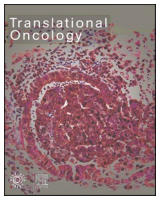
Translational Oncology 68 (2026) 102761

Contents lists available at ScienceDirect

Translational Oncology

journal homepage: www.elsevier.com/locate/tranon



Comprehensive single-cell analysis reveals cellular heterogeneity and immune interactions in colorectal cancer

Zhenyu Chi [a,1,*], Rui Kong [b,1], Song Wang [c,1], Shaofan Qiu [d,*]

[a] People's Hospital of Anji, Anji, Zhejiang, 313300, China
[b] Department of Oncology, The Third Affiliated Hospital of Chongqing Medical University, Chongqing, 401120, China
[c] Department of Colorectal Surgery, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, 310006, China
[d] Department of Gastrointestinal Surgery, The Second Hospital of Hebei Medical University, Shijiazhuang, Hebei, 050000, China

ARTICLE INFO

Keywords:
Colorectal cancer
Single-cell RNA sequencing
Tumor microenvironment
Immune cell subtypes
Prognostic biomarkers

ABSTRACT

Colorectal cancer (CRC) presents considerable therapeutic challenges due to its diverse cellular composition and intricate microenvironment. Our study utilized single-cell RNA sequencing (scRNA-seq) on CRC tissues, examining 18,741 individual cells, which were grouped into six primary cell populations: epithelial, fibroblast, endothelial, T and NK, B, and myeloid. The epithelial cells exhibited notable variations in gene copy numbers. Within T_NK cells, we identified four distinct subsets. CytoTRACE analysis indicated that subtype C3 exhibited lower differentiation potential, whereas subtypes C0 and C1 showed higher differentiation potential. Consistently, Monocle pseudotime trajectory analysis positioned C3 cells at the terminal stage of differentiation, while C0 cells were enriched at the early stage of the developmental trajectory, suggesting functional heterogeneity among T/NK subpopulations. Through functional analyses with GSVA and ssGSEA, subtype C3 displayed the highest inflammation-associated activity scores. Further exploration of transcription factors defined three unique regulatory clusters among T_NK cells, illuminating their gene-regulation networks. We developed a prognostic signature using markers from subtype C3 T_NK cells combined with age-associated genes, revealing a significant correlation with patient survival outcomes. This prognostic model proved effective in categorizing CRC patients according to risk. Additionally, immune profiling employing ESTIMATE, CIBERSORT, and Xcell algorithms underscored the complexity of immune cell populations within CRC tumors. Analysis of tumor mutational burden (TMB) highlighted differential patterns between patient groups and its relationship to prognostic risk levels. Collectively, these insights provide a detailed perspective on CRC cell diversity and immune dynamics, supporting the advancement of targeted and personalized therapeutic interventions.

Introduction

Colorectal cancer (CRC) remains among the leading global causes of cancer-related illness and mortality, posing significant public health challenges worldwide [1,2]. Although considerable progress has been achieved in understanding CRC pathology and improving treatment options, the complex heterogeneity of the disease continues to hinder effective clinical management [3,4]. This complexity arises primarily from the intricate interplay between tumor cells and their microenvironment, including diverse cellular and molecular interactions that collectively influence tumor progression and therapeutic outcomes. Traditional bulk RNA sequencing methods have significantly contributed to the understanding of CRC; however, they often obscure crucial details of tumor heterogeneity by averaging gene expression signals across diverse cell populations [9,10]. This averaging mask rare or critical cell subpopulations and the dynamic intercellular interactions crucial for disease progression and therapeutic response. Consequently, these limitations underline the necessity for advanced approaches capable of dissecting the complex tumor ecosystem in greater detail. Single-cell RNA sequencing (scRNA-seq) has emerged as a transformative technique, enabling unprecedented resolution in cancer research by characterizing the cellular and molecular heterogeneity at the individual cell level [5,6]. Unlike bulk sequencing, scRNA-seq permits the detailed exploration of distinct cellular populations within

