Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrated analysis of single-cell RNA sequencing and bulk RNA sequencing reveals T-cell-associated prognostic risk models and the regulation of the tumour immune microenvironment in colorectal cancer

Xiangli Yin
Beijing University of Chinese Medicine
Yiyan Zhai
Beijing University of Chinese Medicine
Xiaodong Chen
Beijing University of Chinese Medicine
Jiangying Liu
Beijing University of Chinese Medicine
Haojia Wang
Beijing University of Chinese Medicine
Peiying Lu
Beijing University of Chinese Medicine
Junyi Li
Beijing University of Chinese Medicine
Muyassar Mamattohti
Beijing University of Chinese Medicine
Tianzhu Zhang
Beijing University of Chinese Medicine
Xiaojing Zhao
Beijing University of Chinese Medicine
Kailin Liu
Beijing University of Chinese Medicine
Hexuan Liu
Henan University of Chinese Medicine
Jiarui Wu
exogamy@163.com

Beijing University of Chinese Medicine

Research Article

Keywords:

Posted Date: August 31st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10717563/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Integrated analysis of single-cell RNA sequencing and bulk RNA sequencing reveals T-cell-associated prognostic risk models and the regulation of the tumour immune microenvironment in colorectal cancer

Xiangli Yin[1], Yiyan Zhai[1], Xiaodong Chen[1], Jiangying Liu[1], Haojia Wang[1], Peiying Lu[1], Junyi Li[1], Muyassar Mamattohti[1], Tianzhu Zhang[1], Xiaojing Zhao[1], Kailin Liu[1], Hexuan Liu[2] , Jiarui Wu[1*]

[1] School of Chinese Materia Medica, Beijing University of Chinese Medicine, Beijing, 102488, China
[2] School of Pharmacy, Henan University of Chinese Medicine, Zhengzhou, 450045, China
*****Corresponding author:
Jiarui Wu, Email: exogamy@163.com

ABSTRACT

Objective: Colorectal cancer (CRC), a highly prevalent gastrointestinal malignancy, has experienced a paradigm shift in therapeutic strategies driven by recent advances in immunotherapy. Nevertheless, the precise impact of T-cell subpopulation heterogeneity and functional evolution within the tumor microenvironment (TME) on clinical prognosis warrants further elucidation. This study aims to construct a robust T-cell-related prognostic model for CRC by integrating single-cell RNA sequencing (scRNA-seq) and bulk transcriptomic data, thereby evaluating its utility in guiding clinical decision-making.

Methods: We utilized scRNA-seq data to comprehensively analyze T-cell subsets and cellular heterogeneity in patients with CRC. Bulk transcriptomic profiles retrieved from public repositories, including The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases, were integrated to identify core T-cell signature genes for prognostic model development. Furthermore, the CIBERSORT algorithm was applied to evaluate immune infiltration landscapes across distinct risk strata and to explore their associations with immunotherapeutic responsiveness.

Results: A total of 77,167 high-quality single-cell transcriptomes were successfully acquired, yielding 15,450 T lymphocytes that were further classified into central memory T cells, regulatory T cells (Tregs), effector memory T cells, effector T cells, and helper T cells. Integrative bioinformatics analyses identified six core prognostic genes (HSPA1A, TIMP1, HSPA8, ZG16, LGALS4, and CXCL13), based on which a stable prognostic risk model was established. This model effectively stratified CRC patients into high- and low-risk groups, exhibiting superior predictive accuracy across both training and external validation cohorts. Moreover, the high-risk group demonstrated pronounced stromal enrichment and an immunosuppressive TME phenotype characterized by elevated Treg infiltration.

Conclusion: This study successfully identifies and validates a novel prognostic signature capable of accurately predicting overall survival in patients with CRC while reflecting the immunological status of the TME, thereby offering a robust scientific framework for personalized therapeutic strategies and immune stratification in clinical practice.

Introduction

Colorectal cancer (CRC) represents a major global health challenge. According to the latest global cancer statistics for 2024 based on estimates from the International Agency for Research on Cancer's GLOBOCAN, CRC ranks as the third most common malignant tumor worldwide, accounting for 9.9% of all cancer cases, and stands as the second leading cause of cancer mortality globally, responsible for 9.4% of all cancer-related deaths [1]. Established risk factors contributing to the development of CRC include alcohol consumption, high intake of red and processed meat, obesity, physical inactivity, and smoking [2]. Current therapeutic modalities for CRC primarily encompass surgical resection, radiotherapy, chemotherapy, and targeted therapy. Nevertheless, the overall clinical prognosis for patients with advanced-stage disease remains dismal. While immune checkpoint blockade (ICB) therapy has substantially revolutionized the treatment landscape for malignancies such as melanoma and non-small cell lung cancer, its efficacy in CRC is largely restricted to patients harboring tumors with microsatellite instability-high (MSI-H) or mismatch repair deficiency (dMMR) [3]. Because checkpoint inhibitor-based immunotherapy relies heavily on T-cell infiltration into the tumor microenvironment (TME)—making T cells the paramount effector cells within the TME [4]—comprehensive exploration of T-cell heterogeneity, functional characteristics, and prognostic associations is vital for advancing novel immunotherapeutic strategies and identifying reliable prognostic biomarkers.

Although traditional bulk RNA sequencing (bulk RNA-seq) captures average gene expression profiles at the whole-tissue level, it inherently obscures intercellular heterogeneity and exhibits limited sensitivity for detecting low-abundance transcripts. Conversely, single-cell RNA sequencing (scRNA-seq) enables high-resolution transcriptomic profiling at the single-cell level, facilitating the unmasking of cellular diversity, the identification of rare cell populations, the tracking of developmental trajectories, and the dissection of complex microenvironments [5]. Despite its exceptional power in decoding immune responses and tumor ecosystems, scRNA-seq is frequently constrained by small sample sizes and high costs. Translating single-cell-derived molecular signatures into robust, clinically translatable prognostic tools therefore necessitates seamless integration with bulk transcriptomic cohorts. Mapping cell-specific signatures identified at single-resolution onto large-scale batch RNA-seq datasets not only validates the clinical relevance of cellular subpopulations but also bridges deep biological mechanisms with practical clinical utility for prognostic modeling.

Driven by this scientific premise, the present study aims to systematically identify core gene modules closely linked to T-cell function via the integration of single-cell and multi-center bulk RNA-seq data. By constructing a robust T-cell-related prognostic risk model and performing an in-depth evaluation of TME dynamics across distinct risk strata, this work endeavors to furnish a novel theoretical framework for clinical prognostic assessment and precision immunotherapy in CRC **(Figure 1)**.

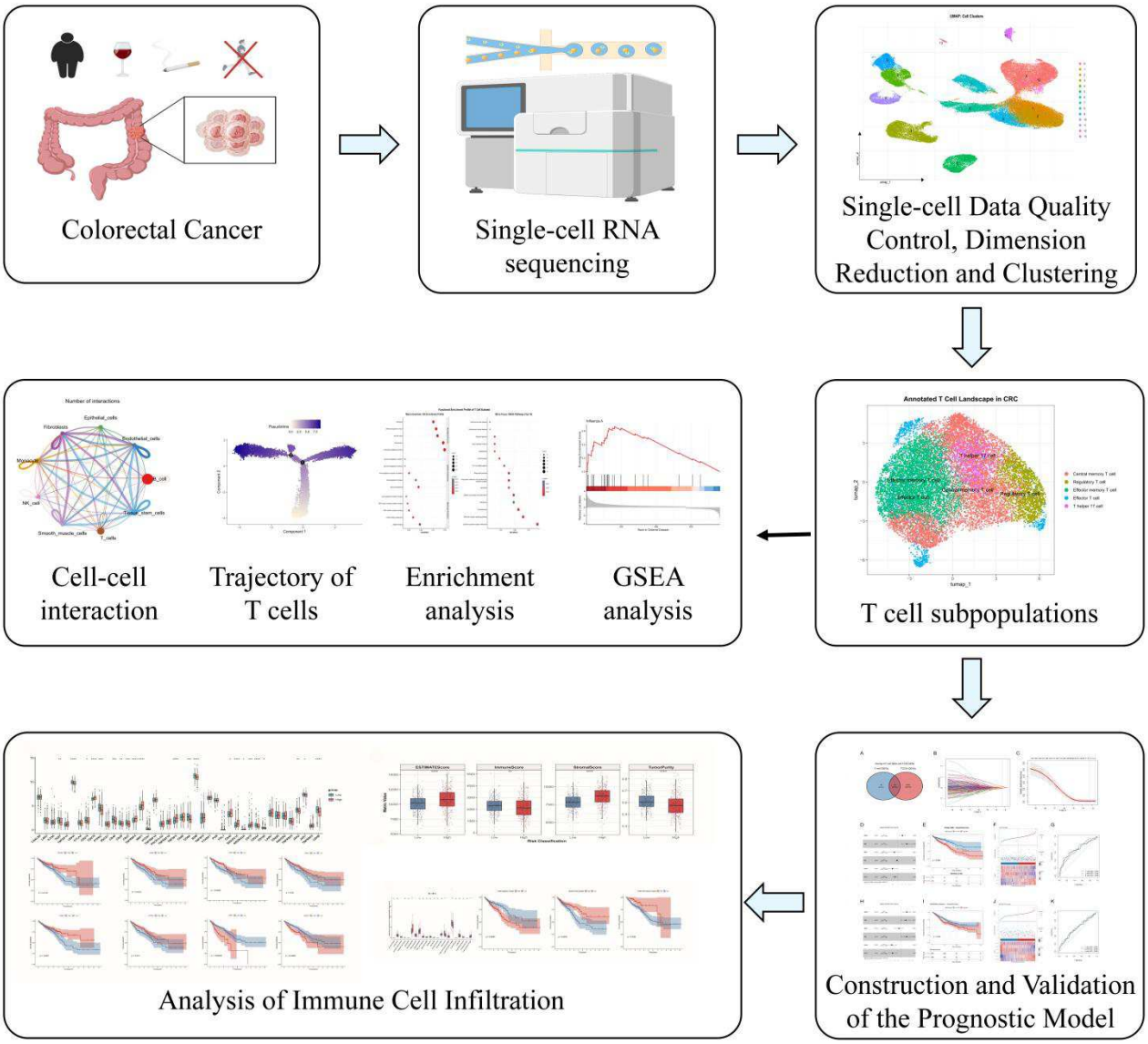


Figure 1. Summary diagram of this study.

1. Materials and Methods

1.1 Data Acquisition and Pre-processing

The scRNA-seq dataset for CRC, designated as GSE315534, was acquired from the Gene Expression Omnibus (GEO) repository (http://www.ncbi.nlm.nih.gov/geo/). This dataset encompasses a total of 15 samples derived from primary tumor tissues and matched adjacent non-tumor tissues of six CRC patients, alongside liver metastasis samples from three additional patients. Because this study focuses primarily on the heterogeneity of T cells within the primary tumor microenvironment and its implications for clinical prognosis, six primary tumor samples and six paired adjacent non-tumor samples were selected for downstream single-cell transcriptomic analysis.

