Journal of Zhejiang University-SCIENCE B (Biomedicine & Biotechnology) *2025 26(4):371-392*
www.jzus.zju.edu.cn; www.springer.com/journal/11585
E-mail: jzus_b@zju.edu.cn

Research Article

https://doi.org/10.1631/jzus.B2300679

Single-cell and spatial transcriptomic analysis reveals that an immune cell-related signature could predict clinical outcomes for microsatellite-stable colorectal cancer patients receiving immunotherapy

Shijin YUAN[1,2], Yan XIA[1,3], Guangwei DAI[1,3], Shun RAO[1,3], Rongrong HU[1,3,4], Yuzhen GAO[1,3], Qing QIU[1,3], Chenghao WU[1,3], Sai QIAO[1,3], Yinghua XU[2], Xinyou XIE[1,3], Haizhou LOU[2], Xian WANG[2]✉, Jun ZHANG[1,3]✉

[1]*Department of Clinical Laboratory, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou 310016, China*
[2]*Department of Medical Oncology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou 310016, China*
[3]*Key Laboratory of Precision Medicine in Diagnosis and Monitoring Research of Zhejiang Province, Hangzhou 310016, China*
[4]*Yongkang Hospital of Traditional Chinese Medicine Medical Community Xicheng Branch, Jinhua 321300, China*

Abstract: Recent data suggest that vascular endothelial growth factor receptor inhibitor (VEGFRi) can enhance the anti-tumor activity of the anti-programmed cell death-1 (anti-PD-1) antibody in colorectal cancer (CRC) with microsatellite stability (MSS). However, the comparison between this combination and standard third-line VEGFRi treatment is not performed, and reliable biomarkers are still lacking. We retrospectively enrolled MSS CRC patients receiving anti-PD-1 antibody plus VEGFRi (combination group, n=54) or VEGFRi alone (VEGFRi group, n=32), and their efficacy and safety were evaluated. We additionally examined the immune characteristics of the MSS CRC tumor microenvironment (TME) through single-cell and spatial transcriptomic data, and an MSS CRC immune cell-related signature (MCICRS) that can be used to predict the clinical outcomes of MSS CRC patients receiving immunotherapy was developed and validated in our in-house cohort. Compared with VEGFRi alone, the combination of anti-PD-1 antibody and VEGFRi exhibited a prolonged survival benefit (median progression-free survival: 4.4 vs. 2.0 months, P=0.0024; median overall survival: 10.2 vs. 5.2 months, P=0.0038) and a similar adverse event incidence. Through single-cell and spatial transcriptomic analysis, we determined ten MSS CRC-enriched immune cell types and their spatial distribution, including naive CD4$^+$ T, regulatory CD4$^+$ T, CD4$^+$ Th17, exhausted CD8$^+$ T, cytotoxic CD8$^+$ T, proliferated CD8$^+$ T, natural killer (NK) cells, plasma, and classical and intermediate monocytes. Based on a systemic meta-analysis and ten machine learning algorithms, we obtained MCICRS, an independent risk factor for the prognosis of MSS CRC patients. Further analyses demonstrated that the low-MCICRS group presented a higher immune cell infiltration and immune-related pathway activation, and hence a significant relation with the superior efficacy of pan-cancer immunotherapy. More importantly, the predictive value of MCICRS in MSS CRC patients receiving immunotherapy was also validated with an in-house cohort. Anti-PD-1 antibody combined with VEGFRi presented an improved clinical benefit in MSS CRC with manageable toxicity. MCICRS could serve as a robust and promising tool to predict clinical outcomes for individual MSS CRC patients receiving immunotherapy.

Key words: Colorectal cancer (CRC); Microsatellite stability (MSS); Immunotherapy; Single-cell RNA sequencing (scRNA-seq); Spatial transcriptomics

1 Introduction

Colorectal cancer (CRC) ranks among all cancer types as the third highest in terms of incidence and second highest in terms of mortality worldwide (Sung et al., 2021). In China, CRC is one of the five most commonly diagnosed cancers and is the leading cause of cancer-related death in both genders, with approximately 408 000 new cases and 195 600 deaths in 2016

✉ Jun ZHANG, jameszhang2000@zju.edu.cn
Xian WANG, wangx118@zju.edu.cn
Jun ZHANG, https://orcid.org/0000-0002-2613-3165
Xian WANG, https://orcid.org/0000-0003-0041-7589

Received Sept. 21, 2023; Revision accepted Jan. 25, 2024;
Crosschecked Apr. 14, 2025

(Zheng et al., 2023). For locally advanced or metastatic CRC, chemotherapy based on 5-fluorouracil is the standard first- and second-line treatment, including FOLFOX (5-fluorouracil, leucovorin plus oxaliplatin) and FOLFIRI (5-fluorouracil, leucovorin plus irinotecan) alone or combined with the vascular endothelial growth factor (VEGF) blocker bevacizumab or the epidermal growth factor receptor (EGFR) blocker cetuximab (if *RAS* wild-type) (Modest et al., 2019). Two small-molecule inhibitors that target the VEGF receptor (VEGFR), namely, regorafenib and fruquintinib, have been approved for treating patients who have progressed after chemotherapy in China. Although regorafenib or fruquintinib presented an improved survival compared with placebo, fewer than 5% of patients achieved an objective response, indicating that the clinical benefit of the VEGFR inhibitor (VEGFRi) is still limited, and exploring other effective treatment strategies for chemotherapy-refractory CRC patients remains an unmet need (Grothey et al., 2013; Li et al., 2018).

Recently, immunotherapies represented by immune checkpoint inhibitors (ICIs), such as the anti-cytotoxic T-lymphocyte antigen 4 (anti-CTLA4) antibody ipilimumab and the anti-programmed cell death 1 (anti-PD-1) antibodies pembrolizumab and nivolumab, have exhibited durable anti-tumor activity in various cancers, revolutionizing the field of cancer treatment. Notably, the objective response rate (ORR) of single agent or doublet ICIs has reached 30%–60% in a handful of CRC patients who are characterized as microsatellite instability-high (MSI-H), a molecular phenotype of CRC that is due to mismatch repair-related gene mutation (Le et al., 2015, 2017, 2020; Overman et al., 2017, 2018; André et al., 2020). Based on these results, pembrolizumab or nivolumab monotherapy and nivolumab combined with ipilimumab have been approved for the treatment of MSI-H CRC. However, immunotherapy is particularly ineffective in patients with microsatellite stability (MSS), a population that comprises approximately 85% of advanced CRC. The treatment failure is primarily due to the immunosuppressive tumor microenvironment (TME) of MSS CRC. Hence, the combination of ICIs and other therapies to reverse the immunosuppressive environment may be the key to treating MSS CRC.

The mechanisms of VEGF-mediated immune suppression have been widely clarified. Blocking the VEGF-related pathway could normalize vascular construction, inhibit the production of immunosuppressive cells, such as regulatory T cells (Tregs) and myeloid-derived suppressor cells (MDSCs), and increase effector T cell infiltration in the TME (Rahma and Hodi, 2019). Therefore, it is reasonable to consider that VEGFRi has a synergistic effect with immunotherapy through restoring immune activity in MSS tumors. Three current single-arm clinical trials have shown that ICIs combined with VEGFRi have a promising anti-tumor activity in MSS CRC (Fukuoka et al., 2020; Cousin et al., 2021; Kim et al., 2022), whereas a comparison between this combination and standard third-line treatment is still lacking. Meanwhile, identifying predictive biomarkers to select suitable patients to receive immunotherapy is critically important. Single-cell RNA sequencing (scRNA-seq) and spatial transcriptomic data provide the comprehensive characteristics of TME cell-type composition, distribution, and heterogeneity, which help to advance our understanding of key features of tumorigenesis, metastasis, and drug resistance.

In this work, we evaluated the efficacy and safety of the combination of anti-PD-1 antibody and VEGFRi compared with VEGFRi alone in MSS CRC. Through scRNA-seq and spatial transcriptomic data, we determined the gene signature of MSS CRC-enriched immune cells and assessed its predictive value for the prognosis, recurrence, and benefit of immunotherapy in MSS CRC patients. Our work may help optimize precision treatment and further improve the clinical outcomes of CRC patients.

2 Materials and methods

2.1 Study design, patient enrollment, data collection, and treatment procedure

This study retrospectively evaluated the efficacy and safety of anti-PD-1 antibody and VEGFRi combination therapy compared with VEGFRi monotherapy in MSS CRC.

Eligible patients were pathologically confirmed to have unresectable locally advanced or metastatic MSS CRC and the progression of previous chemotherapy. The exclusion criteria were as follows: (1) autoimmune disease history or any underlying immune dysfunction disease; (2) an Eastern Cooperative Oncology

Group (ECOG) performance status of 3 or higher; (3) inadequate bone marrow hematopoietic, hepatic, and renal functions. Patients receiving anti-PD-1 antibody and VEGFRi (regorafenib or fruquintinib) were enrolled as the combination group, while patients receiving VEGFRi monotherapy were enrolled as the VEGFRi group. All enrolled participants signed an informed consent form. The baseline clinical characteristics of each enrolled patient were collected from electronic medical records. Peripheral blood lymphocyte and monocyte counts of each patient were obtained within one week before treatment initiation through the Hematology Analyzer System (Beckman Coulter LH780 ANl6104).

With regards to treatment procedure, in the VEGFRi group, patients received regorafenib or fruquintinib orally once daily for three weeks and were off treatment for one week in a four-week cycle. In the combination group, anti-PD-1 antibody was given intravenously starting on Day 1 of VEGFRi initiation.

2.2 Therapeutic efficacy, survival, and safety assessment

Therapeutic efficacy was evaluated by radiologists and investigators every six weeks after treatment initiation according to the Response Evaluation Criteria in Solid Tumors version 1.1, and was documented as complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD). ORR was defined as the percentage of patients with CR or PR, and disease control rate (DCR) was defined as the percentage of patients with CR, PR, or SD. Progression-free survival (PFS) was defined as the duration from the date of treatment initiation to the date of disease progression or death due to any cause, whichever occurred first. Overall survival (OS) was defined as the duration from the date of treatment initiation to the date of death due to any cause. All treatment-related adverse events (AEs) were evaluated and graded according to the Common Terminology Criteria for Adverse Events version 4.0.

2.3 Public scRNA-seq, spatial transcriptomic, and bulk RNA-seq data collection

The scRNA-seq data of CRC were collected from two independent datasets, GSE132465 and GSE146771 (Lee et al., 2020; Zhang et al., 2020). Specifically, GSE132465 contained 65 362 single cells from 33 samples (23 tumors and 10 adjacent normal tissues) of 23 CRC patients (4 with MSI-H and 19 with MSS), and GSE146771 contained 54 285 single cells from the tumors, adjacent normal tissues, and blood samples of 18 CRC patients (3 with MSI-H and 15 with MSS). All CRC patients were treatment-naive and all single cells underwent quality control for further analyses. The spatial transcriptomic data of CRC patients were downloaded from the website (http://www.cancerdiversity.asia/scCRLM) of a spatial transcriptomic study including eight samples from four patients (Wu et al., 2022). Four paired primary and liver metastatic (LM) tumor samples from two treatment-naive patients, namely, CRC-1P, LM-1P, CRC-2P, and LM-2P, were used for spatial analyses. In addition, the bulk RNA-seq data of 1293 MSS CRC patients from seven public datasets were obtained from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO). Among these, the TCGA-CRC cohort was composed of 418 patients from TCGA-colon adenocarcinoma (COAD) and TCGA-rectum adenocarcinoma (READ) datasets, and the GEO-meta cohort was composed of six datasets including GSE39582 (n=459), GSE39084 (n=54), GSE41258 (n=137), GSE92921 (n=53), GSE143985 (n=85), and GSE30378 (n=87). Furthermore, 23 pan-cancer immunotherapy cohorts with therapeutic efficacy, prognosis, and bulk RNA-seq data were also collected, including 1 MSS CRC dataset, 12 melanoma datasets, 2 renal carcinoma datasets, 2 urothelium carcinoma datasets, 2 lung cancer datasets, 1 breast cancer dataset, 1 gastric cancer dataset, 1 esophagus cancer dataset, and 1 hepatocellular carcinoma dataset. The detailed information about each dataset is presented in Table S1.

2.4 scRNA-seq data processing and analysis

The scRNA-seq data analysis was performed using the “Seurat” R package. First, the batch effects among different samples were corrected using the canonical correlation analysis method (Butler et al., 2018). Then, the top 2000 highly variable genes were identified by the “FindVariableFeatures” function for further principal component analysis (PCA) and t-distributed stochastic neighbor embedding (tSNE) dimensionality reduction. Furthermore, we used the “FindClusters” function to obtain different cell clusters, and the “FindAllMarkers” function to determine the differentially expressed genes (DEGs) for each cluster.

Lastly, we annotated different cell types manually according to well-known marker genes and previous research.

2.5 Tissue distribution preference analysis

To characterize the tissue distribution of different cell types, we calculated odds ratios (ORs) to indicate preferences through Fisher's exact test. An OR value of >1.5 indicated that cell type was preferentially distributed in tissue, whereas an OR value of <0.5 indicated that cell type preferred not to be distributed in tissue (Zheng et al., 2021).

2.6 Cell–cell communication analysis

Based on a ligand−receptor interaction database, the cell–cell communication network was analyzed using the "CellChat" R package to explore the communication probability of different cell subtypes, major pathway signaling, and ligand–receptor pairs (Jin et al., 2021). The "netVisual_circle" function was used to visualize the interaction numbers and strength among different cell subtypes.

2.7 Pseudotime trajectory analysis

The "monocle" R package was performed to illustrate the cell state transition of different cell types including CD4$^+$ T cells, CD8$^+$ T cells, and monocytes; a reversed graph-embedding technique was applied to reconstruct single-cell trajectories (Qiu et al., 2017). Briefly, highly variable genes were identified for semi-supervised trajectory reconstruction. Dimensional reduction and cell ordering were performed using the DDRTree method and the "orderCells" function. The "plot_cell_trajectory" function was used to visualize the dynamic pseudotime trajectories of different cell types.

2.8 Spatial transcriptomic data processing and cell type identification

The spatial transcriptomic data of each sample were processed using the "Seurat" R package and analyzed separately. In detail, the data were normalized using the "SCTransform" method and then submitted to perform "RunPCA," "FindNeighbors," and "FindClusters" to obtain clusters of spatial spots with similar transcriptomic characteristics. Considering that each spot contained approximately ten cells, we could not identify the certain cell type for each spot. Hence, using cell type annotation information from scRNA-seq data as a reference, combined with pathological features of hematoxylin and eosin (H&E) staining sections, we predicted the main cell types of spatial spots using the "FindTransferAnchors" and "TransferData" functions of "Seurat." As for immune cell subtype annotation, based on DEGs and the well-known marker genes of each cell subtype, we used a signature-based strategy to score each cell subtype enrichment via the "AddModuleScore" function, which was visualized by the "SpatialFeaturePlot" function.

