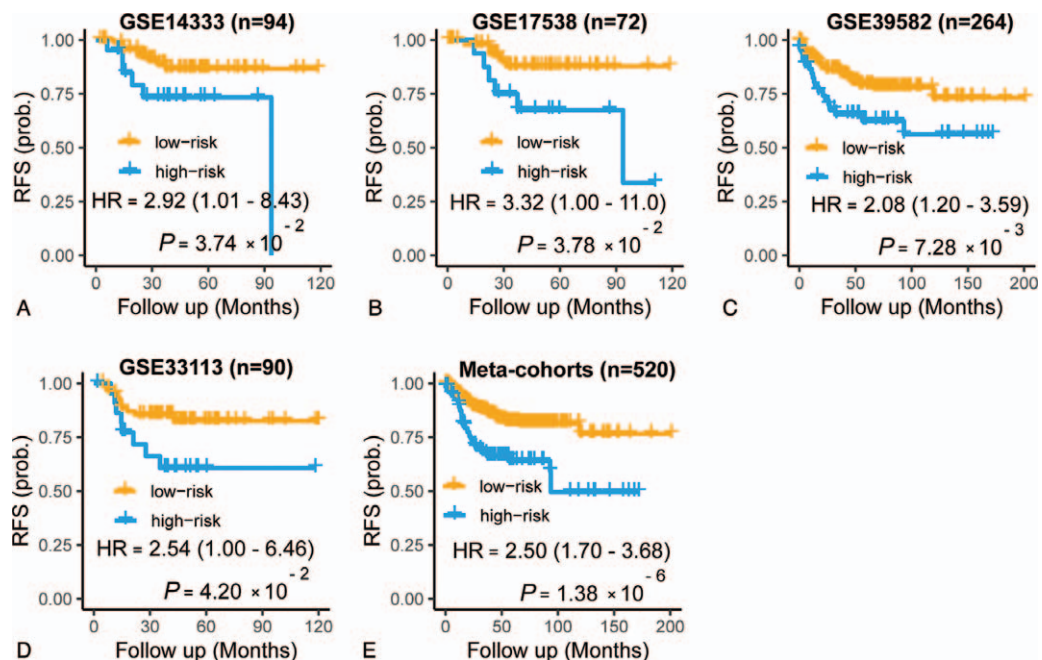

Figure 4. Kaplan-Meier plots showing differences in relapse-free survival among different risk groups for stage II CRC within GSE14333 (A), GSE17538 (B), GSE39582 (C), GSE33113 (D), and (E) the combination of these 4 cohorts. *P*-values are based on log-rank tests. CRC = colorectal cancer.
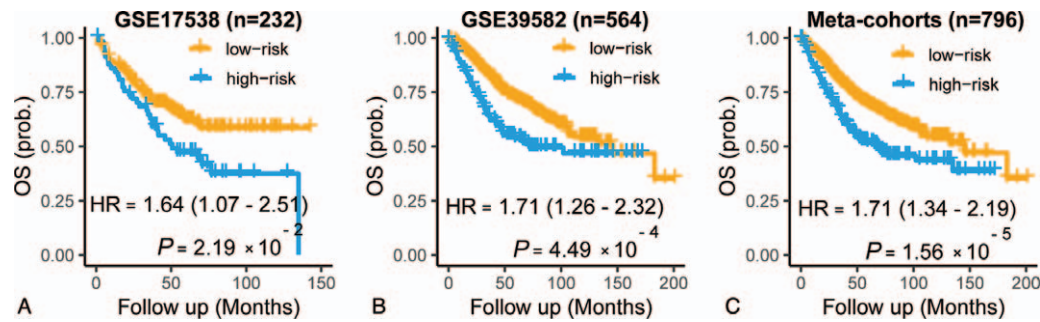
Figure 5. Kaplan-Meier plots showing differences in overall survival among different risk groups within GSE17538 (A), GSE39582 (B), and (C) the combination of these 2 cohorts. P-values are based on log-rank tests.