Bulk RNA sequencing cohorts were retrieved from The Cancer Genome Atlas (TCGA). Specifically, the TCGA-COAD and TCGA-READ RNA-seq datasets were downloaded from the Genomic Data Commons (GDC) portal utilizing the 'TCGAbiolinks' package in R (version 4.4.3). STAR-Counts data served as the raw expression matrix, with unstranded read count data extracted for subsequent computational workflows. Sample classifications were determined based on the 14th and 15th characters of the TCGA sample barcodes,

culminating in 647 primary tumor tissue samples and 51 adjacent normal tissue samples. Comprehensive clinical annotations, including overall survival (OS) time and survival status, were retrieved via the 'GDCquery_clinic' function. Tumor samples were subsequently deduplicated per patient ID, and gene expression profiles were integrated with corresponding clinical metadata. Ultimately, a cohort of 591 CRC patients with complete survival information was established for prognostic model construction and associated analyses.

To rigorously evaluate the stability and generalizability of the developed prognostic signature, the GSE39582 dataset from the GEO database was incorporated as an external independent validation cohort. This cohort comprises samples from 585 CRC patients profiled on the GPL570 [HG-U133 Plus 2.0] Affymetrix Human Genome U133 Plus 2.0 microarray platform. Following rigorous curation and screening of clinical parameters, a total of 573 CRC tumor samples with complete clinical records were ultimately included for external validation analyses.

1.2 Single-cell Data Quality Control, Dimension Reduction and Clustering

The gene expression matrices for each single-cell sample were imported utilizing the Read10X function in R (version 4.4.3), and individual Seurat objects were constructed using the CreateSeuratObject function from the Seurat package. Cells were categorized and annotated based on sample origin metadata, wherein primary tumor tissues were designated as the "Tumor" group and matched adjacent non-tumor tissues as the "Normal" group. This grouping framework was preserved for downstream analyses, including cellular composition comparisons, differential expression evaluations, and data visualization. Subsequently, rigorous quality control (QC) was implemented to filter the single-cell transcriptomic data. Cells were initially screened based on the number of detected genes (nFeature_RNA), retaining those with feature counts ranging between 300 and 6,500. Additionally, each gene was required to be expressed in at least three cells to minimize technical artifacts arising from low-quality cells and poorly expressed genes. Furthermore, the PercentageFeatureSet function was applied to compute the proportional expression levels of mitochondrial and hemoglobin genes per cell. Cells exhibiting a mitochondrial gene proportion exceeding 10% or a hemoglobin gene proportion greater than 2% were excluded to eliminate data bias attributable to cellular damage, apoptosis, or erythrocyte contamination. Following quality control, the FindVariableFeatures function was employed to identify the top 3,000 highly variable genes (HVGs), upon which principal component analysis (PCA) was executed for dimensionality reduction. A cell-to-cell neighborhood matrix was subsequently constructed from the PCA results, and a shared nearest neighbor graph was generated using the FindNeighbors function. Graph-based cell clustering analysis was then performed via the FindClusters function with the resolution parameter set to 0.4. Finally, the Uniform Manifold Approximation and Projection (UMAP) algorithm was utilized to project and visualize the dimension-reduced cell populations in a two-dimensional space, elucidating the global distribution patterns of distinct cellular subsets.

1.3 Cell Type Identification

Based on the single-cell clustering results, the SingleR software package was utilized for automated cell type annotation [6]. Leveraging the Human Primary Cell Atlas as a reference database, predictions were generated for distinct cell clusters at the cluster level, and the annotation outcomes were visually validated using UMAP dimensionality reduction. Subsequently, these automated annotations were manually refined against well-established, classic cell type marker genes reported in previous literature to enhance the accuracy of cellular identification.

Following successful global cell typing, T-cell subpopulations were isolated based on the celltype metadata label. The top 2,000 highly variable genes were re-normalized and selected while filtering out

interfering genes. After ScaleData regression, PCA dimensionality reduction, and Harmony batch effect correction, a clustering resolution of 1.2 was applied. Each resulting subpopulation was subsequently annotated using classic T-cell marker genes established in literature—including FOXP3, RTKN2, CD8B, GZMK, IL7R, CCR7, IL17A, CTSH, and GNLY—in conjunction with dot plots and feature expression maps.

1.4 Identification of differentially expressed genes in T cells and functional enrichment analysis

To elucidate the molecular characteristics and functional evolution of T cells within the colorectal cancer tumor microenvironment, downstream computational analyses were performed using R (utilizing packages such as 'Seurat' and 'clusterProfiler'). First, the 'FindAllMarkers' function was employed to identify differentially expressed genes (DEGs) across T-cell subsets. To balance gene detection sensitivity with biological specificity in high-throughput single-cell data, parameters were configured requiring genes to be expressed in at least 25% of cells within a given subpopulation ('min.pct = 0.25'), alongside significance thresholds of a corrected $P < 0.05$ and $|\log_2FC| > 0.5$. The top 10 characteristic genes for each subpopulation were extracted based on absolute $\log_2FC$ values and visualized using 'DotPlot' and 'DoHeatmap'.

Subsequently, the 'FindMarkers' function (utilizing the Wilcoxon rank-sum test) was applied to evaluate gene expression differences between the tumor group (Tumor) and the normal group (Normal). Employing identical screening thresholds, volcano plots were generated via the 'ggplot2' and 'ggrepel' packages to depict the significantly differentially expressed genes induced by the TME. Furthermore, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were conducted on these DEGs ($P < 0.05$, $q < 0.05$). In parallel, Gene Set Enrichment Analysis (GSEA) was performed on gene lists ranked by $\log_2FC$ to uncover the functional regulatory mechanisms governing T cells during tumor progression.

1.5 Cell-cell Communication Analysis

To investigate ligand–receptor interactions between T cells and other cell types, we used the 'CellChat' package to construct intercellular communication networks [7]. Using single-cell transcriptomic expression matrices and cell type annotations as inputs, we performed inference using the human ligand–receptor database ('CellChatDB.human'). First, we identified genes and ligand–receptor pairs that were overexpressed in each cell type and projected them onto a protein interaction network to enhance prediction accuracy. Subsequently, we calculated the probability of intercellular communication using the truncated mean method ('trim = 0.1') and filtered out interactions detected in fewer than 10 cells. We then aggregated the ligand–receptor pairs to the signaling pathway level to obtain the probability of communication between pathways. A bubble plot is utilized to display the interaction probabilities for all ligand–receptor pairs, with T cells acting as senders and other cell types as receivers, whilst retaining a detailed table of interactions for each pathway. Furthermore, network diagrams and role heatmaps are employed to visualize the global communication patterns of different cell types acting as senders, receivers, mediators, and influencers.

1.6 Pseudotime Analysis

Pseudotime-series analysis of T-cell subsets was performed using the Monocle2 package to explore T-cell differentiation trajectories. A CellDataSet was constructed based on UMI counts; the expression data were fitted using a negative binomial distribution, and the size factor and gene dispersion were estimated. Subsequently, genes for trajectory construction were selected based on gene expression characteristics; the selection criteria were set as highly variable genes with an average expression level ≥ 0.1 and an empirical dispersion higher than the model-fitted dispersion. Dimension reduction was then performed using the DDRTree algorithm; the root node was defined based on early states to infer differentiation states and

pseudotime. Trajectory plots were colored according to clustering, cell type, grouping, and pseudotime to visualize the differentiation process. Pseudotime-correlated differentially expressed genes ($q < 0.01$) were identified using differentialGeneTest; the dynamic expression of the top 50 genes was visualized as a heatmap, and nine T-cell marker genes (CD8B, GZMK, IL7R, CCR7, FOXP3, RTKN2, IL17A, CTSH, GNLY) were selected to plot expression curves, comprehensively revealing the differentiation trajectories.

1.7 Construction and Validation of the Prognostic Model

To pinpoint genes that bridge T-cell function and clinical prognosis, we integrated T-cell-specific transcriptomic signatures with bulk transcriptomic data. First, T-cell signature genes were defined as those exhibiting significant upregulation ($|\log2FC| > 0.5$, adjusted $P < 0.05$) across T-cell subpopulations using the FindAllMarkers function in Seurat. These signature genes were subsequently intersected with the differentially expressed genes (DEGs) identified between tumor and normal tissues in the TCGA-CRC cohort. This intersection strategy ensured that candidate prognostic genes were both highly expressed in T-cell subsets and significantly dysregulated within the CRC tumor microenvironment, thereby capturing the functional interplay between T-cell reprogramming and tumorigenesis. Building upon this candidate gene set, this study constructed a robust T-cell-related prognostic risk model. Candidate genes significantly associated with overall survival (OS) were initially screened using univariate Cox proportional hazards regression. To mitigate multicollinearity and minimize the risk of overfitting, Least Absolute Shrinkage and Selection Operator (LASSO)–Cox regularized regression was implemented for dimensionality reduction, determining the optimal penalty parameter (λ_{min}) via 10-fold cross-validation. Based on the core feature combination selected by LASSO, a multivariate Cox regression model was subsequently established to formulate the risk score algorithm. Patients in the TCGA training cohort were stratified into high- and low-risk groups according to the median risk score, and Kaplan–Meier survival curves were utilized to compare OS differences between the strata. Time-dependent receiver operating characteristic (ROC) curves and their corresponding areas under the curve (AUC) were calculated using the timeROC package to evaluate the 1-, 3-, and 5-year prognostic performance of the model. Finally, to rigorously test the generalizability and robustness of the signature, the independent external cohort GSE39582 was evaluated under identical criteria.

1.8 Analysis of Immune Cell Infiltration

The ESTIMATE algorithm was employed to calculate immune, stromal, and composite scores using sets of immune and stromal marker genes, and tumor purity was estimated via linear normalization [8]. CIBERSORT is a deconvolution method that estimates the distribution of immune cells based on the gene expression profile of the cancer [9]. In this study, CIBERSORT deconvolution was used to calculate the infiltration proportions of 22 types of immune cells and to compare differences between high- and low-risk groups [10]. Based on risk stratification, differential expression analysis was performed on immune checkpoint genes, and expression levels between groups were compared using the Wilcoxon test. The 'surv_cutpoint' function was employed to determine the optimal cut-off value, dividing gene expression into high- and low-expression groups; survival curves were plotted using the Kaplan–Meier method, and the log-rank test was used to assess prognostic significance. Furthermore, the correlation between risk scores and immune checkpoint expression was investigated. All statistical tests were two-sided, with $p < 0.05$ considered statistically significant.