2.9 Bulk RNA-seq data processing

The RNA-seq raw read count data of the TCGA-CRC cohort were downloaded via the "TCGAbiolinks" R package, further converted to the values of transcripts per kilobase million (TPM), and $\log_2$-transformed (Colaprico et al., 2016). For the microarray data of six GSE CRC cohorts, we used normalized matrix files downloaded from the GEO database.

In addition, among 23 immunotherapy cohorts, the RNA-seq data of the IMvigor210 cohort were available from the "IMvigor210" R package (Mariathasan et al., 2018). The expression data of other cohorts were downloaded from the GEO database or supplementary files from corresponding research (Table S1).

2.10 Weighted correlation network analysis (WGCNA)

Co-expression gene networks of the TCGA-CRC cohort were generated via the "WGCNA" R package (Langfelder and Horvath, 2008). An appropriate soft threshold β was calculated for the scale-free network. The weighted adjacency matrix was converted into a topological overlap matrix (TOM), and the corresponding dissimilarity was generated (1−TOM). The dynamic tree-cutting approach was used to identify different modules. The module that displayed the highest correlation with immune cell enrichment was chosen for further analysis, and the genes with both high gene significance (GS) and module membership (MM) were defined as immune cell-related hub genes.

2.11 MSS CRC immune cell-related signatures generated from machine learning-based integrative approaches

To develop a stable and accurate MSS CRC immune cell-related signature (MCICRS), we first used a multivariable Cox proportional hazards regression analysis to obtain OS-related genes in TCGA-CRC, GSE39582, and GSE41258 cohorts adjusted by age, gender, and tumor node metastasis (TNM) stage. Through the "meta" R package, the hazard ratio (HR) of each OS-related gene was pooled in the random-effects model of a systematic meta-analysis, and the stable prognostic genes were determined. Then, we integrated ten machine learning algorithms, including random survival forest (RSF, "randomForestSRC" R package), Lasso ("glmnet" R package), elastic network (Enet, "glmnet" R package), Ridge ("glmnet" R package), stepwise Cox ("survival" R package), CoxBoost ("CoxBoost" R package), partial least squares regression for Cox (plsRcox, "plsRcox" R package), supervised principal components (SuperPC, "superpc" R package), generalized boosted regression modeling (GBM, "superpc" R package), and survival support vector machine (survival-SVM, "survivalsvm" R package), and a total of 96 algorithm combinations to fit prediction models across the TCGA-CRC, GSE39582, GSE39084, and GSE41258 cohorts. The concordance index (*C*-index) was calculated for each model, and the model with the highest average *C*-index was considered optimal.

2.12 Immune infiltration analysis

A total of seven algorithms, including single-sample gene set enrichment analysis (ssGSEA, "GSVA" R package), cell type enrichment analysis (xCell, "xCell" R package), estimation of stromal and immune cells in malignant tumors using expression data (ESTIMATE, "estimate" R package), quantitative tumor immune cell spatial characterization (quanTIseq, "immunedeconv" R package), microenvironment cell populations-counter (MCP-counter, "MCPcounter" R package), estimation of proportions of immune and cancer cells (EPIC, "EPIC" R package), and the tumor immune estimation resource (TIMER) databases, were employed to quantify the relative infiltration abundances of different TME immune cell types in the TCGA-CRC cohort.

Furthermore, the gene set variation analysis (GSVA, "GSVA" R package) and gene set enrichment analysis (GSEA, "clusterProfiler" R package) were performed to investigate the activity of pathway signaling based on hallmark gene sets (h.all.v7.5.1.symbols, Molecular Signatures Database (MSigDB), http://www.gsea-msigdb.org/gsea/msigdb).

2.13 Proteomic analysis through proteomaps

We used proteomaps to analyze and visualize the composition of protein abundances and their functions (https://www.proteomaps.net) (Liebermeister et al., 2014). The proteomaps could be divided into six different functional parts, including genetic information processing, metabolism, organismal systems, environmental information processing, human disease, and cellular processes. Each protein was shown as a polygon-shaped tile with an area representing protein abundance, and function-related proteins appeared in adjacent regions.

2.14 Statistical analysis

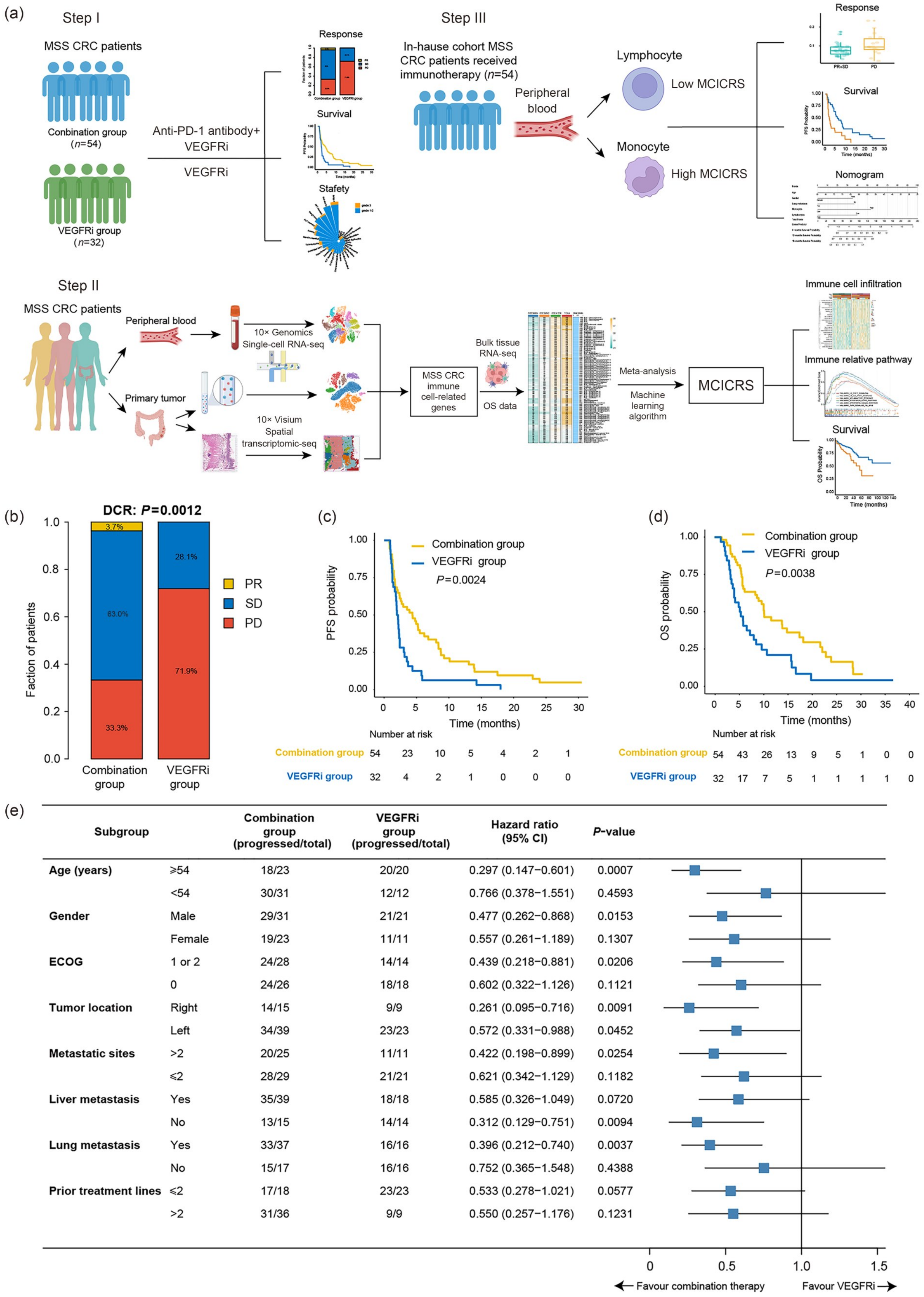
Categorical variables were reported as frequencies and percentages, and their significance was compared using the Chi-square test. For continuous variables, Student's *t*-test or Wilcoxon test was used for comparisons between two groups, and one-way analysis of variance (ANOVA) or the Kruskal-Wallis test for comparisons among three groups. In addition, Pearson's correlation test was used for assessing correlations between two continuous variables.

The best cut-off value was determined by the "survminer" R package. The Kaplan-Meier method and log-rank test were used for survival analysis via the "survival" R package. Multivariable Cox regression analysis was used to identify significant prognostic variables and calculate HR and 95% confidence interval (CI), which was visualized using the "forestplot" R package. The nomogram was developed using the "rms" R package, and the time-dependent area under the receiver operating characteristic (ROC) curve (AUC) was evaluated using the "timeROC" R package. All statistical analyses were performed using the R software version 4.0. All statistical tests were two-tailed and a *P*-value less than 0.05 was considered statistically significant.

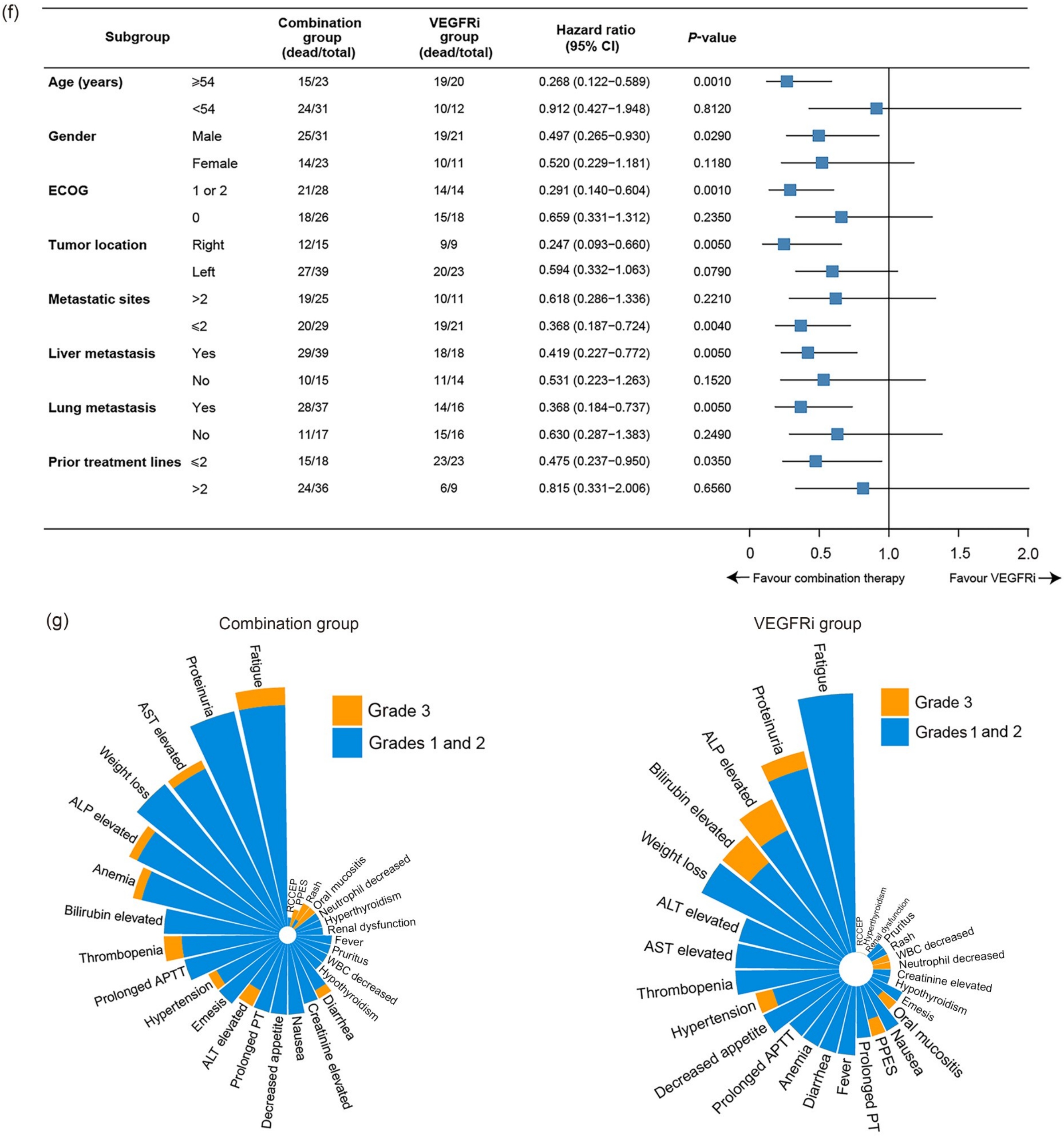
3 Results

3.1 Immune and angiogenic landscape in the MSS CRC tumor microenvironment

The overall design of this study is displayed in Fig. 1a. To better elucidate the TME cellular composition of a CRC tumor, we analyzed a total of 65 362



Subgroup		Combination group (progressed/total)	VEGFRi group (progressed/total)	Hazard ratio (95% CI)	*P*-value
Age (years)	≥54	18/23	20/20	0.297 (0.147−0.601)	0.0007
	<54	30/31	12/12	0.766 (0.378−1.551)	0.4593
Gender	Male	29/31	21/21	0.477 (0.262−0.868)	0.0153
	Female	19/23	11/11	0.557 (0.261−1.189)	0.1307
ECOG	1 or 2	24/28	14/14	0.439 (0.218−0.881)	0.0206
	0	24/26	18/18	0.602 (0.322−1.126)	0.1121
Tumor location	Right	14/15	9/9	0.261 (0.095−0.716)	0.0091
	Left	34/39	23/23	0.572 (0.331−0.988)	0.0452
Metastatic sites	>2	20/25	11/11	0.422 (0.198−0.899)	0.0254
	≤2	28/29	21/21	0.621 (0.342−1.129)	0.1182
Liver metastasis	Yes	35/39	18/18	0.585 (0.326−1.049)	0.0720
	No	13/15	14/14	0.312 (0.129−0.751)	0.0094
Lung metastasis	Yes	33/37	16/16	0.396 (0.212−0.740)	0.0037
	No	15/17	16/16	0.752 (0.365−1.548)	0.4388
Prior treatment lines	≤2	17/18	23/23	0.533 (0.278−1.021)	0.0577
	>2	31/36	9/9	0.550 (0.257−1.176)	0.1231



Subgroup		Combination group (dead/total)	VEGFRi group (dead/total)	Hazard ratio (95% CI)	*P*-value
Age (years)	≥54	15/23	19/20	0.268 (0.122–0.589)	0.0010
	<54	24/31	10/12	0.912 (0.427–1.948)	0.8120
Gender	Male	25/31	19/21	0.497 (0.265–0.930)	0.0290
	Female	14/23	10/11	0.520 (0.229–1.181)	0.1180
ECOG	1 or 2	21/28	14/14	0.291 (0.140–0.604)	0.0010
	0	18/26	15/18	0.659 (0.331–1.312)	0.2350
Tumor location	Right	12/15	9/9	0.247 (0.093–0.660)	0.0050
	Left	27/39	20/23	0.594 (0.332–1.063)	0.0790
Metastatic sites	>2	19/25	10/11	0.618 (0.286–1.336)	0.2210
	≤2	20/29	19/21	0.368 (0.187–0.724)	0.0040
Liver metastasis	Yes	29/39	18/18	0.419 (0.227–0.772)	0.0050
	No	10/15	11/14	0.531 (0.223–1.263)	0.1520
Lung metastasis	Yes	28/37	14/16	0.368 (0.184–0.737)	0.0050
	No	11/17	15/16	0.630 (0.287–1.383)	0.2490
Prior treatment lines	≤2	15/18	23/23	0.475 (0.237–0.950)	0.0350
	>2	24/36	6/9	0.815 (0.331–2.006)	0.6560