6.00; $P = 5.68 \times 10^{-7}$]) (Fig. 3B–E and Supplemental Table 3, http://links.lww.com/MD/E354) and OS (HR range: 1.64 [95% CI: 1.07–2.51; $P = 2.19 \times 10^{-2}$] to 1.71 [95% CI: 1.26–2.32; $P = 4.49 \times 10^{-4}$]) (Fig. 5A–B and Supplemental Table 3, http://links.lww.com/MD/E354) in the 4 validation cohorts. When considering patients with stage II CRC only, the IPSCRC remained prognostic in terms of RFS for validation cohorts (HR range: 2.08 [95% CI: 1.20–3.59; $P = 7.28 \times 10^{-3}$] to 3.32 [95% CI: 1.00–11.0; $P = 3.78 \times 10^{-2}$]) (Fig. 4B–D and Supplemental Table 3, http://links.lww.com/MD/E354). In the meta-analysis for all datasets, the prognostic effects of the IPSCRC are more obvious in terms of RFS (HR: 2.58, 95% CI: 2.08–3.21; $P < 1.00 \times 10^{-22}$) (Fig. 3F), stage II RFS (HR: 2.50, 95% CI: 1.70–3.68; $P = 1.38 \times 10^{-6}$) (Fig. 4E), and OS (HR: 1.71, 95% CI: 1.34–2.19; $P = 1.56 \times 10^{-5}$) (Fig. 5C). Furthermore, it remains an independent predictor of prognosis in the uni- and multivariate Cox model, after adjusting for sex, tumor stages, tumor locations, MSI status, and BRAF mutation status (Table 2).

3.4. In silico functional assessment of the IPSCRC

To gain insight into the biological differences between risk groups, we performed immune cell infiltration and GSEA analyses. We observed a significantly higher proportion of Macrophage M2 in the high-risk group and a significantly higher enrichment of plasma cells in the low-risk group (Fig. 6A–B). Furthermore, these 2 risk groups' specific immune cell infiltration was also validated in validation cohorts (Supplemental Fig. 4, http://links.lww.com/MD/E350). Enrichment analysis between high- and low- risk groups identified that many mesenchymal phenotype-related pathways, including the TGF-beta signaling,

epithelial-mesenchymal transition, and focal adhesion, were positively enriched in the high-risk group (Supplemental Fig. 5, http://links.lww.com/MD/E351 and Supplemental Table 4, http://links.lww.com/MD/E355). When compared with a clinically applicable and commercialized biomarker, the IPSCRC achieved a higher C-index compared with Oncotype Dx Colon Cancer[31] in training and validation cohorts (Fig. 7).

4. Discussion

Colorectal cancer (CRC) is one of the most common types of cancer worldwide and the fourth leading cause of cancer-related death.[1,2] In the past, researchers have constructed numerous multigene-based prognostic signatures that can divide CRC into different risk groups.[7–10] However, due to the genetic heterogeneity of CRC, the prediction effect of most markers is not as expected. Therefore, a new signature that can accurately recognize patients with poor CRC prognosis is urgently needed to give more rigorous treatments.

CRC has been classified into 4 consensus molecular subtypes (CMSs) with different molecular characteristics and clinical features, of which CMS4 has mesenchymal modalities.[11] Prognostic signature screened based on molecular portraits specific to the worst prognosis subtype may be used for risk stratification of CRC patients.[32,33] Recent studies have shown that the tumor microenvironment plays an important role in the occurrence and development of tumors.[34] Previous studies have revealed that the immune system can be used to assess the prognosis of CRC.[16,17] In this study, we established an immune gene-based prognostic signature for CRC (IPSCRC) by integrating mesenchymal modalities and the immune system underlying

Table 2

Univariate and multivariate analysis of immune signature and clinicopathological factors.

| | Meta-validation cohorts | | | |
| | Univariate | | Multivariate | |
	HR (95% CI)	P	HR (95% CI)	P
Gender (male vs female)	1.09 (0.87–1.39)	0.43	1.07 (0.76–1.49)	0.71
Stage (III and IV vs I and II)	3.13 (2.45–3.99)	2E–16	2.42 (1.69–3.47)	1.5E–06
Location (left vs right)	1.24 (0.94–1.64)	0.14	1.26 (0.87–1.81)	2.2E–01
MSI (MSI vs MSS)	2.48 (1.44–4.28)	1.00E–03	3.30 (1.54–7.06)	2.0E–03
BRAF (mut and wild)	1.01 (0.62–1.65)	0.97	1.67 (0.84–3.31)	0.14
Immune signature (high vs low risk)	2.42 (1.91–3.07)	2.33E–13	2.10 (1.48–2.97)	3.0E–05