1.9 Statistical Analysis

Inter-group differences and differentially expressed genes were evaluated utilizing the Wilcoxon rank-sum test. Least absolute shrinkage and selection operator (LASSO) regression coupled with multivariate Cox proportional hazards regression was implemented to establish the prognostic risk model.

Overall survival distributions were assessed via Kaplan–Meier survival estimators, and statistical significance between strata was determined using the log-rank test. The predictive efficacy of the signature was quantified using the area under the time-dependent receiver operating characteristic curve. Two-sided p-values below 0.05 were deemed statistically significant, with method-specific thresholds explicitly defined in respective analytical steps. All computational workflows and statistical procedures were executed in the R environment (version 4.4.3).

2.Results

2.1 Cellular Composition of CRC

This study analyzed a total of 12 single-cell samples from colorectal cancer, comprising six primary tumor samples and six matched adjacent non-tumor (peri-tumor) samples (**Figure 2A**). **Figure 2B** presents a UMAP visualization of the 12 samples categorized into Normal and Tumor groups. Following rigorous quality control and dimensionality reduction clustering of the single-cell RNA sequencing data, a total of 77,167 cells were successfully classified into 14 distinct clusters **(Figure 2C**). The expression profiles of key marker genes for each cell type were illustrated using a dot plot (**Figure 2D**). Subsequently, leveraging automated cell-type annotation via the 'SingleR' package combined with manual refinement, these subclusters were assigned to nine major cell types: T cells (clusters 0 and 12), B cells (clusters 1, 2, 6, and 8), monocytes (cluster 3), tissue stem cells (clusters 4 and 10), endothelial cells (cluster 5), secretory cells (cluster 7), fibroblasts (cluster 9), natural killer (NK) cells (cluster 11), and smooth muscle cells (cluster 13) (**Figure 2E**). Furthermore, **Figure 2F** delineates the proportional distribution of these cell types across individual samples.

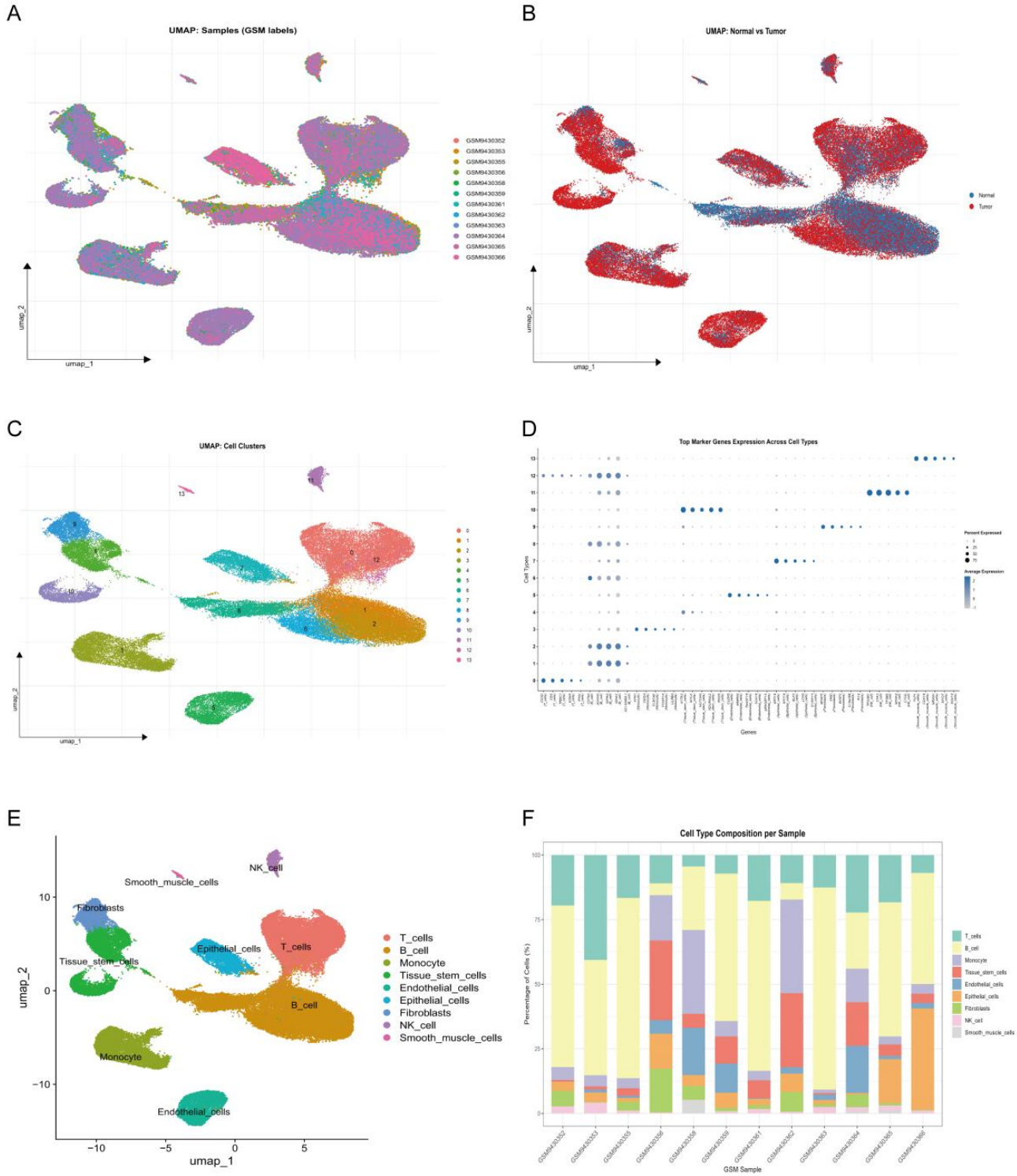


Figure 2. Integration and clustering analysis of CRC single-cell data (GSE315534). (A) Visualization of 12 CRC samples using the UMAP algorithm. (B) Visualization of 12 CRC samples categorized into Normal and Tumor groups using the UMAP algorithm. (C) UMAP plot showing the partitioning of the GSE315534 dataset into 14 distinct cell clusters. (D) A dot plot displaying the expression levels of characteristic marker genes across each cell subtype. (E) Identification of nine major cell types based on canonical marker genes. (F) Bar plot illustrating the proportional distribution of each cell type across individual samples.

2.2 T-cell Subpopulation Identification and Enrichment Analysis

Two of these clusters were designated as T cells and comprised a substantial number of cells. Consequently, 15,450 T cells were extracted for downstream re-clustering, yielding 14 distinct subclusters

(**Figure 3A**), which consisted of 4,530 T cells in the normal group and 10,920 T cells in the tumor group (**Figure 3B**). Based on annotation using canonical marker genes established in literature, the identified T-cell subsets encompassed 6,317 central memory T cells (clusters 2, 3, 4, 5, and 12), 1,836 regulatory T cells (clusters 8 and 9), 5,326 effector memory T cells (clusters 0, 1, 7, and 13), 942 effector T cells (clusters 10 and 11), and 1,029 T helper 17 cells (cluster 6) (**Figure 3C**). **Figures 3D and 3E** illustrate the characteristic marker genes for each T-cell subtype, including IL7R and CCR7 for central memory T cells, FOXP3 and RTKN2 for Tregs, CD8B and GZMK for effector memory T cells, GNLY for effector T cells, and IL17A and CTSH for T helper 17 cells. **Figure 3F** delineates the proportional distribution of each T-cell subset across the normal and tumor groups. Differential expression analysis of T cells between the normal and tumor cohorts identified 232 significantly differentially expressed genes, of which 169 were upregulated and 63 were downregulated in the tumor group, with the top five significant genes highlighted (**Figure 3G**). **Figure 3H** displays a heatmap of the top 10 DEGs across each T-cell subpopulation.

To elucidate the underlying biological functions and pathways associated with T-cell marker genes, GO, KEGG and GSEA were performed. GO enrichment analysis revealed that differentially expressed T-cell genes were significantly enriched in biological processes (BP) such as leukocyte-mediated immunity, viral processes, the viral life cycle, antiviral defense responses, and cell killing. Cellular component (CC) analysis indicated primary enrichment in blood granules and immunoglobulin complexes, whereas molecular function (MF) demonstrated significant enrichment in antigen binding, protein folding chaperones, MHC protein complex binding, and immunoglobulin receptor binding (**Figure 3I**). These findings suggest that differentially expressed T-cell genes are heavily involved in antiviral immune responses and antigen-presentation-related processes, potentially participating in immune regulation within the colorectal cancer microenvironment via MHC class II-mediated antigen recognition and immunoglobulin receptor signaling. KEGG enrichment analysis demonstrated significant enrichment in cytokine–cytokine receptor interactions, NOD-like receptor signaling pathways, and antigen processing and presentation pathways (**Figure 3I**), implying that T cells contribute to immunoreprogramming of the CRC tumor microenvironment by regulating inflammatory signaling and antigen presentation. Furthermore, GSEA results indicated that the Influenza A pathway was significantly upregulated in tumor-infiltrating T cells, whereas the Ribosome and Coronavirus disease pathways were significantly downregulated (**Figure 3J**), pointing toward specific antiviral and interferon response activity accompanied by the remodeling of protein translation mechanisms and targeted transcriptional regulation.