* Corresponding authors.
E-mail addresses: ak47czy1990@163.com (Z. Chi), 28602153@hebmu.edu.cn (S. Qiu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.tranon.2026.102761
Received 28 December 2025; Received in revised form 9 March 2026; Accepted 9 April 2026

tumors, capturing rare cells and their unique gene expression profiles [7,8]. This granular perspective allows for precise delineation of cellular differentiation trajectories, assessment of genomic alterations such as copy number variations (CNVs), and elucidation of intercellular communication pathways [11,12]. A critical gap in current CRC research is the incomplete characterization and understanding of the diverse cell types within the tumor microenvironment. Beyond malignant epithelial cells, CRC tumors comprise numerous stromal and immune cells—such as fibroblasts, endothelial cells, T cells, natural killer (NK) cells, B cells, and various myeloid cells—that significantly influence tumor biology and patient outcomes [13,14]. Each cellular component contributes uniquely to cancer progression, immune modulation, and treatment responsiveness [15]. For example, tumor-associated fibroblasts can actively promote metastasis and tumor growth, whereas immune cells can either exert antitumor activities or facilitate immune escape, depending on their subtype and activation state [16–18]. Recent single-cell studies in CRC have already provided valuable insights, revealing multiple epithelial cell subpopulations with distinctive molecular signatures and diverse immune cell subsets within the tumor microenvironment [19,20]. Additionally, single-cell analyses have identified rare cell populations potentially linked to therapy resistance and disease relapse, thereby highlighting novel therapeutic targets [21,22]. Nevertheless, comprehensive single-cell profiling that integrates cellular differentiation trajectories, detailed CNV analysis, and cell-cell communication within CRC tumors remains relatively unexplored. Therefore, this study aims to address these gaps by performing a comprehensive single-cell analysis of colorectal cancer. Specifically, we seek to characterize cell type-specific gene expression patterns, assess copy number variations, map cellular differentiation trajectories, and elucidate intercellular communication pathways within the CRC tumor microenvironment. Employing advanced computational analyses alongside scRNA-seq technologies, this research will uncover the intricate molecular and cellular dynamics underpinning CRC progression. By deepening our understanding of these mechanisms, we aim to contribute significantly towards identifying novel therapeutic strategies, ultimately improving clinical outcomes for CRC patients [23,24].

Methods

Data acquisition and preparation for transcriptomic analysis

Gene expression data and relevant patient clinical information for colorectal cancer cases were retrieved from the publicly accessible database known as The Cancer Genome Atlas (TCGA). This research analysingzed a set of 606 colorectal cancer patient records, which served as the main dataset to develop a predictive model of patient prognosis. Moreover, a separate dataset was set aside exclusively to assess and confirm the reliability and accuracy of the constructed prognostic model. Initially, the raw RNA sequencing data were standardized by converting gene expression counts into a commonly used measurement called Transcripts Per Million (TPM). Following this normalization step, the TPM values underwent further refinement through a log2 conversion. This transformation allowed the data to be uniformly scaled and facilitated consistent computational analyses.

Acquisition and computational analysis of scRNA-seq data

Single-cell RNA sequencing (scRNA-seq) data were retrieved from six colorectal cancer specimens available in Gene Expression Omnibus (GEO; accession: GSE231559). Samples were stratified by patient age, categorized as either “Young” (<50 years) or “Old” (>50 years). Computational analyses were conducted using R software (version 4.1.3), predominantly through the Seurat package. Strict quality control (QC) parameters were enforced initially. Cells that passed QC exhibited mitochondrial gene expression below 30%, unique molecular identifiers (UMIs) ranging between 200 and 25,000, and gene counts per cell from 200 to 6000. Following QC, normalization was executed via the Seurat NormalizeData function. Subsequently, 2000 highly variable genes were identified using the FindVariableFeatures function in Seurat. Data scaling was performed using the ScaleData function, regressing out the cell-cycle effects denoted by “S.Score” and “G2M.Score.” Batch effects among samples were corrected employing the Harmony integration algorithm. Cell visualization and clustering involved dimensionality reduction techniques, specifically Uniform Manifold Approximation and Projection (UMAP) and t-distributed Stochastic Neighbor Embedding (t-SNE). Cell clusters were defined using Seurat’s integrated Louvain clustering method. Differentially expressed genes (DEGs) among clusters were identified by the FindAllMarkers function in Seurat, applying thresholds of p-value $\langle$ 0.05, log2 fold-change $\rangle$ 0.25, and detection in at least 10% of cells per cluster.

Collection of genes related to inflammation

Genes associated with inflammation were retrieved from an online database called the Molecular Signatures Database (MSigDB). In particular, the analysis utilized a specific set of genes referred to as "HALLMARK_INFLAMMATORY_RESPONSE". The complete list of these inflammation-related genes was downloaded directly from the MSigDB website, available at this link: https://www.gsea-msigdb.org/gsea/msigdb/cards/HALLMARK_INFLAMMATORY_RESPONSE.html, and was subsequently applied in the data analyses.