Fig. 1 Effect of immunotherapy combined with anti-angiogenic therapy on the clinical outcomes of MSS CRC patients. (a) Study workflow. (b) The treatment response comparison of the combination and VEGFRi groups (disease control rate (DCR): 66.7% vs. 28.1%, *P*=0.0012, Chi-square test). (c, d) Kaplan-Meier analyses of progression-free survival (PFS) (c) and overall survival (OS) (d) in the combination and VEGFRi groups (log-rank test). (e, f) Subgroup analyses of PFS (e) and OS (f) in the combination and VEGFRi groups. (g) Adverse events of any grade in the combination and VEGFRi groups. PR: partial response; SD: stable disease; PD: progressive disease; ECOG: Eastern Cooperative Oncology Group; ALP: alkaline phosphatase; ALT: alanine aminotransferase; AST: aspartate aminotransferase; APTT: activated partial thromboplastin time; PT: prothrombin time; PPES: palmar-plantar erythrodysesthesia syndrome; WBC: white blood cell count; RCCEP: reactive cutaneous capillary endothelial proliferation.

cells from the scRNA-seq data, and eight major cell types were annotated (Figs. S1a, S2a, and S2b). Therein, the MSS tumor had the highest proportion of epithelial cells and the lowest immune cell infiltration compared with the MSI tumor and adjacent normal tissue (Fig. S2c). To illustrate the effects of immune suppression and angiogenesis in MSS CRC tumors, we first investigated the immune checkpoint and VEGF pathway-related gene expression among different cell types (Fig. S1b). Compared with adjacent normal tissue, we observed high expression of major immune checkpoint genes in T cells from the MSS tumor, such as PD-1, CD274 (programmed death-ligand 1 (PD-L1)), CTLA4, lymphocyte-activation gene 3 (LAG3), T-cell immunoglobulin and immunoreceptor tyrosine-based inhibitory motif domain (TIGIT), and hepatitis A virus cellular receptor 2 (HAVCR2) (T-cell immunoglobulin and mucin-domain containing-3 (TIM-3)), indicating an exhausted state of T cells in the TME. In addition, VEGFA and VEGFB were highly expressed in malignant epithelial and myeloid cells and fibroblasts in tumor tissue, demonstrating that these cell types were the main sources of angiogenic factor release. In addition, we found high co-expression of CD274, HAVCR2, VEGFA, and VEGFR1 in tumor myeloid cells, suggesting their critical roles in maintaining a suppressed immune and pro-angiogenic TME.

Owing to the complexity of intratumor architecture, we next explored the spatially expressed location of immune checkpoint and VEGF pathway-related genes in spatial transcriptomic data (Fig. S1c). We found a highly spatial co-expression of CD68, endoglin (ENG), and platelet endothelial cell adhesion molecule 1 (PECAM1), indicating a wide distribution of myeloid cells in the stromal region. Meanwhile, VEGFA and VEGFB were also co-expressed in the same area, which was consistent with the scRNA-seq data. Interestingly, there was a subset of exhausted CD8$^+$ T cells (high expression of CD3D, CD8A, PD-1, CTLA4, and TIGIT) with co-localized VEGFB$^+$ myeloid cells wrapped around endothelial cells, indicating a physical interaction between VEGF-driven angiogenesis and immune suppression. Additionally, a significant positive correlation was observed between the immune checkpoint and VEGF-related angiogenic signatures among all spots of spatial transcriptomic data (Fig. S1d). Furthermore, in MSS CRC tumors, the extensive communications among malignant epithelial, myeloid, and endothelial cells and fibroblasts through the VEGF signaling pathway were presented by the cell-chat analysis (Figs. S1e and S1f). Collectively, these findings verified the synergistic effect between immunosuppression and angiogenesis from single-cell and spatial dimensions, and hinted that the combination of immunotherapy and anti-angiogenic therapy might enhance anti-tumor activity in MSS CRC.

3.2 Effect of immunotherapy combined with anti-angiogenic therapy on the clinical outcomes of MSS CRC patients

To evaluate the efficacy and safety of immunotherapy combined with anti-angiogenic therapy in MSS CRC, we retrospectively enrolled 54 patients receiving anti-PD-1 antibody plus VEGFRi as a combination group. For comparison, 32 patients receiving standard VEGFRi monotherapy were enrolled as a VEGFRi group. The baseline clinical characteristics of the patients in the two groups were basically balanced (Table S2). At the cutoff date, the median follow-up time was 5.8 months (interquartile range (IQR) 3.1–10.7 months) in the combination group and 3.8 months (IQR 2.0–6.4 months) in the VEGFRi group (Table S3). PR was observed in two patients in the combination group (ORR: 3.7%) but in no patients in the VEGFRi group (ORR: 0%). The DCR of the combination group was significantly higher than that of the VEGFRi group (66.7% vs. 28.1%, $P=0.0012$; Fig. 1b). For the survival comparison between the two groups, the combination group showed a longer median PFS (4.4 vs. 2.0 months, $P=0.0024$; Fig. 1c) and OS (10.2 vs. 5.2 months, $P=0.0038$; Fig. 1d). Considering the differentiation of targets between fruquintinib and regorafenib, we compared the outcomes of anti-PD-1 antibody plus fruquintinib, anti-PD-1 antibody plus regorafenib, and the VEGFRi group separately, and we found that the treatment response and survival between the anti-PD-1 antibody plus fruquintinib and anti-PD-1 antibody plus regorafenib groups were comparable, while both were superior to those of the VEGFRi group, verifying the reliability of our results (Fig. S3). In addition, the subgroup analysis revealed an improved survival benefit for anti-PD-1 antibody plus VEGFRi combination, therapy when patients were stratified into major subgroups based on different clinical characteristics, such as age$\geqslant$54 years, being male, an ECOG of 1, the right side of the primary tumor, $\leqslant$2 metastatic sites, and lung metastasis (Figs. 1e and 1f).

Treatment-related AEs of any grade were reported equally between the two groups, including 52 patients (96.3%) in the combination group and 30 patients (93.8%) in the VEGFRi group ($P>0.99$; Table S4). Therein, Grade 3 AEs were reported in 16 patients (29.6%) in the combination group and 8 patients (25%) in the VEGFRi group ($P=0.91$). The most common

AEs in two groups were fatigue (32 (50.0%) in the combination group and 15 (46.9%) in the VEGFRi group) and proteinuria (25 (46.3%) in the combination group and 12 (37.5%) in the VEGFRi group) (Fig. 1g). In addition, a total of six patients in the combination group (9.4%) experienced immune-related AEs including pneumonia, myocarditis, and thyroiditis, and all patients resumed treatment after AEs solved. Overall, the above results demonstrated that immunotherapy combined with anti-angiogenic therapy exhibited robust anti-tumor activity and improved the clinical outcomes of MSS CRC patients with manageable toxicity. Hence, we further investigated the particular TME immune characteristics of MSS CRC patients and identified potential subsets suitable for immunotherapy.

3.3 Single-cell transcriptomic atlas in the MSS CRC tumor immune microenvironment

Since immune cells are major effectors in immunotherapy, we first re-clustered lymphocytes (T/B/plasma cells) in scRNA-seq data and identified 14 cell subtypes (Figs. 2a and S2d, Table S5). Based on cell proportion and tissue distribution preference analyses, naive CD4$^+$ T (CD4T_naive), regulatory CD4$^+$ T (CD4_Treg), CD4$^+$ Th17 (CD4_Th17), exhausted CD8$^+$ T (CD8T_exhausted), cytotoxic CD8$^+$ T (CD8T_cytotoxic), proliferated CD8$^+$ T (CD8T_plf), natural killer (NK) cells, and immunoglobulin G (IgG) plasma cells (Plasma_IgG) showed a strong distribution preference in MSS tumors (Figs. 2b and 2c, Table S6). In addition, cell-chat analysis revealed that naive CD4$^+$ T, exhausted CD8$^+$ T, and epithelial cells had more communications with other cells in MSS tumors (Fig. 2d). We next explored the dynamic immune states and cell transitions in MSS CRC-infiltrated CD4$^+$ T and CD8$^+$ T cells by inferring the state trajectories. Pseudotime analysis showed that naive CD4$^+$ T cells were at the beginning of the trajectory path, whereas several CD4$^+$ Th17 cells and most regulatory CD4$^+$ T cells were at a terminal state (Fig. 2e). Similarly, we also found that the proliferated CD8$^+$ T cells primarily aggregated at the initiation of the pseudotime trajectory, and exhausted CD8$^+$ T and cytotoxic CD8$^+$ T cells were located in different directions (Fig. 2f). The trajectory information indicated the different functional divergence of T cells in the MSS tumor; accurately identifying the MSS subtype with activated immune and transcriptional states may help to expand immunotherapy candidates.

Similarly, we re-clustered myeloid cells and identified nine cell subtypes, of which classical monocytes and intermediate monocytes were considered MSS tumor-enriched cell subtypes (Figs. 2g–2i and S2e, Tables S7 and S8). Additionally, a wider crosstalk between monocytes and other cells was observed in MSS tumors through cell-chat analysis (Fig. 2j). Furthermore, the pseudotime trajectory displayed a distinct distribution of monocyte subclusters (Fig. 2k). Collectively, we deciphered the cellular composition, interaction, and transition of the tumor immune environment in MSS CRC at single-cell resolution, which helped us to deeply understand the immune characteristics and identify different subtypes of MSS tumors.

3.4 Spatial and peripheral blood single-cell transcriptomic landscape in MSS CRC

Given the complexity of TME structure, we subsequently analyzed CRC spatial transcriptomic data to explore the spatial distribution of different immune cells. We first used the unsupervised clustering method to classify the spatial spots into different clusters, and then identified distinct regions, including normal epithelial, hepatocyte, tumor, stromal, and immune cell regions, according to H&E staining sections and the integration of previous scRNA-seq annotation data (Figs. 3a, S1a, and S4a). Notably, owing to a spatial spot containing approximately ten cells, we observed several cell types colocalized at the same spots in some regions, such as malignant/immune cells, immune/stromal cells, and malignant/stromal cells, indicating physical interactions between these cell types. Furthermore, we quantified the spatial distribution of MSS tumor-enriched immune cell subtypes based on their DEGs and well-known marker genes from previous scRNA-seq results (Figs. 3a and S4a, Tables S5 and S7). In primary tumors, we found that T, NK, and plasma cells were distributed locally in the stromal region and surrounded by abundant fibroblasts or lamina propria, suggesting that these effective immune cells were difficult to infiltrate into the tumor parenchyma (Fig. 3a). By contrast, monocytes were more widely enriched in tumor regions, as shown by cell-chat results of extensive interaction between monocytes and tumor cells (Fig. 2j). These results demonstrated significant spatial intratumoral immune infiltration heterogeneity in MSS CRC.

Most TME-infiltrated immune cells are derived from peripheral blood; therefore, we also analyzed the

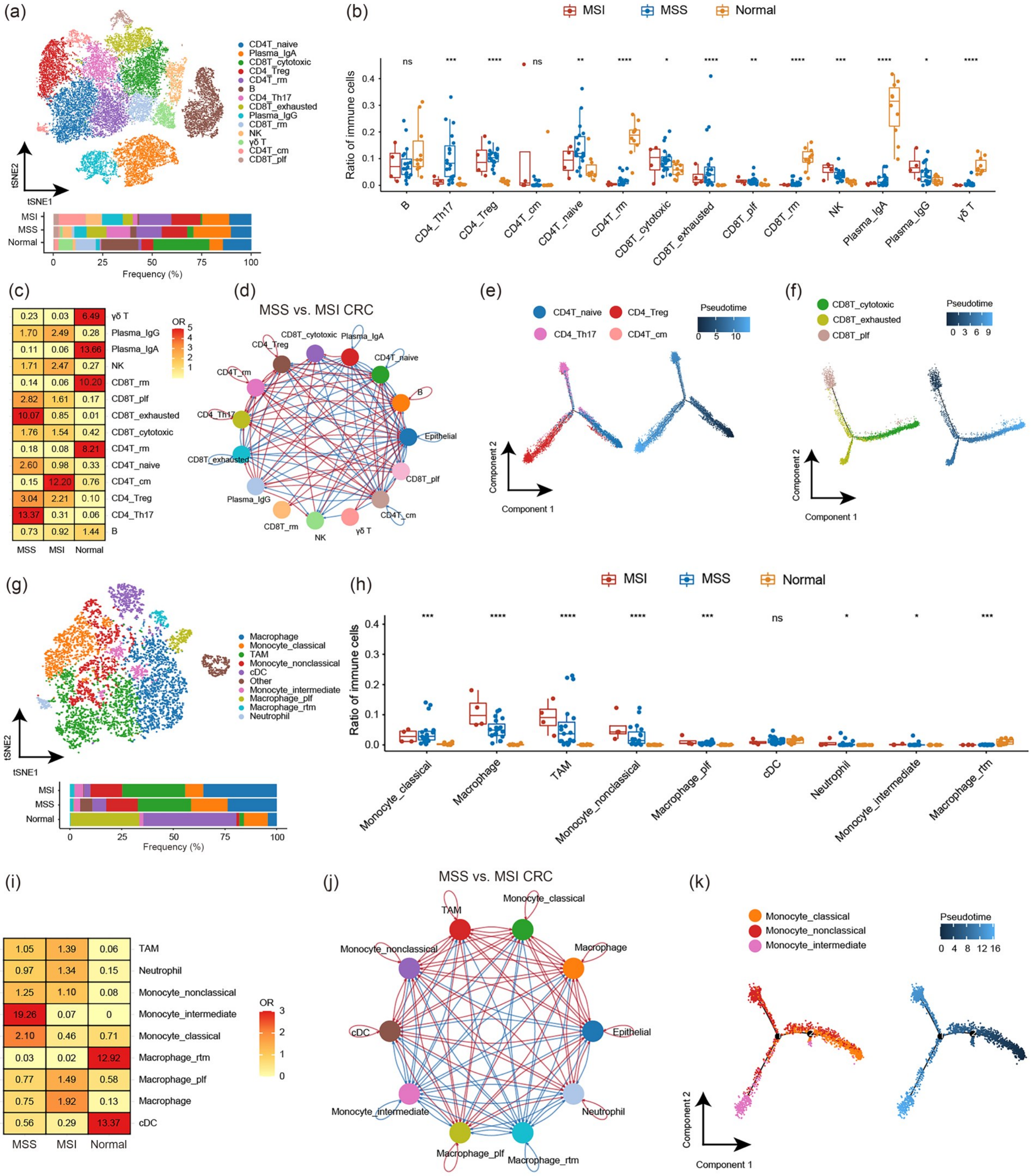


Fig. 2 Single-cell transcriptomic atlas in MSS CRC tumor immune microenvironment. (a, g) tSNE plots of lymphoid-derived (a) and myeloid-derived (g) cell subtypes of CRC patients. Each cluster is color-coded according to cell type. Cluster annotations and comparisons of proportions are indicated in the figure. (b, h) Boxplots of lymphoid-derived (b) and myeloid-derived (h) cell subtype fractions in MSI, MSS, and normal tissues (Kruskal-Wallis test; $^{*}P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$, $^{****}P<0.0001$; ns: not significant). (c, i) Heatmaps showing the odds ratio (OR) value of each lymphoid-derived (c) or myeloid-derived (i) cell subtype in MSI, MSS, and normal tissues. (d, j) Comparisons of cell–cell communications of malignant epithelial and lymphoid-derived (d) or myeloid-derived (j) cell subtypes between MSS and MSI tumors by cell-chat analysis. Blue indicates a higher interaction weight in MSS tumor, and red indicates a higher interaction weight in MSI tumor. (e, f, k) Pseudotime analyses of CD4^{+} T cells (e), CD8^{+} T cells (f), and monocytes (k) from MSS CRC patients. CD4^{+} T cell, CD8^{+} T cell, and monocyte subtypes are labeled by colors. cm: central memory; plf: proliferated; rm: tissue-resident memory; rtm: resident tissue; cDC: conventional dendritic cell; TAM: tumor-associated macrophage.