BRAF = B-Raf proto-oncogene serine/threonine kinase, MSI = microsatellite instability.
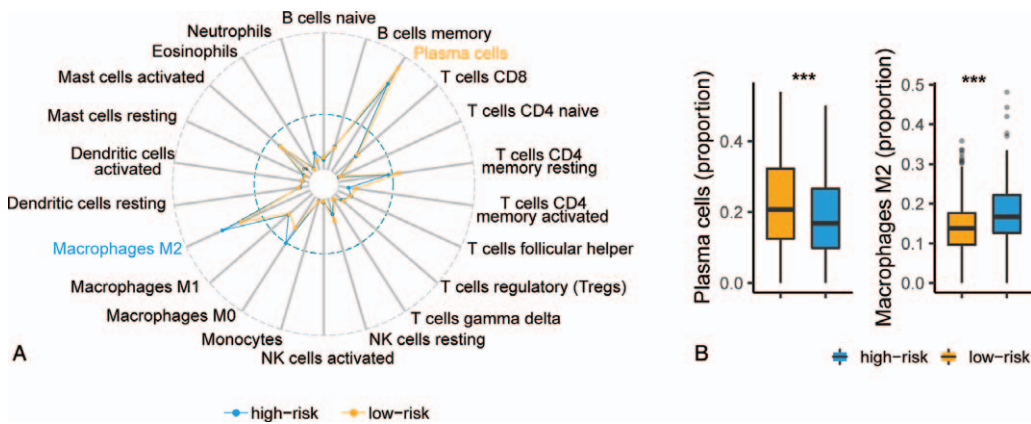
Figure 6. Immune infiltration status between the 2 immune risk groups. (A) Immune infiltration status for different immune risk groups. (B) The proportion level of plasma and Macrophage M2 for different immune risk groups. For every immune cell subset, the Wilcoxon test *P*-value comparing the high- versus low- immune risk groups are shown. ***$P < .001$.

CMS4 and validated it in 4 independent validation cohorts. The large sample size provided sufficient validation for the IPSCRC and makes it more robust. To our knowledge, no research has been done for risk stratification by integrating the immune system and the characteristics of CMS4 in CRC.

The IPSCRC was constructed by 6 immune genes as the key factors of CMS4 and could stratify patients into different risk groups. Within these 6 immune genes, positive expression PROK1 was significantly enriched CRC patients with high recurrence rate, lymphatic invasion, and lymph node metastasis.[35] THBS1 over-expression was significantly associated with regional lymph node involvement and poor survival in esophageal squamous cell carcinoma (ESCC) patients.[36] High expression of CRP is associated with poorer disease-specific survival in CRC patients.[37] S100A14 promotes hepatocellular carcinoma invasion and migration.[38] Elevated expression of CCL19 correlates with progression in cervical cancer.[39] The defined high-risk group showed a worse RFS and OS than the low-risk group. The IPSCRC remained an independent prognostic predictor in multivariate Cox proportional hazards analysis after adjusting for other clinical factors. Previous studies have described evaluated plasma cells counts in CRC,[40] while

Macrophage M2 indicates a poor prognosis in ovarian cancer.[41] We observed a significantly higher proportion of Macrophage M2 in the high-risk group and a significantly higher enrichment of plasma cells in the low-risk group. Moreover, some mesenchymal phenotype-related pathways, such as EMT, TGF-beta, and focal adhesion, were positively enriched in the high-risk group. Our findings inferred the important role of IPSCRC in tumor invasion and therefore, could sever as a robust prognostic signature in CRC.

This study still has some limitations. First of all, the prognostic signature was screened from gene expression profiles generated from microarray platforms, which are expensive, difficult to operate and involves professional bioinformatics expertise, so it is difficult to be popularized in daily clinical application. Second, the training and validation data sets were all from retrospective studies in the study, including fresh frozen samples. Therefore, the efficiency and stability of FFPE samples are still in doubt. In the following improvement process, more datasets containing more clinical characteristics need to be included for more extensive screening and validation.