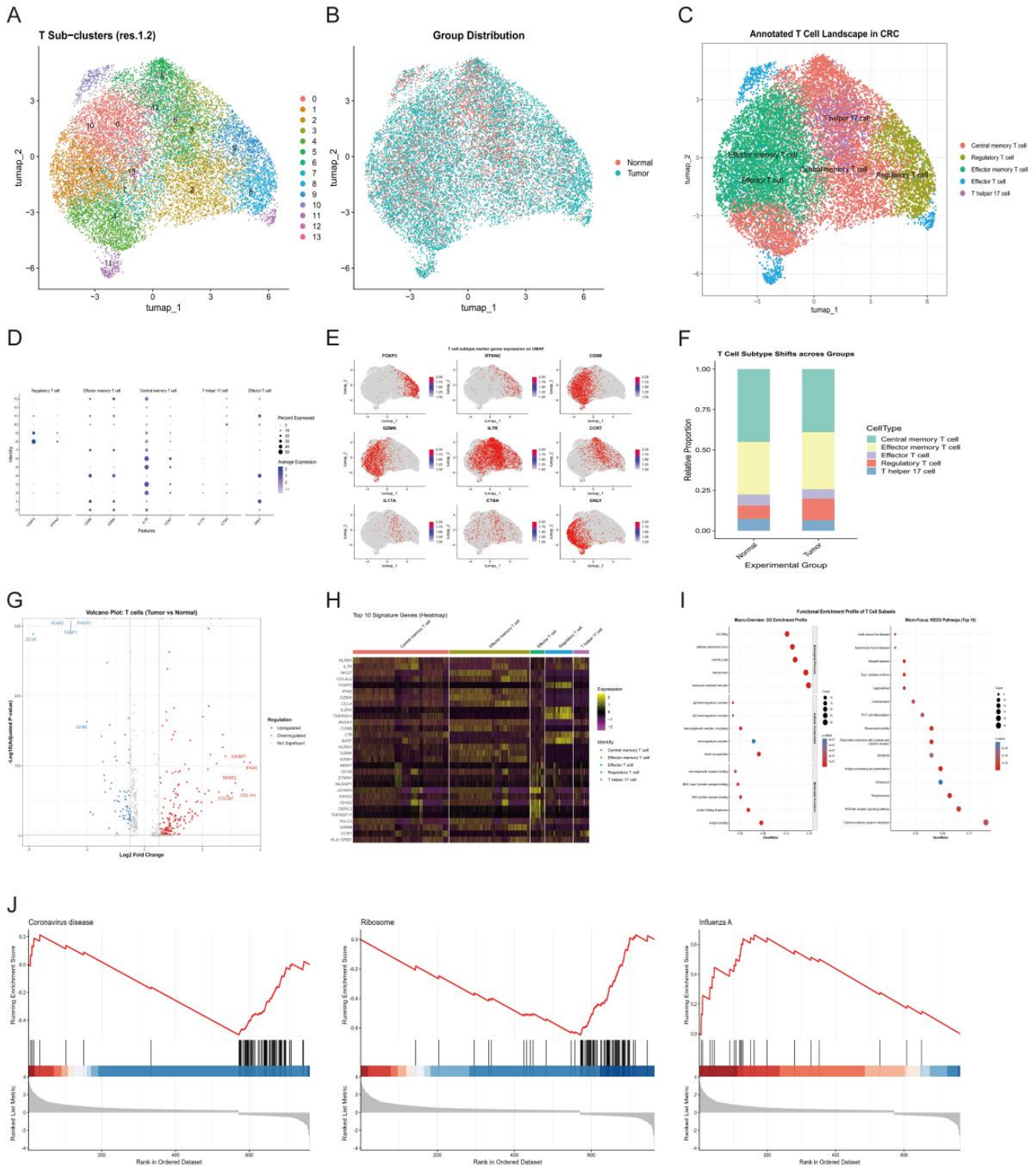


Figure 3. Identification and functional enrichment analysis of T-cell subsets in CRC single-cell data (GSE315534). (A) UMAP plot showing the re-clustering of T cells into 14 distinct clusters. (B) Distribution and visualization of T cells categorized into Normal and Tumor groups. (C) Identification of five major T-cell subsets based on canonical marker genes. (D) A dot plot displaying the expression levels of characteristic marker genes across each T-cell subtype. (E) UMAP feature plots showing the expression distribution of marker genes for each T-cell subtype. (F) Bar plot illustrating the proportional distribution of each T-cell subset in the Normal and Tumor groups. (G) Volcano plot showing differentially expressed genes in T cells between the Normal and Tumor groups. (H) Heatmap showing the top 10 DEGs for each T-cell subtype. (I) GO and KEGG enrichment analyses associated with T-cell marker genes. (J) GSEA associated with T-cell functional characteristics.

2.3 Analysis of Intercellular Communication

To investigate the mechanisms underlying interactions between T cells and other cell types, we utilized 'CellChat' to systematically predict cellular communication networks within the tumor microenvironment. Analysis of the number and intensity of intercellular interactions revealed frequent and high-intensity cell-to-cell communication among fibroblasts, monocytes, endothelial cells, and T cells (**Figures 4A–4B**). Signaling pathway-level analysis indicated that the CCL signaling pathway plays a pivotal role in reshaping intercellular communication within the TME (**Figure 4C).** Further evaluation of ligand–receptor (LR) pairs demonstrated that T cells predominantly signal to other cell types (particularly monocytes and endothelial cells) via interactions such as CCL5–CCR1, CCL5–CCR5, and CCL5–ACKR1 (**Figure 4D**). Gene expression profiling (**Figure 4E**) further confirmed that the ligand gene CCL5 is robustly and specifically expressed in T cells, whereas its receptor genes CCR1 and CCR5 are significantly expressed in monocytes, and ACKR1 is highly and specifically expressed in endothelial cells. Finally, network centrality analysis of the CCL signaling pathway (**Figure 4F**) indicated that T cells act as the primary signal senders and influencers, whereas endothelial cells and monocytes function predominantly as signal recipients.

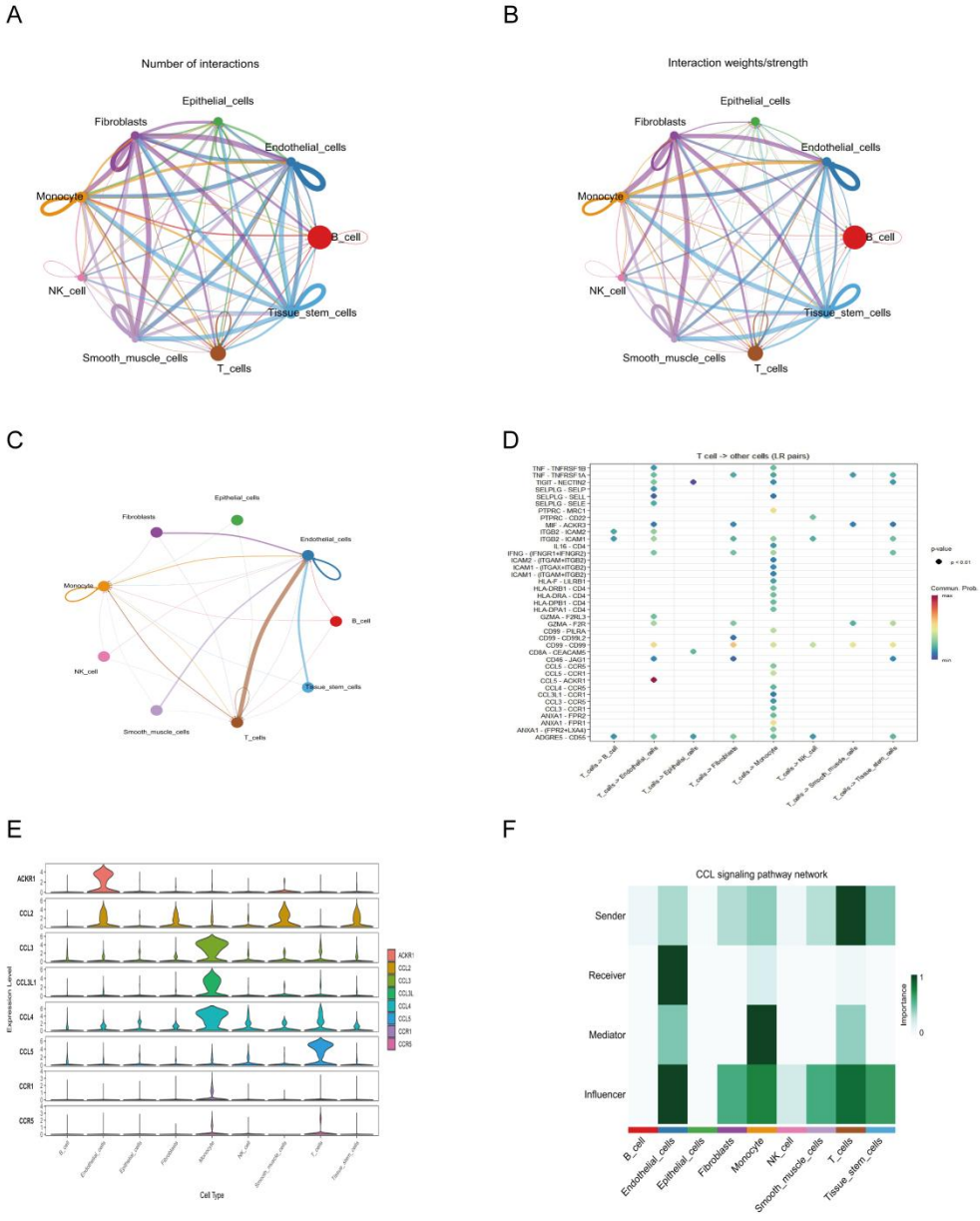


Figure 4. Analysis of T-cell-related intercellular communication. (A) Bar plot showing the number of interactions across the intercellular communication network. (B) Interaction weights and intensities in the intercellular communication network. (C) Intercellular communication interactions mediated by the CCL signaling pathway. (D) Dot plot illustrating the communication probabilities of ligand–receptor pairs between T cells and diverse cell types. (E) Violin plots showing the expression levels of genes associated with the CCL signaling pathway. (F) Heatmap displaying the inferred network roles (senders, receivers, mediators, and influencers) of major cell types within the CCL signaling pathway based on network centrality scores.

2.4 T-cell Trajectory Analysis

To further elucidate the differentiation trajectories and developmental stages of T-cell subsets, pseudotime trajectory analysis was performed using the 'Monocle2' algorithm. The results demonstrated that the T-cell differentiation trajectories exhibited a distinct tree-like branching structure, comprising a total of five separate differentiation states (**Figures 5A–5E**). Combined with the pseudotime initiation points (**Figure 5D**), lighter-colored regions at the base of the trajectory tree denoted the early stages of differentiation. Further evaluation of the proportional distribution of T-cell subsets across distinct states (**Figure 5G**) revealed that State 1 cells predominantly accumulated at the early phases of pseudotime development, subsequently branching out through critical decision nodes to differentiate into State 2 through State 5 cells residing on separate terminal branches. UMAP mapping (**Figure 5F**) further validated the consistency between 'Monocle2'-derived differentiation states and the spatial distribution of T-cell subsets. Furthermore, based on gene expression kinetics along the pseudotime axis, pseudotime-associated genes were clustered into three modules displaying significant dynamic shifts (**Figure 5H**), systematically illustrating the dynamic expression patterns of key T-cell marker genes across differentiation states and pseudotime (**Figure 5I**). Collectively, these findings unveil a continuous developmental trajectory wherein T cells originate from a central memory state, diverge through key branching nodes, and ultimately differentiate into effector and regulatory states, furnishing a developmental foundation for understanding the functional reprogramming and fate determination of T cells within the TME.