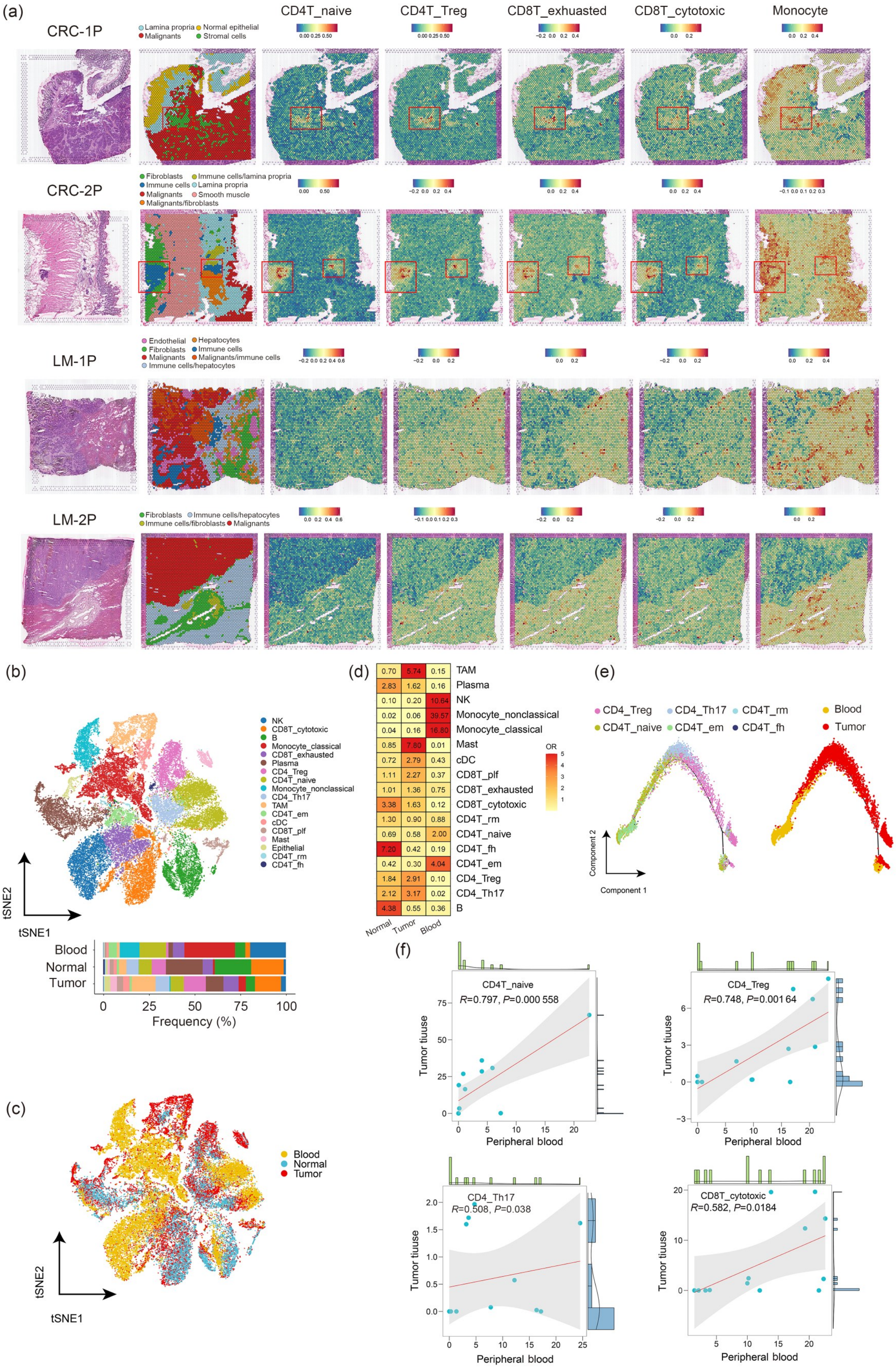

(a)
CD4T_naive
CD4T_Treg
CD8T_exhuasted
CD8T_cytotoxic
Monocyte
CRC-1P
Lamina propria
Normal epithelial
Malignants
Stromal cells
CRC-2P
Fibroblasts
Immune cells/lamina propria
Immune cells
Lamina propria
Malignants
Smooth muscle
Malignants/fibroblasts
LM-1P
Endothelial
Hepatocytes
Fibroblasts
Immune cells
Malignants
Malignants/immune cells
Immune cells/hepatocytes
LM-2P
Fibroblasts
Immune cells/hepatocytes
Immune cells/fibroblasts
Malignants
(b)
NK
CD8T_cytotoxic
B
Monocyte_classical
CD8T_exhausted
Plasma
CD4_Treg
CD4T_naive
Monocyte_nonclassical
CD4_Th17
TAM
CD4T_em
cDC
CD8T_plf
Mast
Epithelial
CD4T_rm
CD4T_fh
tSNE2
tSNE1
Blood
Normal
Tumor
0
25
50
75
100
Frequency (%)
(c)
Blood
Normal
Tumor
tSNE2
tSNE1
(d)
0.70 5.74 0.15 TAM
2.83 1.62 0.16 Plasma
0.10 0.20 10.64 NK
0.02 0.06 39.57 Monocyte_nonclassical
0.04 0.16 16.80 Monocyte_classical
0.85 7.80 0.01 Mast
0.72 2.79 0.43 cDC
1.11 2.27 0.37 CD8T_plf
1.01 1.36 0.75 CD8T_exhausted
3.38 1.63 0.12 CD8T_cytotoxic
1.30 0.90 0.88 CD4T_rm
0.69 0.58 2.00 CD4T_naive
7.20 0.42 0.19 CD4T_fh
0.42 0.30 4.04 CD4T_em
1.84 2.91 0.10 CD4_Treg
2.12 3.17 0.02 CD4_Th17
4.38 0.55 0.36 B
Normal
Tumor
Blood
OR
5
4
3
2
1
(e)
CD4_Treg
CD4_Th17
CD4T_rm
CD4T_naive
CD4T_em
CD4T_fh
Blood
Tumor
Component 2
Component 1
(f)
CD4T_naive
R=0.797, P=0.000 558
Tumor tiuuse
Peripheral blood
CD4_Treg
R=0.748, P=0.001 64
Tumor tiuuse
Peripheral blood
CD4_Th17
R=0.508, P=0.038
Tumor tiuuse
Peripheral blood
CD8T_cytotoxic
R=0.582, P=0.0184
Tumor tiuuse
Peripheral blood

Fig. 3 Spatial and peripheral blood single-cell transcriptomic landscape in MSS CRC. (a) The H&E staining sections, unsupervised clustering analysis, and MSS CRC-enriched immune cell distribution of spatial transcriptomic data from two CRC patients with primary tumors and liver metastasis (LM) (CRC-1P, CRC-2P, LM-1P, and LM-2P). (b) tSNE plot of immune cell subtypes of CRC patients. (c) tSNE plot of immune cell subtypes from the tumor, normal, and peripheral blood samples of CRC patients. Each cluster is color-coded according to cell type. Cluster annotations and comparisons of proportions are indicated in the figure. (d) Heatmap showing the odds ratio (OR) value of each immune cell subtype in the tumor, normal, and peripheral blood samples of CRC patients. (e) Pseudotime analysis of CD4$^+$ T cells from MSS CRC patients. CD4$^+$ T cell subtypes are labeled by color. (f) Pearson correlation of MSS CRC-enriched immune cell type proportions in tumor and peripheral blood samples. TAM: tumor-associated macrophage; cDC: conventional dendritic cell; em: effector memory; fh: follicular helper; plf: proliferated; rm: tissue-resident memory.

scRNA-seq data from blood samples of CRC patients. After unsupervised clustering, we obtained a total of 16 immune cell subtypes, and the majority of these cell types were consistent with our previously annotated cell types from MSS CRC (Figs. 2a, 2g, 3b, and 3c). The NK cells, monocytes, and naive and effector memory CD4$^+$ T cells showed a strong distribution preference in the peripheral blood, while functional and terminally differentiated T cells and macrophages appeared to be enriched in the tissue, and the pseudotime trajectory also presented an immune cell differentiation trend from the peripheral blood to the tissue (Figs. 3d and 3e, Table S9). Interestingly, we found a significantly positive correlation between the abundances of several immune cell subtypes, including naive CD4$^+$ T, regulatory CD4$^+$ T, cytotoxic CD8$^+$ T, CD4$^+$ Th17, and B cells, in blood and tumor tissues (Figs. 3f and S4b). Therefore, we concluded that the content of these cell subsets in the peripheral blood could reflect the cell infiltration level in the TME, which provided the feasibility of liquid biopsy-based biomarker screening. Altogether, we comprehensively analyzed the immune characteristics of MSS CRC tumors from three dimensions including tissue spatial distribution and tissue and peripheral blood single-cell transcriptomics.

3.5 Identification of MSS CRC immune cell-related modules and hub genes

To further identify the MSS CRC immune cell-related modules and hub genes, the WGCNA procedure was used to explore the gene modules related to three main MSS CRC-enriched immune cell types including T cells (composed of naive CD4$^+$ T, regulatory CD4$^+$ T, CD4$^+$ Th17, exhausted CD8$^+$ T, cytotoxic CD8$^+$ T, proliferated CD8$^+$ T, and NK cells), plasma cells, and monocytes (composed of classical and intermediated monocytes). First, to obtain the signature of the three main immune cell types, we selected the top DEGs of each cell subtype from the scRNA-seq data as markers and then calculated the immune cell infiltration level for each patient using the GSVA algorithm on the TCGA-CRC cohort (Tables S5 and S7). Obviously, three signatures could properly represent the three main immune cell types, and the survival analysis showed that higher T and plasma cell infiltration and lower monocyte infiltration were associated with superior OS (Figs. 4a–4f). We then set the soft threshold β as 9 (R^2=0.908) to provide a suitable power value for co-expression network construction (Fig. S5a). Eighteen modules indicated by different colors were determined by dynamic tree (Fig. S5b). In addition, the correlations between modules and the infiltration of three immune cell types were calculated. Interestingly, we observed that the dark green module had the highest correlation with the infiltration of all three immune cell types in the module–trait relationship (Fig. 4g). In this module, the correlation coefficients between GS and MM reached 0.69, 0.67, and 0.82, suggesting a good quality of module construction (Figs. 4h–4j). To identify the hub genes derived from the infiltration of three immune cell types within the dark green module, 816 genes with GS>0.5 and MM>0.6 were considered hub genes. With the intersection of DEGs of MSS CRC-enriched immune cell subtypes from scRNA-seq data, a total of 108 overlapping genes were extracted for subsequent analyses (Fig. 4k).

3.6 Construction of MCICRS

Based on the expression profiles of 108 MSS CRC immune cell-related genes, we performed multivariate Cox regression analysis to identify prognostic genes, adjusting for age, gender, and TNM stage. To ensure the stability of OS-related genes, we used the random-effects model of systematic meta-analysis to pool the HR values of 108 hub genes in three independent MSS CRC cohorts (TCGA-CRC (n=418), GSE39582 (n=459), and GSE41258 (n=137)). Finally, we obtained 39 stable and significant prognostic immune cell-related

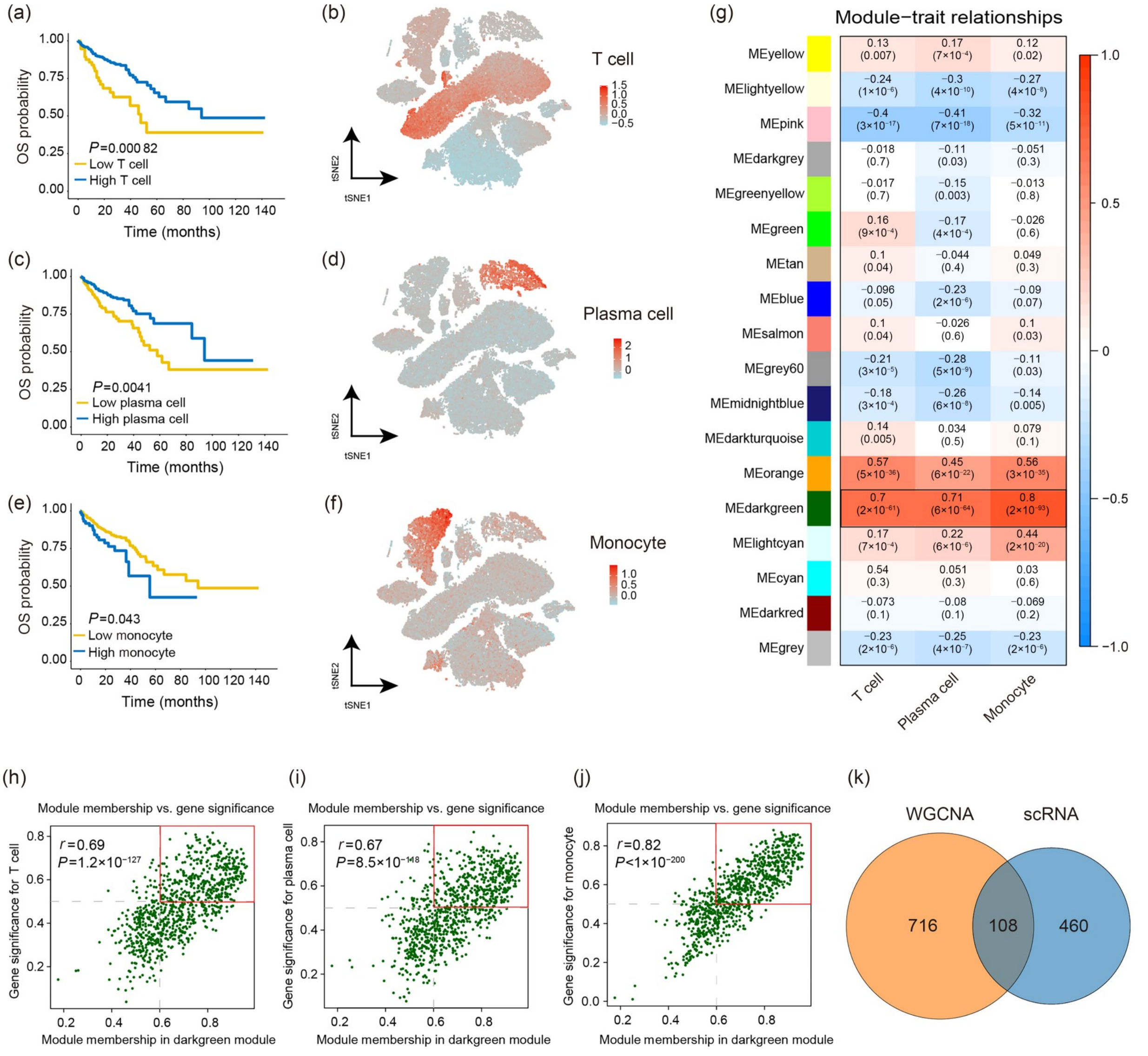


Fig. 4 Identification of MSS CRC immune cell-related modules and hub genes. (a, c, e) The overall survival (OS) analysis for the MSS TCGA-CRC cohort with the infiltration levels of different T cells (a), plasma cells (c), and monocytes (e) (log-rank test). (b, d, f) The expression levels of T cell (b), plasma cell (d), and monocyte (f) signatures in the tSNE plots. (g) Correlation analyses between Module Eigengenes (MEs) and immune infiltration of T cell, plasma cell, and monocyte. (h–j) The high correlations between gene significance (GS) values of T cell (h), plasma cell (i), and monocyte (j) and module membership (MM) in the darkgreen module. Dots within the red rectangle were defined as immune cell-related genes, with both high GS and MM. (k) Overleaping of immune cell-related genes between WGCNA and scRNA-seq data. *r*: correlation coefficient.