Taken together, our network analysis established an immune gene-based signature, which could effectively predict CRC
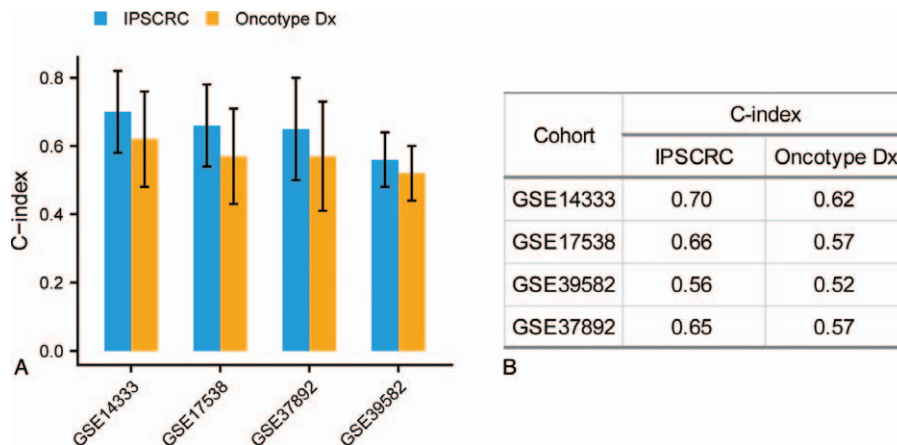

Figure 7. C-index for IPSCRC. C-index comparison between IPSCRC and Oncotype DX colon cancer (A). C-index values in different cohorts (B). IPSCRC = immune-based prognostic signature for CRC.

patients' survival. Our study is the first attempt to integrate tumor heterogeneity and the immune system to develop the prognostic signature for CRC.

Author contributions

Conceptualization: Peng Shu, Bin Zhang.
Data curation: Bin Zhang, Liangbin Wang, Wenliang Jiang.
Formal analysis: Peng Shu, Bin Zhang, Liangbin Wang, Bin Shao.
Funding acquisition: Bin Zhang.
Investigation: Peng Shu, Bin Zhang.
Methodology: Bin Zhang, Liangbin Wang, Zhixian Liu, Wenliang Jiang.
Resources: Zhixian Liu, Bin Shao.
Software: Bin Shao.
Supervision: Peng Shu, Bin Zhang.
Validation: Bin Zhang.
Writing – original draft: Peng Shu, Bin Zhang.
Writing – review & editing: Peng Shu, Bin Zhang, Liangbin Wang, Zhixian Liu, Bin Shao, Wenliang Jiang.

References

[1] Arnold M, Pandeya N, Byrnes G, et al. Global burden of cancer attributable to high body-mass index in 2012: a population-based study. Lancet Oncol 2015;16:36–46.

[2] Ferlay J, Shin H-R, Bray F, et al. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010;127:2893–917.

[3] Arnold M, Sierra MS, Laversanne M, et al. Global patterns and trends in colorectal cancer incidence and mortality. Gut 2017;66:683–91.

[4] Frampton M, Houlston RS. Modeling the prevention of colorectal cancer from the combined impact of host and behavioral risk factors. Genet Med 2017;19:314–21.

[5] Inadomi JM. Screening for colorectal neoplasia. N Engl J Med 2017;376:149–56.

[6] Chang W, Gao X, Han Y, et al. Gene expression profiling-derived immunohistochemistry signature with high prognostic value in colorectal carcinoma. Gut 2014;63:1457–67.

[7] Barrier A, Boelle P-Y, Roser F, et al. Stage II colon cancer prognosis prediction by tumor gene expression profiling. J Clin Oncol 2006;24:4685–91.

[8] Wang Y, Jatkoe T, Zhang Y, et al. Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. J Clin Oncol 2004;22:1564–71.

[9] Yothers G, O'Connell MJ, Lee M, et al. Validation of the 12-gene colon cancer recurrence score in NSABP C-07 as a predictor of recurrence in patients with stage II and III colon cancer treated with fluorouracil and leucovorin (FU/LV) and FU/LV plus oxaliplatin. J Clin Oncol 2013;31:4512–9.

[10] Salazar R, Roepman P, Capella G, et al. Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer. J Clin Oncol 2011;29:17–24.

[11] Guinney J, Dienstmann R, Wang X, et al. The consensus molecular subtypes of colorectal cancer. Nat Med 2015;21:1350–6.

[12] Gentles AJ, Newman AM, Liu CL, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med 2015;21:938–45.

[13] Angell H, Galon J. From the immune contexture to the Immunoscore: the role of prognostic and predictive immune markers in cancer. Curr Opin Immunol 2013;25:261–7.

[14] Taieb J, Moehler M, Boku N, et al. Evolution of checkpoint inhibitors for the treatment of metastatic gastric cancers: current status and future perspectives. Cancer Treat Rev 2018;66:104–13.

[15] Ralph C, Elkord E, Burt DJ, et al. Modulation of lymphocyte regulation for cancer therapy: a phase II trial of tremelimumab in advanced gastric and esophageal adenocarcinoma. Clin Cancer Res 2010;16:1662–72.