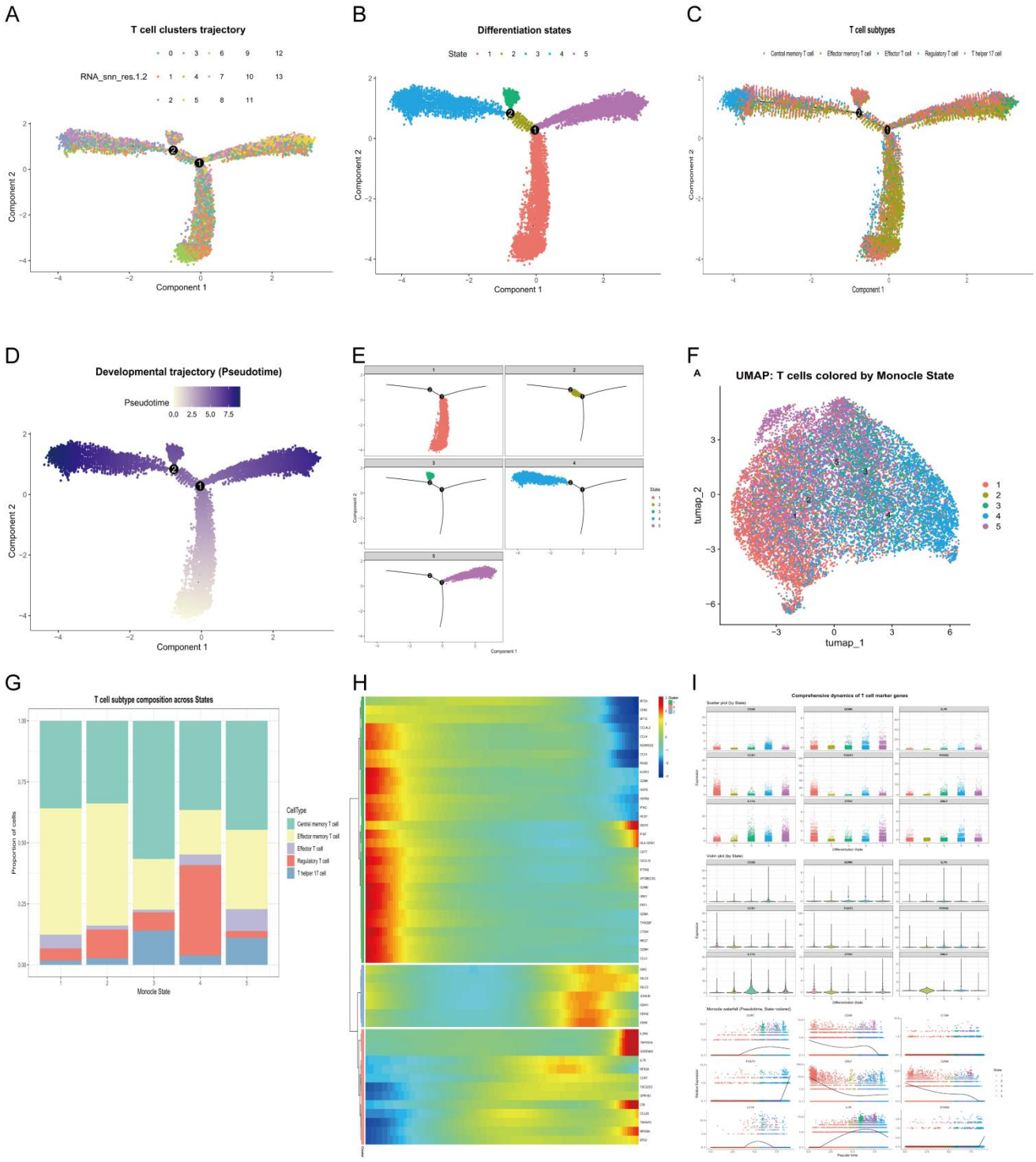


Figure 5. Pseudotime trajectory analysis of T cells in CRC. (A) Trajectory plots showing the distribution of distinct T-cell clusters. (B) Pseudotime trajectory illustrating different differentiation states of T cells. (C) Distribution of five major T-cell subtypes along the trajectory. (D) Pseudotime timeline illustrating the developmental progression of T-cell subsets. (E) Differentiation pathways of T-cell subsets projected onto the trajectory. (F) UMAP plot illustrating the spatial distribution of pseudotemporally resolved differentiation states. (G) Proportional composition of T-cell subsets across different pseudotime states. (H) Heatmap displaying differentially expressed genes along pseudotime. (I) Dynamic expression curves of nine canonical marker genes throughout the pseudotime continuum.

2.5 Construction and Validation of Prognostic Models

First, an intersection analysis was performed between DEGs from the The Cancer Genome Atlas colorectal cancer cohort and T-cell-associated genes from single-cell transcriptomic data, identifying a total of 140 overlapping genes (**Figure 6A**). Subsequently, candidate prognostic genes significantly associated with overall survival were screened using univariate Cox regression, and Least Absolute Shrinkage and Selection Operator –Cox regression was employed to mitigate the risk of multicollinearity; core modeling features were selected based on the optimal λ value (**Figures 6B–6C**). Further multivariate Cox regression analysis was conducted to construct a risk score model comprising six genes (HSPA1A, TIMP1, HSPA8, ZG16, LGALS4, and CXCL13), with its forest plot displayed in **Figure 6D**. The risk score was calculated for each patient using the model, and patients were categorized into high- and low-risk groups based on the median value. Kaplan–Meier survival analysis indicated that OS in the high-risk group was significantly inferior to that in the low-risk group ($P < 0.0001$, **Figure 6E**). A distribution heatmap of risk scores and gene expression profiles further revealed distinct expression patterns of the model genes across the risk strata (**Figure 6F**). Time-dependent receiver operating characteristic analysis showed that the model achieved areas under the curve of 0.672, 0.646, and 0.696 at 1, 3, and 5 years, respectively, in the TCGA training set, demonstrating robust predictive performance (**Figure 6G**). To validate the model's generalizability and clinical utility, independent validation was conducted using the GSE39582 external cohort (**Figures 6H–6K**). The results confirmed that overall survival in the high-risk group remained significantly worse than that in the low-risk group ($P = 0.022$), fully substantiating the reliability and robustness of this prognostic model across multicenter datasets.

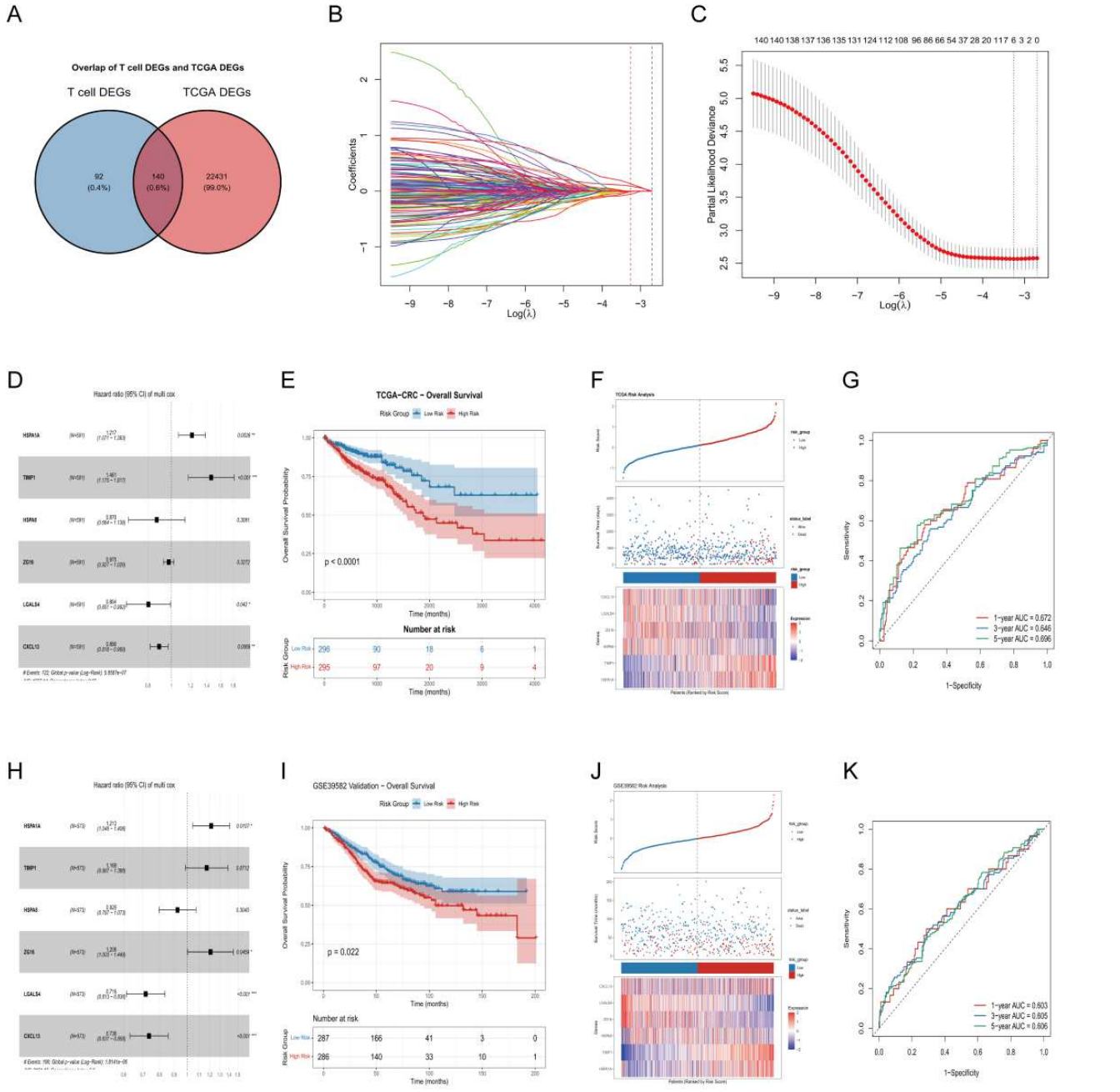


Figure 6. Construction and validation of a prognostic risk model based on T-cell marker genes. (A) Venn diagram showing the intersection of CRC-associated DEGs and T-cell-related signature genes. (B) LASSO coefficient path diagram for the 140 candidate genes. (C) Ten-fold cross-validation for parameter optimization in LASSO regression. (D) Forest plot of optimal prognostic genes derived from multivariate Cox regression analysis in the training cohort. (E) Kaplan–Meier survival curves comparing overall survival between high- and low-risk groups in the training cohort. (F) Heatmap displaying the risk score distribution, survival status, and expression profiles of model genes in the training cohort. (G) Time-dependent ROC curves and AUC values for 1-, 3-, and 5-year survival predictions in the training cohort. (H) Multivariate Cox regression analysis for the validation cohort. (I) Kaplan–Meier survival curves for the external validation cohort (GSE39582). (J) Risk score distribution and expression heatmap for the validation cohort. (K) Time-dependent ROC curves and AUC values for 1-, 3-, and 5-year survival predictions in the validation cohort.