genes for MSS CRC patients (meta-P<0.001; Fig. 5a). Thirty-nine genes were then subjected to the machine learning-based integrative procedure to develop an MCICRS. Based on ten machine learning algorithms, we fitted a total of 96 kinds of prediction models with 10-fold cross-validation, and calculated the *C*-index of each model across the TCGA-CRC training dataset and three independent validation datasets (GSE39582 (n=459), GSE39084 (n=54), and GSE41258 (n=137)). Here, we found that the combination of RSF and stepwise Cox (direction=forward) algorithms had the best performance with the highest average *C*-index (0.681) in all datasets (Fig. 5b, Table S10). A final set of 13 top genes was filtrated in the RSF model and then subjected to the stepwise Cox proportional hazards regression model to obtain their corresponding regression coefficients (Figs. 5c–5e). Next, based on the expression of 13 genes weighted by their regression

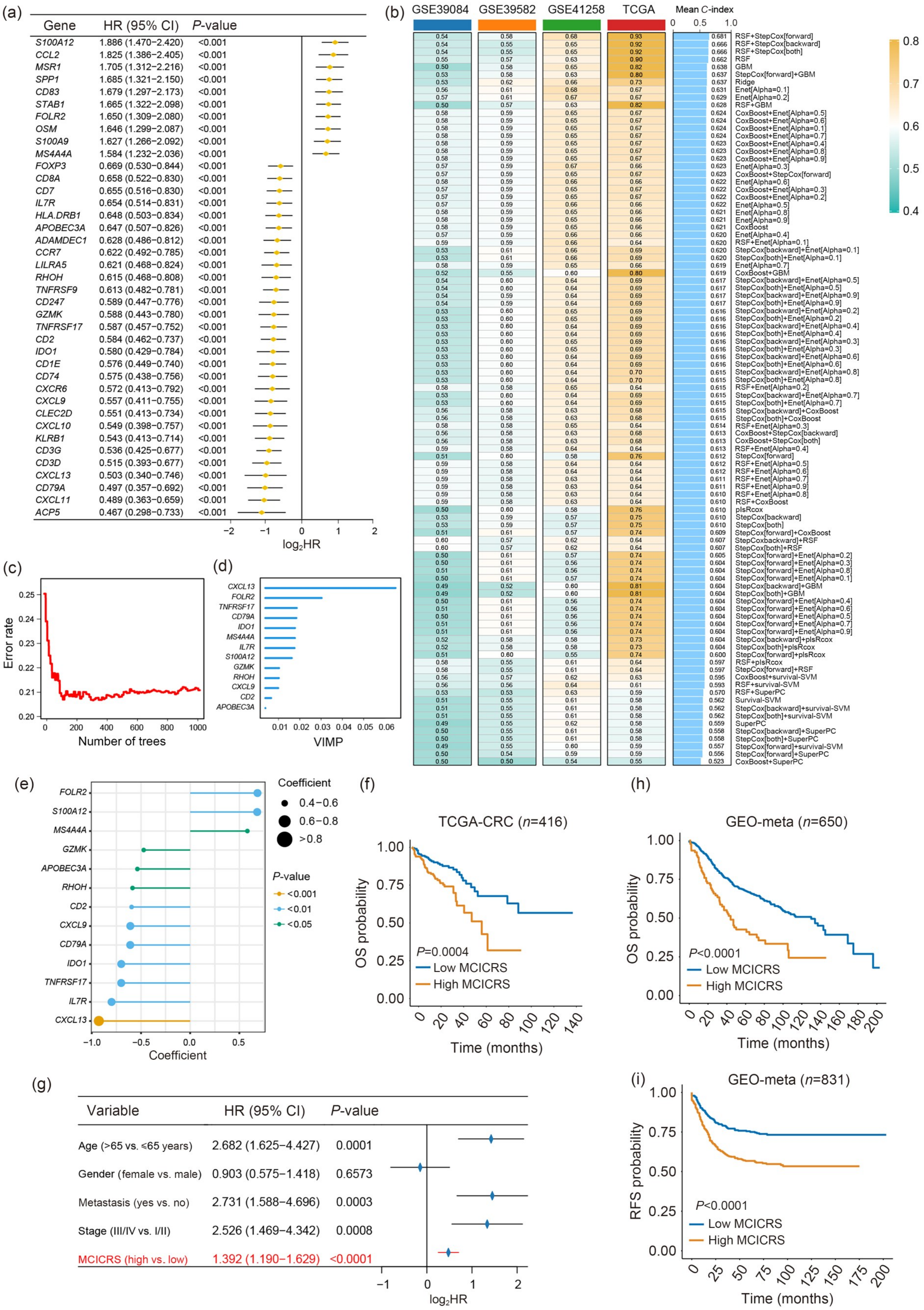

(a)
Gene
HR (95% CI)
P-value
S100A12 1.886 (1.470−2.420) <0.001
CCL2 1.825 (1.386−2.405) <0.001
MSR1 1.705 (1.312−2.216) <0.001
SPP1 1.685 (1.321−2.150) <0.001
CD83 1.679 (1.297−2.173) <0.001
STAB1 1.665 (1.322−2.098) <0.001
FOLR2 1.650 (1.309−2.080) <0.001
OSM 1.646 (1.299−2.087) <0.001
S100A9 1.627 (1.266−2.092) <0.001
MS4A4A 1.584 (1.232−2.036) <0.001
FOXP3 0.669 (0.530−0.844) <0.001
CD8A 0.658 (0.522−0.830) <0.001
CD7 0.655 (0.516−0.830) <0.001
IL7R 0.654 (0.514−0.831) <0.001
HLA.DRB1 0.648 (0.503−0.834) <0.001
APOBEC3A 0.647 (0.507−0.826) <0.001
ADAMDEC1 0.628 (0.486−0.812) <0.001
CCR7 0.622 (0.492−0.785) <0.001
LILRA5 0.621 (0.468−0.824) <0.001
RHOH 0.615 (0.468−0.808) <0.001
TNFRSF9 0.613 (0.482−0.781) <0.001
CD247 0.589 (0.447−0.776) <0.001
GZMK 0.588 (0.443−0.780) <0.001
TNFRSF17 0.587 (0.457−0.752) <0.001
CD2 0.584 (0.462−0.737) <0.001
IDO1 0.580 (0.429−0.784) <0.001
CD1E 0.576 (0.449−0.740) <0.001
CD74 0.575 (0.438−0.756) <0.001
CXCR6 0.572 (0.413−0.792) <0.001
CXCL9 0.557 (0.411−0.755) <0.001
CLEC2D 0.551 (0.413−0.734) <0.001
CXCL10 0.549 (0.398−0.757) <0.001
KLRB1 0.543 (0.413−0.714) <0.001
CD3G 0.536 (0.425−0.677) <0.001
CD3D 0.515 (0.393−0.677) <0.001
CXCL13 0.503 (0.340−0.746) <0.001
CD79A 0.497 (0.357−0.692) <0.001
CXCL11 0.489 (0.363−0.659) <0.001
ACP5 0.467 (0.298−0.733) <0.001
−2 −1 0 1 2
log₂HR
(b)
GSE39084
GSE39582
GSE41258
TCGA
Mean C-index
0 0.5 1.0
0.8
0.7
0.6
0.5
0.4
0.681 RSF+StepCox[forward]
0.666 RSF+StepCox[backward]
0.666 RSF+StepCox[both]
0.662 RSF
0.638 GBM
0.637 StepCox[forward]+GBM
0.637 Ridge
0.631 Enet[Alpha=0.1]
0.629 Enet[Alpha=0.2]
0.628 RSF+GBM
0.624 CoxBoost+Enet[Alpha=0.5]
0.624 CoxBoost+Enet[Alpha=0.6]
0.624 CoxBoost+Enet[Alpha=0.1]
0.624 CoxBoost+Enet[Alpha=0.7]
0.623 CoxBoost+Enet[Alpha=0.4]
0.623 CoxBoost+Enet[Alpha=0.8]
0.623 CoxBoost+Enet[Alpha=0.9]
0.623 Enet[Alpha=0.3]
0.623 CoxBoost+StepCox[forward]
0.622 Enet[Alpha=0.6]
0.622 CoxBoost+Enet[Alpha=0.3]
0.622 CoxBoost+Enet[Alpha=0.2]
0.622 Enet[Alpha=0.5]
0.621 Enet[Alpha=0.8]
0.621 Enet[Alpha=0.9]
0.621 CoxBoost
0.620 Enet[Alpha=0.4]
0.620 RSF+Enet[Alpha=0.1]
0.620 StepCox[backward]+Enet[Alpha=0.1]
0.620 StepCox[both]+Enet[Alpha=0.1]
0.619 Enet[Alpha=0.7]
0.619 CoxBoost+GBM
0.617 StepCox[backward]+Enet[Alpha=0.5]
0.617 StepCox[both]+Enet[Alpha=0.5]
0.617 StepCox[backward]+Enet[Alpha=0.9]
0.617 StepCox[both]+Enet[Alpha=0.9]
0.616 StepCox[backward]+Enet[Alpha=0.2]
0.616 StepCox[both]+Enet[Alpha=0.2]
0.616 StepCox[backward]+Enet[Alpha=0.4]
0.616 StepCox[both]+Enet[Alpha=0.4]
0.616 StepCox[backward]+Enet[Alpha=0.3]
0.616 StepCox[both]+Enet[Alpha=0.3]
0.616 StepCox[backward]+Enet[Alpha=0.6]
0.616 StepCox[both]+Enet[Alpha=0.6]
0.615 StepCox[backward]+Enet[Alpha=0.8]
0.615 StepCox[both]+Enet[Alpha=0.8]
0.615 RSF+Enet[Alpha=0.2]
0.615 StepCox[backward]+Enet[Alpha=0.7]
0.615 StepCox[both]+Enet[Alpha=0.7]
0.615 StepCox[backward]+CoxBoost
0.615 StepCox[both]+CoxBoost
0.614 RSF+Enet[Alpha=0.3]
0.613 CoxBoost+StepCox[backward]
0.613 CoxBoost+StepCox[both]
0.613 RSF+Enet[Alpha=0.4]
0.612 StepCox[forward]
0.612 RSF+Enet[Alpha=0.5]
0.612 RSF+Enet[Alpha=0.6]
0.611 RSF+Enet[Alpha=0.7]
0.611 RSF+Enet[Alpha=0.9]
0.610 RSF+Enet[Alpha=0.8]
0.610 RSF+CoxBoost
0.610 plsRcox
0.610 StepCox[backward]
0.610 StepCox[both]
0.609 StepCox[forward]+CoxBoost
0.607 StepCoxbackward]+RSF
0.607 StepCox[both]+RSF
0.605 StepCox[forward]+Enet[Alpha=0.2]
0.604 StepCox[forward]+Enet[Alpha=0.3]
0.604 StepCox[forward]+Enet[Alpha=0.8]
0.604 StepCox[forward]+Enet[Alpha=0.1]
0.604 StepCox[backward]+GBM
0.604 StepCox[both]+GBM
0.604 StepCox[forward]+Enet[Alpha=0.4]
0.604 StepCox[forward]+Enet[Alpha=0.6]
0.604 StepCox[forward]+Enet[Alpha=0.5]
0.604 StepCox[forward]+Enet[Alpha=0.7]
0.604 StepCox[forward]+Enet[Alpha=0.9]
0.604 StepCox[backward]+plsRcox
0.604 StepCox[both]+plsRcox
0.600 StepCox[forward]+plsRcox
0.597 RSF+plsRcox
0.597 StepCox[forward]+RSF
0.595 CoxBoost+survival-SVM
0.593 RSF+survival-SVM
0.570 RSF+SuperPC
0.562 Survival-SVM
0.562 StepCox[backward]+survival-SVM
0.562 StepCox[both]+survival-SVM
0.559 SuperPC
0.558 StepCox[backward]+SuperPC
0.558 StepCox[both]+SuperPC
0.557 StepCox[forward]+survival-SVM
0.556 StepCox[forward]+SuperPC
0.523 CoxBoost+SuperPC
(c)
Error rate
0.25
0.24
0.23
0.22
0.21
0.20
0 200 400 600 800 1000
Number of trees
(d)
CXCL13
FOLR2
TNFRSF17
CD79A
IDO1
MS4A4A
IL7R
S100A12
GZMK
RHOH
CXCL9
CD2
APOBEC3A
0.01 0.02 0.03 0.04 0.05 0.06
VIMP
(e)
FOLR2
S100A12
MS4A4A
GZMK
APOBEC3A
RHOH
CD2
CXCL9
CD79A
IDO1
TNFRSF17
IL7R
CXCL13
−1.0 −0.5 0.0 0.5
Coefficient
Coefficient
● 0.4−0.6
● 0.6−0.8
● >0.8
P-value
<0.001
<0.01
<0.05
(f)
TCGA-CRC (n=416)
OS probability
1.00
0.75
0.50
0.25
0.00
P=0.0004
Low MCICRS
High MCICRS
0 20 40 60 80 100 120 140
Time (months)
(h)
GEO-meta (n=650)
OS probability
1.00
0.75
0.50
0.25
0.00
P<0.0001
Low MCICRS
High MCICRS
0 20 40 60 80 100 120 140 160 180 200
Time (months)
(g)
Variable
HR (95% CI)
P-value
Age (>65 vs. ≤65 years) 2.682 (1.625−4.427) 0.0001
Gender (female vs. male) 0.903 (0.575−1.418) 0.6573
Metastasis (yes vs. no) 2.731 (1.588−4.696) 0.0003
Stage (III/IV vs. I/II) 2.526 (1.469−4.342) 0.0008
MCICRS (high vs. low) 1.392 (1.190−1.629) <0.0001
−1 0 1 2
log₂HR
(i)
GEO-meta (n=831)
RFS probability
1.00
0.75
0.50
0.25
0.00
P<0.0001
Low MCICRS
High MCICRS
0 25 50 75 100 125 150 175 200
Time (months)