[16] Kong JC, Guerra GR, Pham T, et al. Prognostic impact of tumor-infiltrating lymphocytes in primary and metastatic colorectal cancer. Dis Colon Rectum 2019;62:498–508.

[17] Pagès F, Mlecnik B, Marliot F, et al. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. Lancet 2018;391:2128–39.

[18] Peng D, Wang L, Li H, et al. An immune infiltration signature to predict the overall survival of patients with colon cancer. IUBMB Life 2019;71:1760–70.

[19] Jorissen RN, Gibbs P, Christie M, et al. Metastasis-associated gene expression changes predict poor outcomes in patients with Dukes stage B and C colorectal cancer. Clin Cancer Res 2009;15:7642–51.

[20] Smith JJ, Deane NG, Wu F, et al. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. Gastroenterology 2010;138:958–68.

[21] Marisa L, de Reyniès A, Duval A, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. PLoS Med 2013;10:e1001453.

[22] de Sousa E, Melo F, Colak S, et al. Methylation of cancer-stem-cell-associated Wnt target genes predicts poor prognosis in colorectal cancer patients. Cell Stem Cell 2011;9:476–85.

[23] Laibe S, Lagarde A, Ferrari A, et al. A seven-gene signature aggregates a subgroup of stage II colon cancers with stage III. OMICS 2012;16:560–5.

[24] Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics 2007;23:1846–7.

[25] Miller JA, Cai C, Langfelder P, et al. Strategies for aggregating gene expression data: the collapse Rows R function. BMC Bioinformatics 2011;12:322.

[26] Bhattacharya S, Andorf S, Gomes L, et al. ImmPort: disseminating data to the public for the future of immunology. Immunol Res 2014;58:234–9.

[27] Fletcher MN, Castro MA, Wang X, et al. Master regulators of FGFR2 signalling and breast cancer risk. Nat Commun 2013;4:2464.

[28] Zhao M, Kong L, Liu Y, et al. dbEMT: an epithelial-mesenchymal transition associated gene resource. Sci Rep 2015;5:11459.

[29] Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015;12:453–7.

[30] Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 2005;102:15545–50.

[31] O'Connell MJ, Lavery I, Yothers G, et al. Relationship between tumor gene expression and recurrence in four independent studies of patients with stage II/III colon cancer treated with surgery alone or surgery plus adjuvant fluorouracil plus leucovorin. J Clin Oncol 2010;28:3937–44.

[32] Fessler E, Jansen M, De Sousa E, et al. A multidimensional network approach reveals microRNAs as determinants of the mesenchymal colorectal cancer subtype. Oncogene 2016;35:6026–37.

[33] Trinh A, Trumpi K, De Sousa E, et al. Practical and Robust identification of molecular subtypes in colorectal cancer by immunohistochemistry. Clin Cancer Res 2017;23:387–98.

[34] Liotta LA, Kohn EC. The microenvironment of the tumour-host interface. Nature 2001;411:375–9.

[35] Nakazawa T, Goi T, Hirono Y, et al. Prokineticin 1 protein expression is a useful new prognostic factor for human sporadic colorectal cancer. Ann Surg Oncol 2015;22:1496–503.

[36] Zhou Z-Q, Cao W-H, Xie J-J, et al. Expression and prognostic significance of THBS1, Cyr61 and CTGF in esophageal squamous cell carcinoma. BMC Cancer 2009;9:291.

[37] Kersten C, Louhimo J, Ålgars A, et al. Increased C-reactive protein implies a poorer stage-specific prognosis in colon cancer. Acta Oncol 2013;52:1691–8.

[38] Zhao F-T, Jia Z-S, Yang Q, et al. S100A14 Promotes the Growth and Metastasis of Hepatocellular Carcinoma. Asian Pac J Cancer Prev 2013;14:3831–6.

[39] Zhang X, Wang Y, Cao Y, et al. Increased CCL19 expression is associated with progression in cervical cancer. Oncotarget 2017;8:73817–25.

[40] Nielsen HJ, Hansen U, Christensen IJ, et al. Independent prognostic value of eosinophil and mast cell infiltration in colorectal cancer tissue. J Pathol 1999;189:487–95.

[41] Lan C, Huang X, Lin S, et al. Expression of M2-polarized macrophages is associated with poor prognosis for advanced epithelial ovarian cancer. Technol Cancer Res Treat 2013;12:259–67.