2.6 Analysis of the Correlation between Immune Cell Infiltration and Immune Checkpoints in Colorectal Cancer

This study comprehensively analyzed the TME profile and immunotherapy potential of CRC by integrating the 'ESTIMATE' algorithm, the 'CIBERSORT' algorithm, and survival analysis. First, the 'ESTIMATE' algorithm was used to evaluate the TME characteristics of patients in the high- and low-risk groups (**Figure 7A**). The results showed that patients in the high-risk group had significantly higher ESTIMATE scores, Immune Scores, and Stromal Scores than those in the low-risk group; conversely, patients in the low-risk group exhibited significantly higher tumor purity. Subsequently, the proportions of 22 tumor-infiltrating immune cell subsets were quantitatively predicted using the 'CIBERSORT' algorithm. Kaplan–Meier (KM) survival analysis indicated that the infiltration levels of different immune cells were significantly associated with patient prognosis: high infiltration levels of regulatory T cells predicted poorer overall survival ($P = 0.0037$), whereas high levels of infiltration of activated dendritic cells ($P = 0.0038$) and resting $CD4^+$ memory T cells ($P = 0.039$) were significantly associated with a favorable prognosis (**Figure 7B**). Furthermore, we compared the differences in the expression of immune checkpoint and immune-related genes between the two groups and assessed their prognostic value. Box plots revealed significant differences in the expression of several key immune checkpoint genes between the groups. Kaplan–Meier survival analysis indicated that high expression of the immune checkpoint genes CD244, CD48, CD28, CD27, CD8A, and PTPRC was significantly associated with better overall survival in patients; conversely, high expression of NRP1 and CD276 was associated with a significantly poor prognosis (**Figure 7C**).

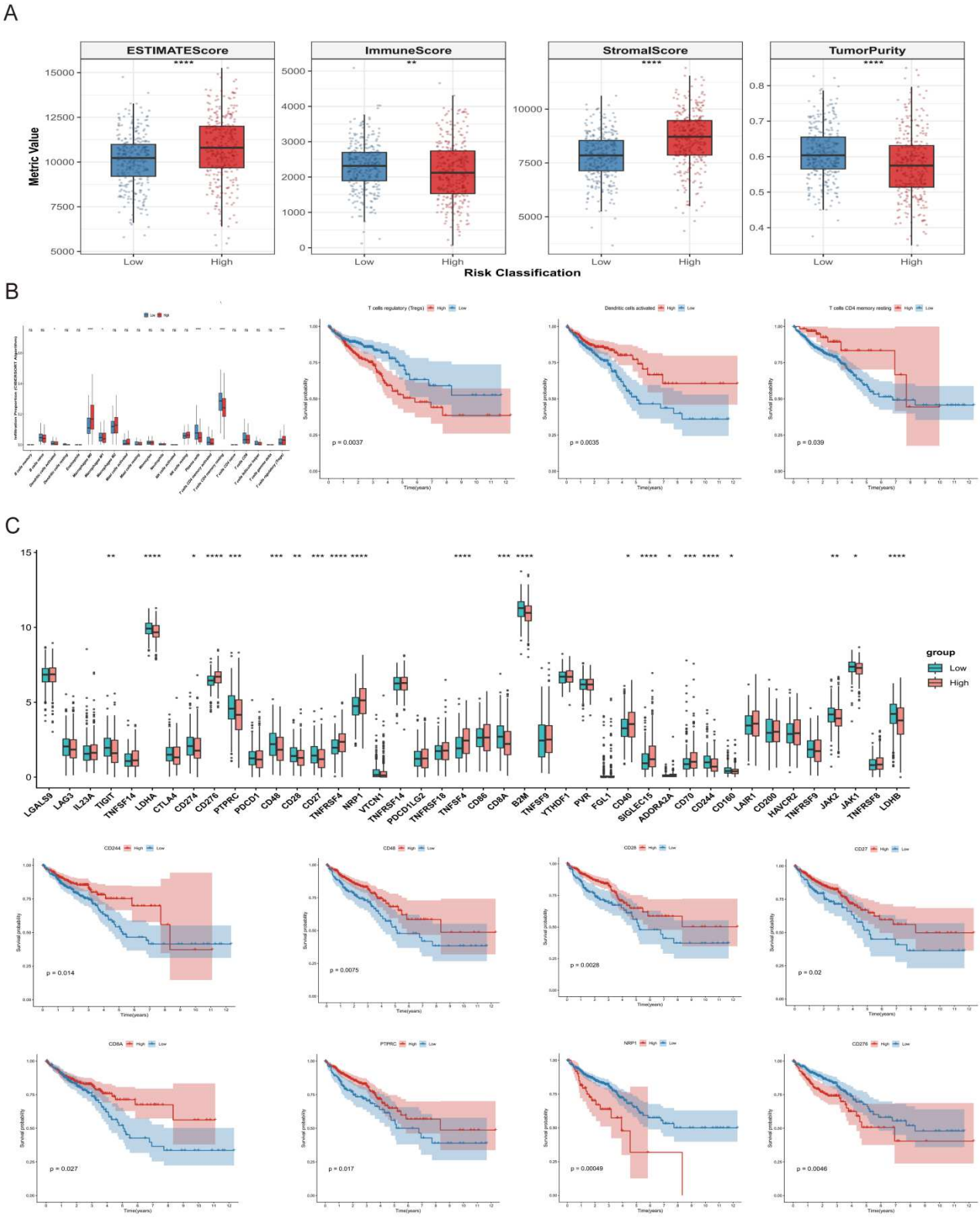


Figure 7. Correlation between the cellular prognostic risk model and immune cell infiltration and immune checkpoints in the TCGA cohort. (A) Box plots showing the ratio of stromal cells to immune cells and tumor purity across risk strata in colorectal cancer tissues. (B) Kaplan–Meier survival curves illustrating the prognostic significance of infiltrating immune cell subsets, including regulatory T cells, activated dendritic cells, and resting $CD4^+$ memory T cells. (C) Comparison of immune checkpoint expression levels and their corresponding prognostic stratification between high- and low-risk groups. ns, $P > 0.05$; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

3.Discussion

Colorectal cancer remains one of the most prevalent malignant tumors of the digestive tract, presenting persistently high incidence and mortality rates worldwide [11-12]. The pathogenesis of CRC stems from cumulative genetic and epigenetic alterations in colonic or rectal epithelial cells, driven by multi-factorial influences such as the tumor microenvironment, environmental conditions, and lifestyle, which collectively promote the gradual malignant transformation of normal mucosa [13-14]. Furthermore, high intratumoral heterogeneity and complex TME compositions are primary factors contributing to significant variations in patient prognosis and low response rates to immunotherapy [15-16]. Although ICB therapies—typified by programmed cell death protein 1 (PD-1) or programmed death-ligand 1 (PD-L1) inhibitors—have demonstrated significant clinical efficacy in patients with MSI-H or dMMR CRC [17-18], evidence from current clinical studies indicates that monotherapy with PD-1 or PD-L1 inhibitors yields minimal efficacy in microsatellite-stable (MSS) CRC, necessitating combinatorial immunotherapy strategies to achieve meaningful clinical benefit [19]. As core effector cells mediating specific anti-tumor immune responses within the TME, T lymphocytes dictate the intensity of local anti-tumor immunity and overall clinical outcomes through their subpopulation differentiation states [20], functional exhaustion trajectories, and intercellular communication networks. By integrating single-cell transcriptomics with large-scale bulk transcriptomics data, this study systematically mapped the T-cell immune landscape within the CRC microenvironment, elucidated pseudotemporal differentiation trajectories and paracrine communication mechanisms, and constructed a robust T-cell-related prognostic risk model capable of predicting survival and reflecting immune microenvironment phenotypes.

Utilizing the GSE315534 scRNA-seq dataset, this study identified nine major cell types from primary CRC lesions and matched normal tissues, encompassing T cells, B cells, monocytes, tissue stem cells, endothelial cells, secretory goblet cells, fibroblasts, NK cells, and smooth muscle cells. Based on canonical marker genes, 15,450 T lymphocytes were successfully annotated into five primary functional subsets: central memory T cells, regulatory T cells, effector memory T cells, effector T cells, and T helper 17 cells. To further uncover the functional states of tumor-infiltrating T cells and their roles within the CRC microenvironment, GO, KEGG, and GSEA were performed on DEGs across T-cell subsets. The results demonstrated that T-cell-associated DEGs were significantly enriched in critical immune processes, including leukocyte-mediated immunity, antiviral defense responses, cytotoxic activity, MHC protein complex binding, and antigen processing and presentation. Notably, viral infection-related pathways enriched in the KEGG analysis—such as "Influenza A" and "Coronavirus disease"—reflect a significant upregulation of interferon-stimulated genes (ISGs) and antiviral-like effector programs within tumor-infiltrating T cells, rather than indicating active tissue viral infection. This transcriptional signature mirrors a common molecular response of host immune cells to chronic antigenic stimulation and inflammatory cues within the TME [21], signifying an active activation state of tumor-infiltrating T cells. Furthermore, the differential expression of the Ribosome pathway suggests that protein translation mechanisms within tumor-infiltrating T cells undergo dynamic remodeling during immune response evolution, potentially correlating with metabolic reprogramming and functional exhaustion driven by sustained immune stimulation. Collectively, these enrichment findings provide robust transcriptional evidence that T cells in the CRC microenvironment undergo active immune responses and functional remodeling.

Intercellular communication analysis further revealed extensive signaling cross-talk between T cells and diverse stromal and immune constituents within the TME, with fibroblasts, monocytes, and endothelial cells exhibiting particularly intense interactions with T cells. Specifically, the CCL chemokine signaling pathway played a pivotal role, indicating that chemokine-driven immune cell recruitment, spatial localization, and

microenvironmental remodeling are essential for CRC immune regulation [22]. These observations confirm that T-cell functional states are governed not only by intrinsic transcriptional programs, but are also co-regulated by microenvironmental signals originating from tumor stromal and myeloid cells. Additionally, pseudotime trajectory analysis demonstrated that T cells undergo continuous developmental dynamics, categorized into three principal expression modules closely associated with protein homeostasis and memory maintenance, effector cytotoxic execution, and terminal immunosuppressive states (Tregs), respectively. These findings indicate that tumor-infiltrating T cells do not reside in a static functional state; rather, they undergo dynamic functional transitions driven by continuous antigen exposure and microenvironmental signaling, progressing from early immune maintenance toward effector cytotoxicity or terminal exhaustion and immunosuppression [21]. In summary, the systematic dissection of T-cell functional lineages and intercellular communication networks uncovers profound heterogeneity within the CRC tumor immune microenvironment, furnishing a theoretical framework for understanding T-cell-mediated anti-tumor immunity and immunotherapy resistance.