Fig. 5 Construction of MSS CRC immune cell-related signature. (a) Forest plot showing the pooled hazard ratio (HR) from meta-analysis of prognostic immune cell-related genes. (b) A total of 96 kinds of prediction models based on machine learning methods and further calculations of the *C*-index of each model across the training and validation datasets. (c, d) The number of trees for determining the minimal error (c) and importance (d) of the 13 most valuable genes based on the random survival forest (RSF) algorithm. (e) Coefficients of 13 genes finally obtained in the stepwise Cox regression model. (f) Overall survival (OS) analysis according to MCICRS for the TCGA-CRC cohort (log-rank test). (g) Multivariable Cox regression analysis indicates that MCICRS is an independent risk factor for OS in the TCGA-CRC cohort. (h) OS analysis according to MCICRS for the GEO-meta cohort (log-rank test). (i) RFS analysis according to MCICRS for the GEO-meta cohort (log-rank test). VIMP: variable importance.

coefficients, we calculated the MCICRS for each MSS CRC patient, and assigned them to the high- and low-MCICRS groups according to the optimal cut-off value. In the TCGA-CRC cohort, patients with a low MCICRS had a significantly prolonged OS ($P<0.05$), and the multivariate Cox regression analysis indicated that MCICRS was an independent and significant risk factor for OS after adjusting for age, gender, metastasis, and TNM stage (Figs. 5f and 5g). In addition, we evaluated the predictive value of MCICRS for OS and RFS in all six GEO datasets and the GEO-meta cohort, and found the same trend (Figs. 5h, 5i, and S6). Moreover, we compared the performance of MCICRS with that of other existing signatures from published literature and prior research and, ultimately, 104 signatures associated with various biological processes were enrolled, including immune response, autophagy, and epithelial–mesenchymal transition (Table S11). Based on univariate Cox regression, we compared the *C*-index of MCICRS with that of other signatures across four datasets. Notably, MCICRS displayed a better performance than most other signatures in the three datasets except GSE39084, which demonstrated the stability of MCICRS (Fig. S7).

3.7 Immune infiltration characteristics of MCICRS and its implication in immunotherapy

Since the construction of MCICRS is based on immune cell-related genes, we assumed that there existed distinct immune characteristics between the high- and low-MCICRS groups, which may explain their prognostic differences. According to the ssGSEA method, we found higher infiltration abundances of 28 immune cells in the low-MCICRS group (Fig. 6a). To ensure that the assessments were not biased by the single algorithm, five other methods, including EPIC, MCP-counter, quanTIseq, TIMER, and xCell, were used to verify the stability and robustness of the ssGSEA results (Fig. S8). In addition, the ESTIMATE method also showed a higher immune score in the low-MCICRS group (Fig. 6b). Furthermore, GSEA and GSVA analyses demonstrated that several immune-related pathways, such as cytokine, chemokine, tumor necrosis factor α (TNFA), and interferon signaling pathways, were activated in the low-MCICRS group (Figs. 6c and 6d). In addition, scatter plots of MCICRS and CD274 demonstrated a positive correlation in the scRNA-seq data (Fig. S9a). We also calculated the MCICRS score of each spot in the spatial transcriptomic data, and found a consistent spatial distribution with the immune cells (Fig. S9b). We also used proteomap, a graphical tool to quantitate and visualize the composition and function of proteomes, to compare the differences between the high- and low-MCICRS groups. Interestingly, the pattern of the proteomap in the low-MCICRS group showed a higher enrichment of proteins related to the JAK/signal transducer and activator of transcription (STAT), cytokine–cytokine receptor interaction, and nuclear factor-κB (NF-κB) and tumor necrosis factor (TNF) signaling pathways compared with the high-MCICRS group (Fig. 6e). Collectively, these results confirmed an activated immune status in low-MCICRS patients; hence, we inferred that MCICRS could predict immunotherapy effects in MSS CRC. The GSE179351 study evaluated the efficacy of combining radiation, ipilimumab, and nivolumab in MSS CRC patients (Parikh et al., 2021). After applying the MCICRS to the RNA-seq data from pre-treatment biopsy samples, we observed prolonged OS in low-MCICRS patients, although the statistical difference was not significant because of the small sample size (Fig. 6f). Furthermore, we enrolled 22 public immunotherapy cohorts with treatment response or survival data to assess the predictive value of MCICRS in immunotherapy across pan-cancer types using a systematic meta-analysis. Strikingly, we observed that a higher MCICRS was significantly associated with the progression of treatment response and dismal survival when the OR and HR values of each cohort were pooled in the random-effects model, indicating that MCICRS was a reliable biomarker for immunotherapy (Figs. 6g and 6h).

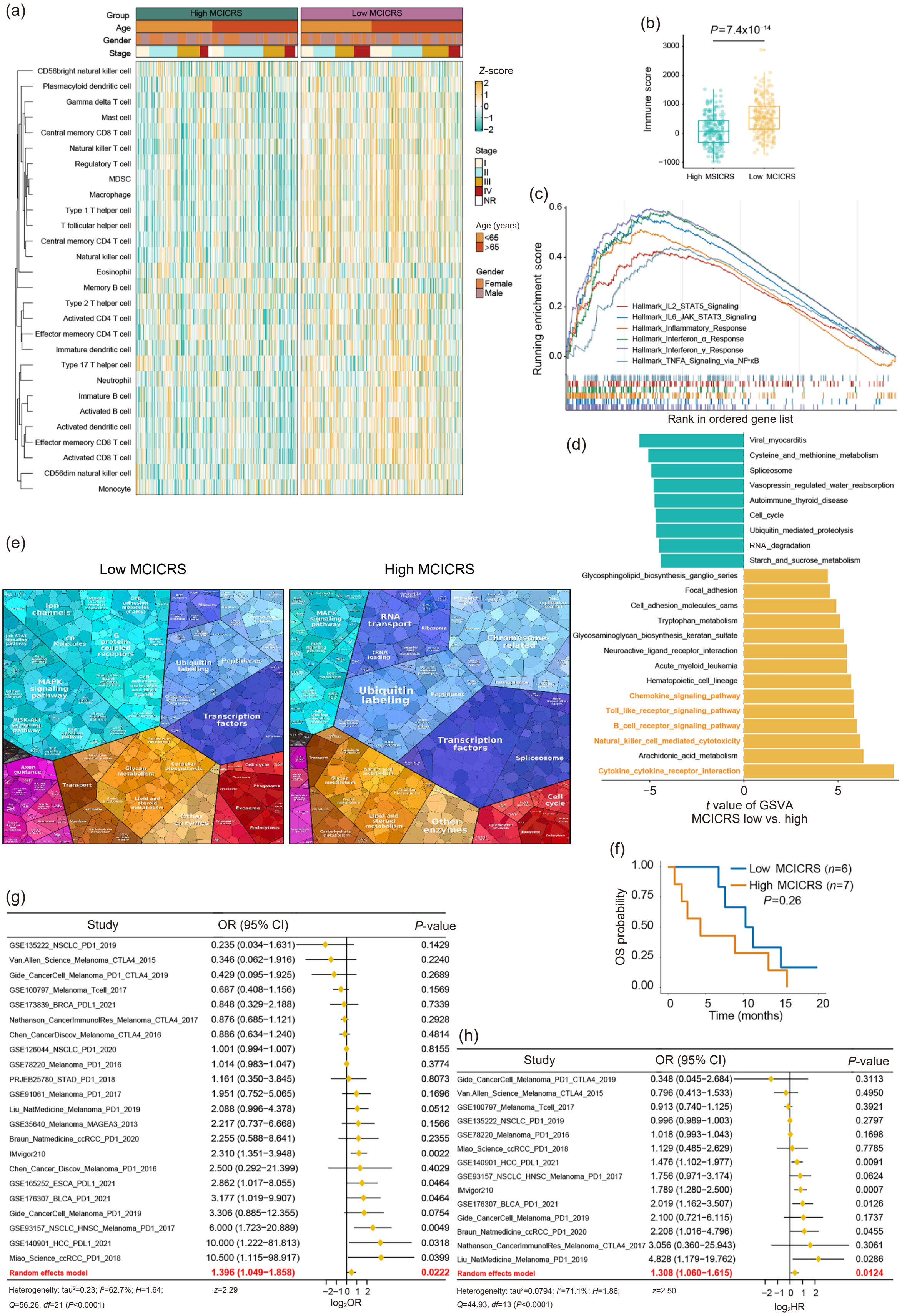

(a)
Group
Age
Gender
Stage
High MCICRS
Low MCICRS
CD56bright natural killer cell
Plasmacytoid dendritic cell
Gamma delta T cell
Mast cell
Central memory CD8 T cell
Natural killer T cell
Regulatory T cell
MDSC
Macrophage
Type 1 T helper cell
T follicular helper cell
Central memory CD4 T cell
Natural killer cell
Eosinophil
Memory B cell
Type 2 T helper cell
Activated CD4 T cell
Effector memeory CD4 T cell
Immature dendritic cell
Type 17 T helper cell
Neutrophil
Immature B cell
Activated B cell
Activated dendritic cell
Effector memeory CD8 T cell
Activated CD8 T cell
CD56dim natural killer cell
Monocyte
Z-score
2
1
0
−1
−2
Stage
I
II
III
IV
NR
Age (years)
<65
>65
Gender
Female
Male
(b)
P=7.4x10−14
Immune score
3000
2000
1000
0
−1000
High MSICRS
Low MCICRS
(c)
Running enrichment score
0.6
0.4
0.2
0.0
Hallmark_IL2_STAT5_Signaling
Hallmark_IL6_JAK_STAT3_Signaling
Hallmark_Inflammatory_Response
Hallmark_Interferon_α_Response
Hallmark_Interferon_γ_Response
Hallmark_TNFA_Signaling_via_NF-κB
Rank in ordered gene list
(d)
Viral_myocarditis
Cysteine_and_methionine_metabolism
Spliceosome
Vasopressin_regulated_water_reabsorption
Autoimmune_thyroid_disease
Cell_cycle
Ubiquitin_mediated_proteolysis
RNA_degradation
Starch_and_sucrose_metabolism
Glycosphingolipid_biosynthesis_ganglio_series
Focal_adhesion
Cell_adhesion_molecules_cams
Tryptophan_metabolism
Glycosaminoglycan_biosynthesis_keratan_sulfate
Neuroactive_ligand_receptor_interaction
Acute_myeloid_leukemia
Hematopoietic_cell_lineage
Chemokine_signaling_pathway
Toll_like_receptor_signaling_pathway
B_cell_receptor_signaling_pathway
Natural_killer_cell_mediated_cytotoxicity
Arachidonic_acid_metabolism
Cytokine_cytokine_receptor_interaction
−5
0
5
t value of GSVA
MCICRS low vs. high
(e)
Low MCICRS
High MCICRS
Ion channels
Cell adhesion molecules (CAMs)
CD Molecules
G protein-coupled receptors
MAPK signaling pathway
PI3K-Akt signaling pathway
Ubiquitin labeling
Peptidases
Transcription factors
Axon guidance
Transport
Glycan metabolism
Cofactor biosynthesis
Lipid and steroid metabolism
Other enzymes
Cell cycle
Phagosome
Exosome
Endocytosis
RNA transport
tRNA loading
Chromosome-related
Spliceosome
Carbohydrate metabolism
(f)
Low MCICRS (n=6)
High MCICRS (n=7)
P=0.26
OS probability
1.00
0.75
0.50
0.25
0.00
0
5
10
15
20
Time (months)
(g)
Study
OR (95% CI)
P-value
GSE135222_NSCLC_PD1_2019 0.235 (0.034−1.631) 0.1429
Van.Allen_Science_Melanoma_CTLA4_2015 0.346 (0.062−1.916) 0.2240
Gide_CancerCell_Melanoma_PD1_CTLA4_2019 0.429 (0.095−1.925) 0.2689
GSE100797_Melanoma_Tcell_2017 0.687 (0.408−1.156) 0.1569
GSE173839_BRCA_PDL1_2021 0.848 (0.329−2.188) 0.7339
Nathanson_CancerImmunolRes_Melanoma_CTLA4_2017 0.876 (0.685−1.121) 0.2928
Chen_CancerDiscov_Melanoma_CTLA4_2016 0.886 (0.634−1.240) 0.4814
GSE126044_NSCLC_PD1_2020 1.001 (0.994−1.007) 0.8155
GSE78220_Melanoma_PD1_2016 1.014 (0.983−1.047) 0.3774
PRJEB25780_STAD_PD1_2018 1.161 (0.350−3.845) 0.8073
GSE91061_Melanoma_PD1_2017 1.951 (0.752−5.065) 0.1696
Liu_NatMedicine_Melanoma_PD1_2019 2.088 (0.996−4.378) 0.0512
GSE35640_Melanoma_MAGEA3_2013 2.217 (0.737−6.668) 0.1566
Braun_Natmedicine_ccRCC_PD1_2020 2.255 (0.588−8.641) 0.2355
IMvigor210 2.310 (1.351−3.948) 0.0022
Chen_Cancer_Discov_Melanoma_PD1_2016 2.500 (0.292−21.399) 0.4029
GSE165252_ESCA_PDL1_2021 2.862 (1.017−8.055) 0.0464
GSE176307_BLCA_PD1_2021 3.177 (1.019−9.907) 0.0464
Gide_CancerCell_Melanoma_PD1_2019 3.306 (0.885−12.355) 0.0754
GSE93157_NSCLC_HNSC_Melanoma_PD1_2017 6.000 (1.723−20.889) 0.0049
GSE140901_HCC_PDL1_2021 10.000 (1.222−81.813) 0.0318
Miao_Science_ccRCC_PD1_2018 10.500 (1.115−98.917) 0.0399
Random effects model 1.396 (1.049−1.858) 0.0222
Heterogeneity: tau²=0.23; I²=62.7%; H=1.64;
Q=56.26, df=21 (P<0.0001)
z=2.29
−2 −1 0 1 2
log₂OR
(h)
Study
OR (95% CI)
P-value
Gide_CancerCell_Melanoma_PD1_CTLA4_2019 0.348 (0.045−2.684) 0.3113
Van.Allen_Science_Melanoma_CTLA4_2015 0.796 (0.413−1.533) 0.4950
GSE100797_Melanoma_Tcell_2017 0.913 (0.740−1.125) 0.3921
GSE135222_NSCLC_PD1_2019 0.996 (0.989−1.003) 0.2797
GSE78220_Melanoma_PD1_2016 1.018 (0.993−1.043) 0.1698
Miao_Science_ccRCC_PD1_2018 1.129 (0.485−2.629) 0.7785
GSE140901_HCC_PDL1_2021 1.476 (1.102−1.977) 0.0091
GSE93157_NSCLC_HNSC_Melanoma_PD1_2017 1.756 (0.971−3.174) 0.0624
IMvigor210 1.789 (1.280−2.500) 0.0007
GSE176307_BLCA_PD1_2021 2.019 (1.162−3.507) 0.0126
Gide_CancerCell_Melanoma_PD1_2019 2.100 (0.721−6.115) 0.1737
Braun_Natmedicine_ccRCC_PD1_2020 2.208 (1.016−4.796) 0.0455
Nathanson_CancerImmunolRes_Melanoma_CTLA4_2017 3.056 (0.360−25.943) 0.3061
Liu_NatMedicine_Melanoma_PD1_2019 4.828 (1.179−19.762) 0.0286
Random effects model 1.308 (1.060−1.615) 0.0124
Heterogeneity: tau²=0.0794; I²=71.1%; H=1.86;
Q=44.93, df=13 (P<0.0001)
z=2.50
−2 −1 0 1 2
log₂HR