To evaluate the clinical prognostic value of T-cell-associated signatures in colorectal cancer, LASSO-Cox dimensionality reduction and multivariate Cox proportional hazards regression were employed to construct a six-gene prognostic model (TIMP1, ZG16, LGALS4, CXCL13, HSPA1A, and HSPA8) using the TCGA cohort, which was subsequently validated in an independent GEO cohort. These signature genes play critical roles in CRC etiology, tumor progression, extracellular matrix remodeling, and immune microenvironment regulation. TIMP1, a soluble glycoprotein secreted by stromal and cancer cells, is widely established as a biomarker of poor prognosis across multiple malignancies, with elevated expression correlating with reduced survival in CRC patients [23-24]. ZG16 and LGALS4 serve as core structural components maintaining intestinal epithelial and mucosal barrier integrity, and their frequent downregulation in CRC is intimately linked to compromised barrier function, increased tumor aggressiveness, and poor overall survival [25-28]. Specifically, the downregulation of ZG16 may facilitate the infiltration of pro-inflammatory factors into the TME by impairing barrier integrity, thereby indirectly augmenting Treg-mediated immunosuppression. CXCL13, functioning as a master chemokine recruiting CXCR5-expressing lymphocytes, is indispensable for the formation of tertiary lymphoid structures (TLS) and the maintenance of localized anti-tumor immunity [29]. Members of the heat shock protein family, HSPA1A and HSPA8, participate in protein folding and cellular stress responses [30]. As demonstrated by Guan et al., HSPA1A shields tumor cells against oxidative stress, inflammatory cytokines, and hypoxia, while promoting tumor proliferation, invasion, and metastasis, with its inhibition successfully inducing apoptosis in CRC cells [31]. Furthermore, HSPA8 correlates strongly with therapeutic responses across diverse oncological regimens, highlighting its viability as an actionable pharmacological target [32]. Together, TIMP1, ZG16, LGALS4, CXCL13, HSPA1A, and HSPA8 represent robust prognostic biomarkers for CRC and offer novel entry points for targeted interventions. Furthermore, Kaplan–Meier survival curves and time-dependent ROC analyses across both the TCGA training set and the external independent validation set (GSE39582) consistently confirmed that overall survival was significantly diminished in the high-risk group compared with the low-risk group, substantiating the model's robustness, cross-platform generalizability, and high clinical value for prognostic stratification.It is noteworthy that the six identified signature genes—TIMP1, ZG16, LGALS4, CXCL13, HSPA1A, and HSPA8—encompass a spectrum of genes expressed in both T cells and the broader tumor microenvironment. While some, such as CXCL13, are hallmark indicators of T-cell activation and tertiary lymphoid structure formation, others like HSPA1A and ZG16 predominantly reflect the stress-adaptation and epithelial-barrier states of the tumor tissue. We propose that this prognostic signature should be interpreted as a composite metric of the 'T-cell-conditioned TME'. Rather than representing exclusively T-cell-derived transcripts, this model captures the functional dialogue between T-cell-mediated immune surveillance and the stress-responsive tumor ecosystem, which collectively dictates clinical outcomes.

The tumor microenvironment dictates cancer progression and governs clinical responses to immunotherapy. Given that the TME heavily influences both prognosis and therapeutic efficacy, this study investigated immune cell infiltration landscapes across risk strata. Patients in the high-risk group exhibited significantly elevated ESTIMATE scores, Immune Scores, and Stromal Scores alongside reduced tumor purity, presenting a hallmark "stroma-rich" and "immune-suppressive" TME phenotype. Utilizing the CIBERSORT algorithm coupled with Kaplan–Meier survival analyses, we established that high infiltration levels of activated dendritic cells and resting $CD4^+$ memory T cells predicted a favorable prognosis, underscoring that targeted antigen presentation and immune memory maintenance are pivotal for tumor suppression [33-34]. Conversely, elevated regulatory T cell infiltration was significantly associated with inferior overall survival, confirming that Tregs enriched within high-risk TMEs mediate potent local immunosuppression [35]. Furthermore, our analysis revealed a significant positive correlation between the risk score and Treg infiltration. Given that Tregs mediate potent local immunosuppression and expand markedly in advanced CRC, their enrichment in the high-risk group provides a direct mechanistic link to disease progression, immune evasion, and the observed inferior overall survival. With the rapid expansion of immune checkpoint therapy in CRC, optimizing patient stratification remains critical. Our microenvironmental analyses revealed that low-risk patients harbor a significantly more active and coordinated immune phenotype. Evaluation of immune-related genes demonstrated that CD244, CD48, CD28, CD27, CD8A, and PTPRC were significantly upregulated in low-risk patients and associated with superior overall survival. These transcripts synergistically reflect heightened immune infiltration, robust T-cell activation, and active immune signaling networks. Specifically, CD244 modulates immune cell cytotoxicity and regulates the generation and survival of effector and memory T cells when co-expressed with its ligand CD48 [36], reflecting coordinated innate and adaptive immune cross-talk [37]. Upregulation of CD28 and CD27 denotes an enhanced T-cell co-stimulatory network; CD28 signaling drives proliferation, cytokine secretion, and effector maintenance [38-39], whereas CD27 promotes memory T-cell generation and sustained immune recall [40]. Elevated CD8A expression further confirms robust cytotoxic immune surveillance—aligning with evidence from Zheng et al [41]. linking low CD8A expression to dismal survival outcomes in immunotherapy-treated cohorts. Additionally, PTPRC (CD45), a critical regulator of antigen receptor-mediated lymphocyte activation and survival, reflects pronounced immune cell infiltration in low-risk tumors [42-43]. In contrast, NRP1 and CD276 were significantly overexpressed in high-risk patients and associated with adverse survival. NRP1 promotes tumor angiogenesis, cellular migration, and immunosuppressive niche establishment [44-46], whereas CD276 (B7-H3) acts as a prominent immune checkpoint molecule driving immune evasion by dampening anti-tumor T-cell responses [47-48]. Consequently, the overexpression of NRP1 and CD276 synergistically fosters an immune-tolerant microenvironment in high-risk patients, predisposing them to immune checkpoint inhibitor resistance. Overall, our prognostic risk model effectively discriminates patient survival outcomes while exposing profound immunological heterogeneity between risk tiers, offering a strong theoretical foundation for immune stratification, prognostic evaluation, and precision immunotherapy design in CRC.

Notwithstanding these clinically valuable insights derived from multi-omics integration, several limitations warrant acknowledgment. First, our prognostic models and microenvironmental inferences rely predominantly on retrospective transcriptomic data sourced from public repositories; their clinical utility necessitates rigorous prospective validation across large-scale, multi-center clinical cohorts. Second, the precise molecular mechanisms governing the six core signature genes within CRC cells and distinct T-cell compartments, their paracrine signaling dynamics, and the precise biochemical pathways driving T-cell exhaustion require further functional elucidation through in vitro co-culture assays and in vivo murine tumor models.

4. Conclusion

In summary, this study explored the T-cell landscape and the TME in CRC using scRNA-seq and bulk RNA-seq, successfully identifying key potential prognostic genes. This investigation elucidated the functional enrichment profiles, cellular differentiation trajectories, and intercellular communication heterogeneity of T cells within CRC. Furthermore, the established T-cell-related prognostic risk model demonstrated robust independent predictive value. Our work not only expands the current understanding of CRC-associated T cells but also provides novel insights into the complex immunological characteristics of the TME. We hope that this integrative scRNA-seq and bulk RNA-seq approach will advance the development of precision immunotherapies for CRC, while acknowledging that these findings warrant further validation through comprehensive experimental and clinical studies.

Acknowledgements

We would like to express our sincere gratitude to the researchers who have provided raw, open-access single-cell data. We would also like to thank the research teams that have played a key role in generating data using colorectal cancer (CRC) samples from the TCGA and GEO databases.

Author contributions

X. Y. , Y. Z. participated in the design of research ideas and the drafting of manuscripts, analysed the data. X. C. , J. L. , H. W. , P. L. made some well-intentioned suggestions. J. L. , M. M. , T. Z. , J. X. , K. L. , H. L. helped draw relevant pictures.

Funding

This work was supported by the Program for Innovative Talents of the National Natural Science Foundation of China (Grant No. TP-001-2025).

Data Availability Statement

This study only conduct analysis based on GEO and TCGA public database, with no experiments on humans and/or the use of human tissue samples involved. GSE315534 ([GSE315534[Accession]-GEO DataSets - NCBI] (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE315534)) and GSE39582 ([GSE39582[Accession]-GEO DataSets - NCBI] (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39582)). The bulk RNA-seq datasets analyzed in this study were obtained from The Cancer Genome Atlas (TCGA) program, specifically the TCGA-COAD (Colon Adenocarcinoma) and TCGA-READ (Rectum Adenocarcinoma) cohorts. These datasets are publicly available through the National Cancer Institute (NCI) Genomic Data Commons (GDC) Data Portal (https://portal.gdc.cancer.gov/). The data were accessed and downloaded using the 'TCGAbiolinks' R package (version 4.4.3). The specific projects can be accessed directly at https://portal.gdc.cancer.gov/projects/TCGA-COAD and https://portal.gdc.cancer.gov/projects/TCGA-READ. All data used in this study are publicly available and no restrictions apply.

Declarations

Ethical Approval and Accordance

Not applicable.

Consent to participate

Not applicable.

Consent to publish

Not applicable.