Fig. 6 Immune infiltration characteristics of MCICRS and its implication in immunotherapy. (a) The infiltration abundances of 28 immune cell subsets evaluated by ssGSEA for the low-MCICRS and high-MCICRS groups. (b) The distribution of immune score inferred by the ESTIMATE algorithm between the low-MCICRS and high-MCICRS groups (Wilcoxon test). (c, d) GSEA enrichment analysis (c) and GSVA analysis (d) showing the activation status of immune-related pathways in the low-MCICRS group. (e) Online proteomaps analysis showing the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway activation between the low-MCICRS and high-MCICRS groups. (f) Overall survival (OS) analysis according to MCICRS for MSS CRC patients receiving immunotherapy (log-rank test). (g, h) Meta-analysis showing that a high MCICRS is significantly associated with the progression (g) and inferior survival (h) of immunotherapy response across pan-cancer types.

3.8 Validation of MCICRS in clinical outcomes of in-house immunotherapy cohort

To further validate the prognostic value of MCICRS in MSS CRC patients receiving immunotherapy, we first calculated the MCICRS in 38 508 cells from 19 MSS CRC tumor samples at single-cell resolution. Surprisingly, the MCICRS value was higher in myeloid cells but lower in T and B cells, indicating that high lymphocyte counts and low myeloid cell counts in patients may be associated with superior survival (Figs. 7a and 7b). In addition, the majority of TME-enriched immune cells were derived and differentiated from peripheral blood immune cells, and the abundances of these cell types between blood and tumor tissues had a positive correlation (Fig. 3f). Therefore, we considered that peripheral blood lymphocytes and monocytes could reflect the infiltration of TME immune cells and serve as biomarker candidates for MSS CRC patients receiving immunotherapy. Here, we detected the peripheral blood lymphocyte and monocyte counts of 54 MSS CRC patients at the initiation of immunotherapy via a hematology analyzer. Likewise, we found that patients with more lymphocytes and fewer monocytes exhibited a better treatment response (PR and SD) and preferable PFS (Figs. 7c–7g). After adjusting for confounding factors, including age, gender, ECOG performance status, primary tumor location, lung metastasis, and liver metastasis, the lymphocyte and monocyte counts remained statistically significant for PFS, supporting the supposition that these two immune cell types were independent predictors for immunotherapy in MSS CRC patients (Fig. 7h). Finally, we developed a nomogram based on lymphocytes, monocytes, and three significant clinical characteristics in a multivariable Cox model with excellent predictive accuracy: the AUCs for predicting PFS at 6, 12, and 18 months were 0.798, 0.797, and 0.845, respectively (Figs. 7i and 7j).

4 Discussion

In this study, we confirmed the improved efficacy and survival benefit of an anti-PD-1 antibody combined with VEGFRi in MSS CRC patients, with manageable toxicity. Based on CRC scRNA-seq and spatial transcriptomic data, we comprehensively deciphered the cell composition, interaction, differentiation, and spatial distribution characteristics, and further identified MSS CRC-enriched immune cell subtypes. Through a machine learning-based integration, MCICRS was developed to accurately predict the efficacy and clinical outcomes for MSS CRC patients receiving immunotherapy.

VEGF-driven angiogenesis affects immune suppression at many different levels, including inhibiting the antigen-presenting cells (APCs) and effector T cells while augmenting the expansion of immune suppressive cells such as Tregs and MDSCs (Rahma and Hodi, 2019). Emerging evidence indicates that the combination of immunotherapy and anti-angiogenic agents could reverse the immune suppression driven by VEGF-related pathways in melanoma, renal cell carcinoma, non-small cell lung cancer, and hepatocellular carcinoma (Hodi et al., 2014; McDermott et al., 2018; Socinski et al., 2018; Finn et al., 2020). However, the application of this combination strategy is limited in clinical practice for MSS CRC patients. Recently, three clinical trials, including regorafenib plus nivolumab in patients with advanced gastric or colorectal cancer (REGONIVO), regorafenib-avelumab combination in patients with microsatellite-stable colorectal cancer (REGOMUNE), and Kim group's study, presented encouraging anti-tumor activity of regorafenib combined with nivolumab or avelumab (anti-PD-L1 antibody) in MSS CRC, with DCRs ranging from 54% to 86% (Fukuoka et al., 2020; Cousin et al., 2021; Kim et al., 2022). Mechanically, our scRNA-seq and spatial transcriptomic data demonstrated a complex interaction

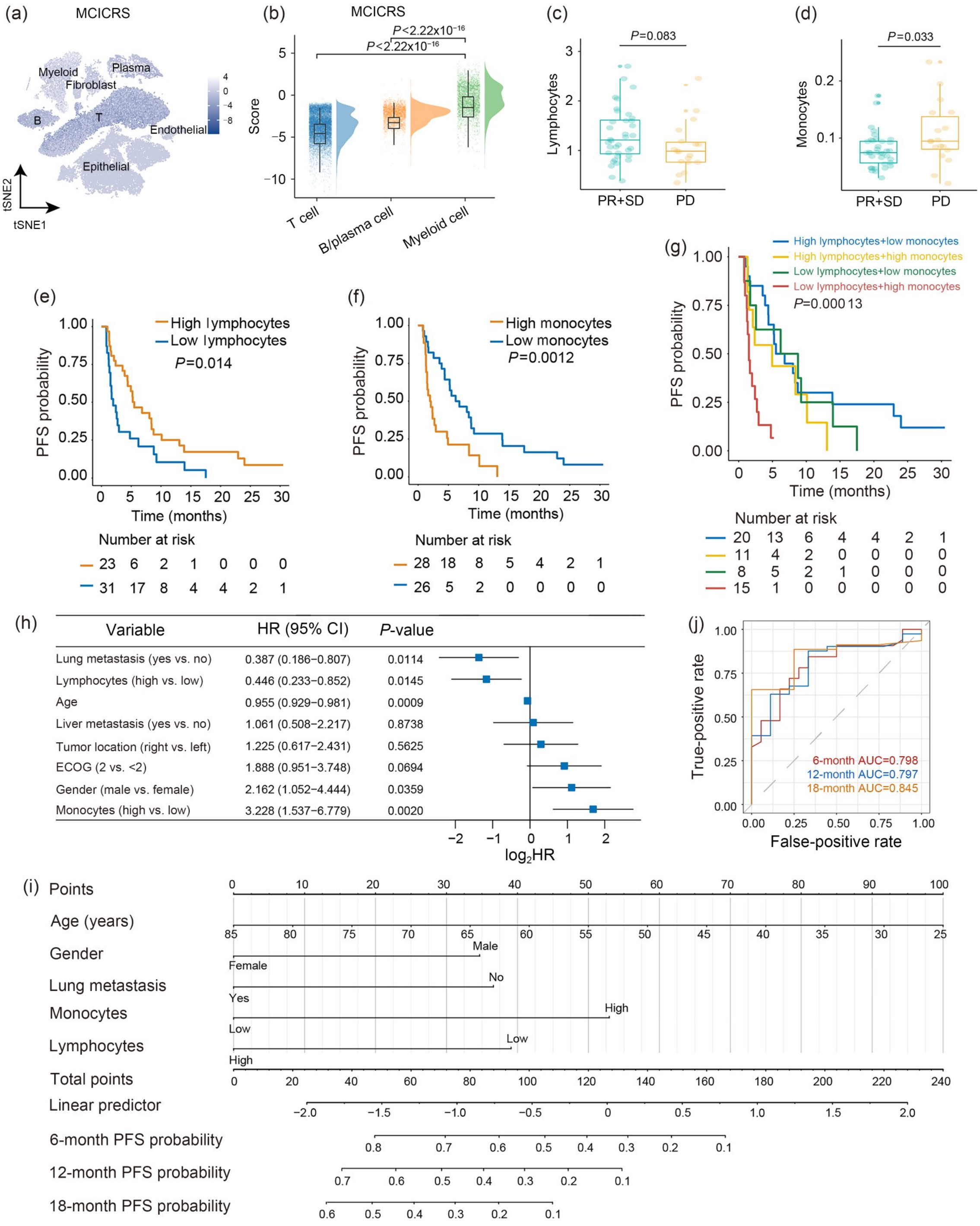


Fig. 7 Validation of MCICRS in clinical outcomes of in-house immunotherapy cohort. (a) The MCICRS level across main cell types of CRC in tSNE plot. (b) Comparison of MCICRS between myeloid cells and T and B/plasma cells (Wilcoxon test). (c, d) Comparisons of peripheral blood lymphocytes (c) and monocytes (d) between the partial response (PR)+stable disease (SD) group and the progressive disease (PD) group for MSS CRC patients receiving immunotherapy (*t*-test). (e–g) Progression-free survival (PFS) analysis according to lymphocyte (c), monocyte (d), and both (g) counts for MSS CRC patients receiving immunotherapy (log-rank test). (h) Multivariable Cox regression analysis indicates that the high lymphocyte and low monocyte counts are two independent factors for superior overall survival (OS) in MSS CRC patients receiving immunotherapy. (i) Nomogram for predicting progression probability for MSS CRC patients receiving immunotherapy. (j) Time-dependent receiver operating characteristic (ROC) analysis for predicting PFS at 6, 12, and 18 months for MSS CRC patients receiving immunotherapy.

between VEGF-driven angiogenic signaling and immune checkpoint-related immunosuppression, shown by the co-localization of VEGFB$^+$ myeloid cells and exhausted CD8$^+$ T cells, which provided a theoretical basis for the potential effect of an immunotherapy and anti-angiogenic therapy combination. As expected, our retrospective study also manifested an improved clinical benefit of anti-PD-1 antibody plus VEGFRi treatment. For MSS CRC patients treated with combination therapy, we observed 63% of DCR, 4.4 months of median PFS, and 10.2 months of median OS, which were highly consistent with the results of REGOMUNE (53% of DCR, 3.6 months of median PFS, and 10.8 months of median OS) and Kim group's study (63% of DCR, 4.3 months of median PFS, and 11.1 months of median OS) (Cousin et al., 2021; Kim et al., 2022). These outcomes were all inferior to those of the REGONIVO study; the main reasons we considered were shown below. First, more than 70% of the patients in our study had liver metastasis compared with 50% of those in the REGONIVO study, which was an independent risk factor for CRC patients (Engstrand et al., 2018). Second, *RAS* mutation could impair the treatment effectiveness and prognosis of CRC and, in our study, the number of patients carrying the *RAS* mutation was almost twice that reported in the REGONIVO study (46.3% vs. 24.0%) (Karapetis et al., 2008). Third, the ECOG performance status of all patients in REGONIVO was 0, whereas that of half of the patients in this study was 2. In this current study, more than half of the patients with SD had a certain degree of tumor shrinkage, and these patients might achieve PR after receiving further cycles of treatment. In addition, in the survival analysis, anti-PD-1 antibody plus VEGFRi combination brought about a 50% decrease in both disease progression risk and death risk compared with VEGFRi alone. In addition, the treatment-related AE profile in this study was consistent with that reported in previous clinical trials (Fukuoka et al., 2020; Cousin et al., 2021; Kim et al., 2022), and the incidence rates between the combination and VEGFRi groups were similar across the majority of AEs. Based on the above results, we considered that anti-PD-1 antibody combined with VEGFRi presented better efficacy and controllable toxicity in MSS CRC compared with VEGFRi alone.

In most scenarios, only 20%–40% of patients responded to immunotherapy and even fewer had long-term disease remission. Therefore, a biomarker-based guide to identify subsets who could benefit from immunotherapy is vitally critical. Currently, PD-L1 expression is the most widely accepted biomarker for guiding the selection of patients to receive anti-PD-1 or anti-PD-L1 antibody (Doroshow et al., 2021). However, REGONIVO did not observe a clear relationship between PD-L1 expression and efficacy outcomes in MSS CRC and, therefore, additional analyses were necessary to clarify the optimal patient population for immunotherapy (Fukuoka et al., 2020). In consideration of the fact that TME-infiltrated immune cells are the main effectors of immunotherapy, in this study, we identified ten MSS CRC-enriched immune cell subtypes and their spatial distribution, including naive CD4$^+$ T, regulatory CD4$^+$ T, CD4$^+$ Th17, exhausted CD8$^+$ T, cytotoxic CD8$^+$ T, proliferated CD8$^+$ T, NK, and IgG plasma cells, and classical and intermediate monocytes, through scRNA-seq along with spatial transcriptomic data. In addition, through WGCNA and marker genes of these MSS CRC-enriched immune cell subtypes, we identified MSS CRC immune cell-related hub genes. Based on the expression profile of hub genes, we first determined 39 stable prognostic genes via meta-analysis in three independent cohorts. Then, a total of 96 kinds of prediction models were fitted in the TCGA-CRC training set and three validation sets. After calculating the *C*-index across all four datasets, we obtained the optimal model of a combination of RSF and stepwise Cox (direction=forward) algorithms, and further developed the MCICRS. Based on a variety of machine learning methods, the integrative procedures used in this study presented the advantage of a fitting prognostic model for MSS CRC, and method combinations could further reduce the dimensionality of variables, making the model more simplified and translational. The following analyses proved that a higher MCICRS was a deleterious indicator of OS and RFS in MSS CRC patients, with inferior immune infiltration and immune-related pathway activity. The meta-analysis of pan-cancer immunotherapy also demonstrated that MCICRS maintained high accuracy and stable performance in predicting immunotherapy efficacy and survival, indicating a great potential for the clinical application of MCICRS.