References

1. Sung H, Filho AM, Laversanne M, et al. Global cancer statistics 2024: GLOBOCAN estimates of incidence and mortality worldwide for 34 cancers in 186 countries. *CA: A Cancer Journal for Clinicians*. 2026;76(4):e70090. https://doi.org/10.3322/caac.70090.
2. Arnold M, Abnet CC, Neale RE, et al. Global Burden of 5 Major Types of Gastrointestinal Cancer. *Gastroenterology*. 2020;159(1):335-349.e15. https://doi.org/10.1053/j.gastro.2020.02.068.
3. Zheng Z, Wieder T, Mauerer B, Schäfer L, Kesselring R, Braumüller H. T Cells in Colorectal Cancer: Unravelling the Function of Different T Cell Subsets in the Tumor Microenvironment. *International Journal of Molecular Sciences*. 2023;24(14):11673. https://doi.org/10.3390/ijms241411673.
4. Binnewies M, Roberts EW, Kersten K, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nature Medicine*. 2018;24(5):541-550. https://doi.org/10.1038/s41591-018-0014-x.
5. Chen X, Liu J, Lu P, et al. Unraveling Traditional Chinese Medicine with single-cell RNA sequencing: Current applications and future frontiers. *Phytomedicine*. 2025;149:157556. https://doi.org/10.1016/j.phymed.2025.157556.
6. Aran D, Looney AP, Liu L, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature Immunology*. 2019;20(2):163-172. https://doi.org/10.1038/s41590-018-0276-y.
7. Jin S, Guerrero-Juarez CF, Zhang L, et al. Inference and analysis of cell-cell communication using CellChat. *Nature Communications*. 2021;12(1):1088. https://doi.org/10.1038/s41467-021-21246-9.
8. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature Communications*. 2013;4(1):2612. https://doi.org/10.1038/ncomms3612.
9. Wang Y, Shi T, Song X, Liu B, Wei J. Gene fusion neoantigens: Emerging targets for cancer immunotherapy. *Cancer Letters*. 2021;506:45-54. https://doi.org/10.1016/j.canlet.2021.02.023.
10. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods*. 2015;12(5):453-457. https://doi.org/10.1038/nmeth.3337.
11. Siegel RL, Kratzer TB, Wagle NS, Sung H, Jemal A. Cancer statistics, 2026. *CA: A Cancer Journal for Clinicians*. 2026;76(1):e70043. https://doi.org/10.3322/caac.70043.
12. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*. 2021;71(3):209-249. https://doi.org/10.3322/caac.21660.
13. Abedizadeh R, Majidi F, Khorasani HR, Abedi H, Sabour D. Colorectal cancer: a comprehensive review of carcinogenesis, diagnosis, and novel strategies for classified treatments. *Cancer and Metastasis Reviews*. 2024;43(2):729-753. https://doi.org/10.1007/s10555-023-10158-3.
14. Brabletz T, Spaderna S, Kolb J, et al. Down-Regulation of the Homeodomain Factor Cdx2 in Colorectal Cancer by Collagen Type I: An Active Role for the Tumor Environment in Malignant Tumor Progression.
15. Binnewies M, Roberts EW, Kersten K, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nature Medicine*. 2018;24(5):541-550.

https://doi.org/10.1038/s41591-018-0014-x.
16. Siveen KS, Raza A, Ahmed EI, et al. The Role of Extracellular Vesicles as Modulators of the Tumor Microenvironment, Metastasis and Drug Resistance in Colorectal Cancer. *Cancers*. 2019;11(6):746. https://doi.org/10.3390/cancers11060746.
17. Elfishawy M, Abd-ELaziz SA, Hegazy A, El-Yasergy DF. Immunohistochemical Expression of Programmed Death Ligand-1 (PDL-1) in Colorectal carcinoma and Its Correlation with Stromal Tumor Infiltrating Lymphocytes. *Asian Pacific Journal of Cancer Prevention*. 2020;21(1):225-232. https://doi.org/10.31557/APJCP.2020.21.1.225.
18. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency.
19. Wu H, Deng M, Xue D, et al. PD-1/PD-L1 inhibitors for early and middle stage microsatellite high-instability and stable colorectal cancer: a review. *International Journal of Colorectal Disease*. 2024;39(1):83. https://doi.org/10.1007/s00384-024-04654-3.
20. Kang TG, Johnson JT, Zebley CC, Youngblood B. Epigenetic regulation of T cell exhaustion in cancer. *Nature Reviews Cancer*. 2026;26(1):46-61. https://doi.org/10.1038/s41568-025-00883-y.
21. Schietinger A, Greenberg PD. Tolerance and exhaustion: defining mechanisms of T cell dysfunction. *Trends in Immunology*. 2014;35(2):51-60. https://doi.org/10.1016/j.it.2013.10.001.
22. Nagarsheth N, Wicha MS, Zou W. Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy. *Nature Reviews Immunology*. 2017;17(9):559-572. https://doi.org/10.1038/nri.2017.49.
23. Song G, Xu S, Zhang H, et al. TIMP1 is a prognostic marker for the progression and metastasis of colon cancer through FAK-PI3K/AKT and MAPK pathway. *Journal of Experimental & Clinical Cancer Research*. 2016;35(1):148. https://doi.org/10.1186/s13046-016-0427-7.
24. Inagaki D, Oshima T, Yoshihara K, et al. Overexpression of Tissue Inhibitor of Metalloproteinase-1 Gene Correlates with Poor Outcomes in Colorectal Cancer. *ANTICANCER RESEARCH*. Published online 2010.
25. Chen X, Du P, She J, Cao L, Li Y, Xia H. Loss of ZG16 is regulated by miR-196a and contributes to stemness and progression of colorectal cancer. *Oncotarget*. 2016;7(52):86695-86703. https://doi.org/10.18632/oncotarget.13435.
26. Meng H, Li W, Boardman LA, Wang L. Loss of ZG16 is associated with molecular and clinicopathological phenotypes of colorectal cancer. *BMC Cancer*. 2018;18(1):433. https://doi.org/10.1186/s12885-018-4337-2.
27. Michalak M, Golde V, Helm D, Kaltner H, Gebert J, Kopitz J. Combining Recombinase-Mediated Cassette Exchange Strategy with Quantitative Proteomic and Phosphoproteomic Analyses to Inspect Intracellular Functions of the Tumor Suppressor Galectin-4 in Colorectal Cancer Cells. *International Journal of Molecular Sciences*. 2022;23(12):6414. https://doi.org/10.3390/ijms23126414.
28. Satelli A, Rao PS, Thirumala S, Rao US. Galectin-4 functions as a tumor suppressor of human colorectal cancer. *International Journal of Cancer*. 2011;129(4):799-809. https://doi.org/10.1002/ijc.25750.
29. Sautès-Fridman C, Lawand M, Giraldo NA, et al. Tertiary Lymphoid Structures in Cancers: Prognostic Value, Regulation, and Manipulation for Therapeutic Intervention. *Frontiers in Immunology*. 2016;7. https://doi.org/10.3389/fimmu.2016.00407.
30. Jiang W, Pan X, Yan H, Wang G. Prognostic Significance of the Hsp70 Gene Family in Colorectal Cancer. *Medical Science Monitor*. 2021;27. https://doi.org/10.12659/MSM.928352.
31. Guan Y, Zhu X, Liang J, Wei M, Huang S, Pan X. Upregulation of HSPA1A/HSPA1B/HSPA7 and Downregulation of HSPA9 Were Related to Poor Survival in Colon Cancer. *Frontiers in Oncology*. 2021;11:749673. https://doi.org/10.3389/fonc.2021.749673.
32. Li B, Ming H, Qin S, et al. HSPA8 Activates Wnt/β-Catenin Signaling to Facilitate BRAF V600E Colorectal Cancer Progression by CMA-Mediated CAV1 Degradation. *Advanced Science*. 2024;11(3):2306535. https://doi.org/10.1002/advs.202306535.

33. Dadabayev AR, Sandel MH, Menon AG, et al. Dendritic cells in colorectal cancer correlate with other tumor-infiltrating immune cells. *Cancer Immunology, Immunotherapy*. 2004;53(11):978-986. https://doi.org/10.1007/s00262-004-0548-2.
34. DiToro D, Basu R. Emerging Complexity in CD4+T Lineage Programming and Its Implications in Colorectal Cancer. *Frontiers in Immunology*. 2021;12:694833. https://doi.org/10.3389/fimmu.2021.694833.
35. Bonertz A, Weitz J, Pietsch DHK, et al. Antigen-specific Tregs control T cell responses against a limited repertoire of tumor antigens in patients with colorectal carcinoma. *Journal of Clinical Investigation*. Published online October 5, 2009:JCI39608. https://doi.org/10.1172/JCI39608.
36. Sun L, Gang X, Li Z, et al. Advances in Understanding the Roles of CD244 (SLAMF4) in Immune Regulation and Associated Diseases. *Frontiers in Immunology*. 2021;12:648182. https://doi.org/10.3389/fimmu.2021.648182.
37. McArdel SL, Terhorst C, Sharpe AH. Roles of CD48 in regulating immunity and tolerance. *Clinical Immunology*. 2016;164:10-20. https://doi.org/10.1016/j.clim.2016.01.008.
38. Kamphorst AO, Wieland A, Nasti T, et al. Rescue of exhausted CD8 T cells by PD-1–targeted therapies is CD28-dependent. *Science*. 2017;355(6332):1423-1427. https://doi.org/10.1126/science.aaf0683.
39. Honikel MM, Olejniczak SH. Co-Stimulatory Receptor Signaling in CAR-T Cells. *Biomolecules*. 2022;12(9):1303. https://doi.org/10.3390/biom12091303.
40. Ochsenbein AF, Riddell SR, Brown M, et al. CD27 Expression Promotes Long-Term Survival of Functional Effector–Memory CD8 + Cytotoxic T Lymphocytes in HIV-infected Patients. *The Journal of Experimental Medicine*. 2004;200(11):1407-1417. https://doi.org/10.1084/jem.20040717.
41. Zheng Z, Guo Y, Huang X, et al. CD8A as a Prognostic and Immunotherapy Predictive Biomarker Can Be Evaluated by MRI Radiomics Features in Bladder Cancer. *Cancers*. 2022;14(19):4866. https://doi.org/10.3390/cancers14194866.
42. Al Barashdi MA, Ali A, McMullin MF, Mills K. Protein tyrosine phosphatase receptor type C (PTPRC or CD45). *Journal of Clinical Pathology*. 2021;74(9):548-552. https://doi.org/10.1136/jclinpath-2020-206927.
43. Li X, Yue Z, Wang D, Zhou L. PTPRC functions as a prognosis biomarker in the tumor microenvironment of cutaneous melanoma. *Scientific Reports*. 2023;13(1):20617. https://doi.org/10.1038/s41598-023-46794-6.
44. Staton CA, Koay I, Wu JM, Hoh L, Reed MWR, Brown NJ. Neuropilin-1 and neuropilin-2 expression in the adenoma–carcinoma sequence of colorectal cancer. *Histopathology*. 2013;62(6):908-915. https://doi.org/10.1111/his.12098.
45. Ma L li, Guo L li, Luo Y, et al. Cdc42 subcellular relocation in response to VEGF/NRP1 engagement is associated with the poor prognosis of colorectal cancer. *Cell Death & Disease*. 2020;11(3):171. https://doi.org/10.1038/s41419-020-2370-y.
46. Ochiumi T, Kitadai Y, Tanaka S, Akagi M, Yoshihara M, Chayama K. Neuropilin-1 is involved in regulation of apoptosis and migration of human colon cancer. *International Journal of Oncology*. Published online July 1, 2006. https://doi.org/10.3892/ijo.29.1.105.
47. Lu Z, Zhao ZX, Cheng P, et al. B7-H3 immune checkpoint expression is a poor prognostic factor in colorectal carcinoma. *Modern Pathology*. 2020;33(11):2330-2340. https://doi.org/10.1038/s41379-020-0587-z.
48. Soares J, Ferreira D, Miranda A, et al. CD276 immature glycosylation drives colorectal cancer aggressiveness and T cell mediated immune escape. *Cell Communication and Signaling*. 2026;24(1):113. https://doi.org/10.1186/s12964-026-02672-y.