Several hub genes of MCICRS have been reported to relate to the efficacy of immunotherapy. For example, S100A12 is an S100/calgranulin protein, and its high expression has been proven to correlate with the survival of patients with human papillomavirus

(HPV)-negative hypopharyngeal squamous cell carcinoma receiving immunotherapy (Mints et al., 2021). MS4A4A is a member of the membrane-spanning 4A family and is selectively highly expressed in tumor-associated macrophages, which can regulate the immune escape of tumor cells by promoting M2 polarization of macrophages by activating the phosphoinositide 3-kinase (PI3K)/protein kinase B (AKT) and JAK/STAT6 pathways (Li et al., 2023). A previous study demonstrated that *GZMK*$^+$ effector memory T cells and *CXCL13*$^+$*BHLHE40*$^+$ TH1-like cells were preferentially enriched in MSI-H CRC, which explained their favorable responses to immunotherapy (Zhang et al., 2018). APOBEC3A is a member of the apolipoprotein B messenger RNA (mRNA)-editing, enzyme-catalytic, polypeptide-like 3 family and plays an important role in the innate immune response, and a higher enrichment score of APOBEC3A mutagenesis correlates with favorable prognosis, immune activation, and potential immunotherapy response in bladder cancer (Shi et al., 2022). The glycoprotein CD2 is a costimulatory receptor expressed mainly on T and NK cells, and plays an important role in the formation and organization of the immunological synapse. The interaction of CD58 and CD2 is required for anti-tumor immunity and is predictive of the immunotherapy response; defects in this axis promote immune evasion through diminished T cell activation and impaired intra-tumoral T cell infiltration and proliferation, and concurrently increase PD-L1 protein stabilization (Ho et al., 2023). A previous study also revealed that CXCL9 was significantly upregulated following dual PD-1/CTLA4 blockade (House et al., 2020). Our research demonstrated that those immune-related genes are associated with the survival of CRC, and are potential biomarkers of immunotherapy.

Peripheral blood-derived biomarkers have the advantage of being non-invasive and easily accessible compared with tissue biopsy (Zhou et al., 2022). To further validate and expand the implementation of MCICRS, we identified a higher MCICRS in monocytes while a lower MCICRS in lymphocytes at single-cell resolution. Additionally, in a house cohort of MSS CRC patients treated with immunotherapy, we verified that patients with higher lymphocyte and lower monocyte counts exhibited a better treatment response and preferable PFS, suggesting that the content of both immune cell types could reflect the MCICRS level and be favorable surrogates for routine immunotherapy biomarkers like PD-L1 expression. The above results transferred the MCICRS model into peripheral blood lymphocyte and monocyte abundance, making it easily available in clinical practice to expand its usage, and validated it as a potential candidate biomarker for selecting suitable MSS CRC patients who could benefit from immunotherapy.

There are several limitations of this study that need to be acknowledged. First, it was a retrospective study with a small sample size. A prospective study with large populations is necessary to further evaluate the efficacy and safety of the anti-PD-1 antibody and VEGFRi combination in MSS CRC. Second, owing to the restriction of tissue samples, we could not assess the MCICRS for each MSS CRC patient. Our future prospective study would collect tissue samples and validate the predictive value of MCICRS. Finally, the detailed functions of the constituent genes of MCICRS in CRC need to be revealed in further in vivo and in vitro experiments.

5 Conclusions

In conclusion, anti-PD-1 antibody combined with VEGFRi presented improved efficacy and survival in MSS CRC patients with manageable toxicity. Based on analyses of scRNA-seq, spatial transcriptomic data, and machine learning algorithms, we developed a stable and powerful signature for accurately assessing the efficacy and prognosis of MSS CRC patients receiving immunotherapy, which could be a promising tool for optimizing decision-making for individual patients.

Data availability statement

The in-house datasets used and analyzed in this study are available from the corresponding authors upon reasonable request. Other public data analyzed in this study are available from TCGA (https://portal.gdc.cancer.gov) and GEO (https://www.ncbi.nlm.nih.gov/geo).

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 81972012). We thank to KS official accounts team (Zhejiang University, Hangzhou, China) for their help in data analysis and visual codes. We also thank Dr. Siyuan HUANG (Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China) for his help in data analysis and visual codes.

Author contributions

Conception, design, and study supervision: Jun ZHANG, Xian WANG, and Haizhou LOU; Development of methodology and interpretation of results: Shijin YUAN, Yan XIA, and Guangwei DAI; Acquisition of data: Shijin YUAN, Yan XIA, Guangwei DAI, Shun RAO, Rongrong HU, Yuzhen GAO, Qing QIU, Chenghao WU, Sai QIAO, and Yinghua XU; Analysis: Shijin YUAN, Yan XIA, Guangwei DAI, Shun RAO, and Rongrong HU; Writing, revision, and review of the manuscript: Shijin YUAN, Yan XIA, Guangwei DAI, Shun RAO, Rongrong HU, Xinyou XIE, Haizhou LOU, Xian WANG, and Jun ZHANG. All authors have read and approved the final manuscript, and therefore, have full access to all the data in the study and take responsibility for the integrity and security of the data.

Compliance with ethics guidelines

Shijin YUAN, Yan XIA, Guangwei DAI, Shun RAO, Rongrong HU, Yuzhen GAO, Qing QIU, Chenghao WU, Sai QIAO, Yinghua XU, Xinyou XIE, Haizhou LOU, Xian WANG, and Jun ZHANG declare that they have no conflicts of interest.

This study was approved by the Ethics Committee of Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China (Certificate No. 20190211-55), and the project was carried out in accordance with the Helsinki Declaration of 1975, as revised in 2013.

References

André T, Shiu KK, Kim TW, et al., 2020. Pembrolizumab in microsatellite-instability-high advanced colorectal cancer. *N Engl J Med*, 383(23):2207-2218. https://doi.org/10.1056/NEJMoa2017699

Butler A, Hoffman P, Smibert P, et al., 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol*, 36(5):411-420. https://doi.org/10.1038/nbt.4096

Colaprico A, Silva TC, Olsen C, et al., 2016. *TCGAbiolinks*: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*, 44(8):e71. https://doi.org/10.1093/nar/gkv1507

Cousin S, Cantarel C, Guegan JP, et al., 2021. Regorafenib-avelumab combination in patients with microsatellite stable colorectal cancer (REGOMUNE): a single-arm, open-label, phase II trial. *Clin Cancer Res*, 27(8):2139-2147. https://doi.org/10.1158/1078-0432.CCR-20-3416

Doroshow DB, Bhalla S, Beasley MB, et al., 2021. PD-L1 as a biomarker of response to immune-checkpoint inhibitors. *Nat Rev Clin Oncol*, 18(6):345-362. https://doi.org/10.1038/s41571-021-00473-5

Engstrand J, Nilsson H, Strömberg C, et al., 2018. Colorectal cancer liver metastases − a population-based study on incidence, management and survival. *BMC Cancer*, 18:78. https://doi.org/10.1186/s12885-017-3925-x

Finn RS, Qin SK, Ikeda M, et al., 2020. Atezolizumab plus bevacizumab in unresectable hepatocellular carcinoma. *N Engl J Med*, 382(20):1894-1905. https://doi.org/10.1056/NEJMoa1915745

Fukuoka S, Hara H, Takahashi N, et al., 2020. Regorafenib plus nivolumab in patients with advanced gastric or colorectal cancer: an open-label, dose-escalation, and dose-expansion phase Ib trial (REGONIVO, EPOC1603). *J Clin Oncol*, 38(18):2053-2061. https://doi.org/10.1200/JCO.19.03296

Grothey A, van Cutsem E, Sobrero A, et al., 2013. Regorafenib monotherapy for previously treated metastatic colorectal cancer (CORRECT): an international, multicentre, randomised, placebo-controlled, phase 3 trial. *Lancet*, 381(9863):303-312. https://doi.org/10.1016/S0140-6736(12)61900-X

Ho P, Melms JC, Rogava M, et al., 2023. The CD58-CD2 axis is co-regulated with PD-L1 via CMTM6 and shapes antitumor immunity. *Cancer Cell*, 41(7):1207-1221.e12. https://doi.org/10.1016/j.ccell.2023.05.014

Hodi FS, Lawrence D, Lezcano C, et al., 2014. Bevacizumab plus ipilimumab in patients with metastatic melanoma. *Cancer Immunol Res*, 2(7):632-642. https://doi.org/10.1158/2326-6066.CIR-14-0053

House IG, Savas P, Lai JY, et al., 2020. Macrophage-derived CXCL9 and CXCL10 are required for antitumor immune responses following immune checkpoint blockade. *Clin Cancer Res*, 26(2):487-504. https://doi.org/10.1158/1078-0432.CCR-19-1868

Jin SQ, Guerrero-Juarez CF, Zhang LH, et al., 2021. Inference and analysis of cell-cell communication using CellChat. *Nat Commun*, 12:1088. https://doi.org/10.1038/s41467-021-21246-9

Karapetis CS, Khambata-Ford S, Jonker DJ, et al., 2008. *K-ras* mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med*, 359(17):1757-1765. https://doi.org/10.1056/NEJMoa0804385

Kim RD, Kovari BP, Martinez M, et al., 2022. A phase I/Ib study of regorafenib and nivolumab in mismatch repair proficient advanced refractory colorectal cancer. *Eur J Cancer*, 169:93-102. https://doi.org/10.1016/j.ejca.2022.03.026

Langfelder P, Horvath S, 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf*, 9:559. https://doi.org/10.1186/1471-2105-9-559

Le DT, Uram JN, Wang H, et al., 2015. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med*, 372(26):2509-2520. https://doi.org/10.1056/NEJMoa1500596

Le DT, Durham JN, Smith KN, et al., 2017. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science*, 357(6349):409-413. https://doi.org/10.1126/science.aan6733

Le DT, Kim TW, van Cutsem E, et al., 2020. Phase II open-label study of pembrolizumab in treatment-refractory, microsatellite instability-high/mismatch repair-deficient metastatic colorectal cancer: KEYNOTE-164. *J Clin Oncol*, 38(1):11-19. https://doi.org/10.1200/JCO.19.02107

Lee HO, Hong Y, Etlioglu HE, et al., 2020. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nat Genet*, 52(6):594-603. https://doi.org/10.1038/s41588-020-0636-z

Li J, Qin SK, Xu RH, et al., 2018. Effect of fruquintinib vs placebo on overall survival in patients with previously treated metastatic colorectal cancer: the FRESCO randomized clinical trial. *JAMA*, 319(24):2486-2496. https://doi.org/10.1001/jama.2018.7855

Li YS, Shen ZY, Chai Z, et al., 2023. Targeting MS4A4A on tumour-associated macrophages restores CD8$^+$ T-cell-mediated antitumour immunity. *Gut*, 72(12):2307-2320. https://doi.org/10.1136/gutjnl-2022-329147

Liebermeister W, Noor E, Flamholz A, et al., 2014. Visual account of protein investment in cellular functions. *Proc Natl Acad Sci USA*, 111(23):8488-8493. https://doi.org/10.1073/pnas.1314810111

Mariathasan S, Turley SJ, Nickles D, et al., 2018. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature*, 554(7693):544-548. https://doi.org/10.1038/nature25501

McDermott DF, Huseni MA, Atkins MB, et al., 2018. Clinical activity and molecular correlates of response to atezolizumab alone or in combination with bevacizumab versus sunitinib in renal cell carcinoma. *Nat Med*, 24(6):749-757. https://doi.org/10.1038/s41591-018-0053-3

Mints M, Landin D, Näsman A, et al., 2021. Tumour inflammation signature and expression of S100A12 and HLA class I improve survival in HPV-negative hypopharyngeal cancer. *Sci Rep*, 11:1782. https://doi.org/10.1038/s41598-020-80226-z

Modest DP, Pant S, Sartore-Bianchi A, 2019. Treatment sequencing in metastatic colorectal cancer. *Eur J Cancer*, 109:70-83. https://doi.org/10.1016/j.ejca.2018.12.019

Overman MJ, McDermott R, Leach JL, et al., 2017. Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. *Lancet Oncol*, 18(9):1182-1191. https://doi.org/10.1016/S1470-2045(17)30422-9

Overman MJ, Lonardi S, Wong KYM, et al., 2018. Durable clinical benefit with nivolumab plus ipilimumab in DNA mismatch repair-deficient/microsatellite instability-high metastatic colorectal cancer. *J Clin Oncol*, 36(8):773-779. https://doi.org/10.1200/JCO.2017.76.9901

Parikh AR, Szabolcs A, Allen JN, et al., 2021. Radiation therapy enhances immunotherapy response in microsatellite stable colorectal and pancreatic adenocarcinoma in a phase II trial. *Nat Cancer*, 2(11):1124-1135. https://doi.org/10.1038/s43018-021-00269-7

Qiu XJ, Mao Q, Tang Y, et al., 2017. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods*, 14(10):979-982. https://doi.org/10.1038/nmeth.4402

Rahma OE, Hodi FS, 2019. The intersection between tumor angiogenesis and immune suppression. *Clin Cancer Res*, 25(18):5449-5457. https://doi.org/10.1158/1078-0432.CCR-18-1543

Shi R, Wang X, Wu Y, et al., 2022. APOBEC-mediated mutagenesis is a favorable predictor of prognosis and immunotherapy for bladder cancer patients: evidence from pan-cancer analysis and multiple databases. *Theranostics*, 12(9): 4181-4199. https://doi.org/10.7150/thno.73235

Socinski MA, Jotte RM, Cappuzzo F, et al., 2018. Atezolizumab for first-line treatment of metastatic nonsquamous NSCLC. *N Engl J Med*, 378(24):2288-2301. https://doi.org/10.1056/NEJMoa1716948

Sung H, Ferlay J, Siegel RL, et al., 2021. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, 71(3):209-249. https://doi.org/10.3322/caac.21660

Wu YC, Yang SX, Ma JQ, et al., 2022. Spatiotemporal immune landscape of colorectal cancer liver metastasis at single-cell level. *Cancer Discov*, 12(1):134-153. https://doi.org/10.1158/2159-8290.CD-21-0316

Zhang L, Yu X, Zheng LT, et al., 2018. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature*, 564(7735):268-272. https://doi.org/10.1038/s41586-018-0694-x

Zhang L, Li ZY, Skrzypczynska KM, et al., 2020. Single-cell analyses inform mechanisms of myeloid-targeted therapies in colon cancer. *Cell*, 181(2):442-459.e29. https://doi.org/10.1016/j.cell.2020.03.048

Zheng LT, Qin SS, Si W, et al., 2021. Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science*, 374(6574): abe6474. https://doi.org/10.1126/science.abe6474

Zheng RS, Zhang SW, Sun KX, et al., 2023. Cancer statistics in China, 2016. *Chin J Oncol*, 45(3):212-220 (in Chinese). https://doi.org/10.3760/cma.j.cn112152-20220922-00647

Zhou H, Zhu LY, Song J, et al., 2022. Liquid biopsy at the frontier of detection, prognosis and progression monitoring in colorectal cancer. *Mol Cancer*, 21:86. https://doi.org/10.1186/s12943-022-01556-2

Supplementary information

Tables S1–S11; Figs. S1–S9