Classification and annotation of cell populations

Cell types were categorized based on the expression of specific marker genes characteristic of distinct cellular populations. Gene-expression profiles of established markers unique to each cell category guided this classification, as listed below:

- **Epithelial Cells:** EPCAM, KRT18, KRT19, CDH1
- **Fibroblasts:** DCN, THY1, COL1A1, COL1A2
- **Endothelial Cells:** PECAM1, CLDN5, FLT1, RAMP2
- **T Cells:** CD3D, CD3E, CD3G, TRAC
- **Natural Killer (NK) Cells:** NKG7, GNLY, NCAM1, KLRD1
- **B Cells:** CD79A, IGHM, IGHG3, IGHA2
- **Myeloid Cells:** LYZ, MARCO, CD68, FCGR3A
- **Mast Cells:** KIT, MS4A2, GATA2

After initial classification, dimensionality reduction visualization via t-distributed Stochastic Neighbor Embedding (t-SNE) confirmed accurate cell-type annotation. Furthermore, violin plots illustrating marker gene expression within each cell group validated these annotations, clearly demonstrating distinct gene-expression patterns across identified cell populations.

Immune cell subtype analysis

Immune cells were extracted from the primary single-cell RNA-seq dataset based on previously annotated major cell types. Subtype classification was performed using *Sc*-Type (version 1.0.0), an automated cell-type identification tool that assigns cell identities by comparing gene expression profiles with curated reference marker gene sets for immune cell populations. For the analysis, we used default parameters, including a minimum expression of 200 genes per cell, a minimum of 3 cells expressing each marker gene, and a confidence threshold of 0.5 for cell-type assignment.

Analysis of copy number variation (CNV) at single-cell level

Copy number variations (CNVs) at the single-cell level were inferred using InferCNV v1.10.1. Natural killer T (NKT) cells were used as the normal reference. Genes expressed in at least 3 cells were included.

CNVs were calculated using a 100-gene sliding window with smoothing set to 0.1, and hierarchical clustering (Ward's method) was applied to group cells with similar CNV profiles. CNV heatmaps were generated to identify potential malignant epithelial cells, and results were validated by comparing known cancer-associated CNV regions and marker gene expression patterns.

Pseudotime analysis of cellular trajectories at single-cell level

The developmental pathways of T and NK cell populations were studied using pseudotime trajectory analysis. For this purpose, the Monocle2 software package was applied. Within Monocle2, dimensionality reduction was performed using a method known as DDRTree. Throughout the entire analysis, the software's original default settings were used without modifications. This allowed the visualization of how T and NK cells developed and transitioned between different states clearly and systematically.

Single-cell transcription factor analysis

The transcription factors active within the T and NK cell subsets were identified and characterized using a computational tool called SCENIC. For analysis, SCENIC employed two primary methods: the GRNBoost algorithm for gene regulatory network (GRN) inference and the RcisTarget method for detecting enriched transcription factor binding motifs. In particular, RcisTarget examined our gene sets to highlight significantly enriched transcription factor motifs. After identifying the enriched motifs, the AUCell algorithm was applied. This step calculated activity scores, clearly indicating the strength of each transcription factor's influence within different cell groups.

Cell-to-cell communication analysis

Interactions between different cell groups were investigated using a computational tool called CellChat. First, gene expression data that had been normalized previously were loaded into the CellChat software. Next, CellChat's standard functions—specifically, identifyOverExpressedGenes, identifyOverExpressedInteraction, and ProjectData—were applied without altering any default parameters. These functions helped find genes and interactions with notably increased expression. After these steps, the software was used to calculate interactions involving ligands and receptors among the different cell clusters. This was accomplished through several built-in CellChat functions, including computeCommunProb, filterCommunication, and computeCommunProbPathway. Lastly, all identified interactions were summarized, and a detailed network of cell-to-cell communications was visually represented using the CellChat aggregateNet function.

Calculation of inflammatory signature scores

Inflammatory responses at the single-cell level were evaluated using a predefined set of inflammation-associated genes. These genes served as inputs for computing enrichment scores, indicative of cellular inflammatory activity. Enrichment scoring was conducted using Gene Set Variation Analysis (GSVA) and Single-Sample Gene Set Enrichment Analysis (ssGSEA), both methods implemented within the GSVA R package.

Immune cell infiltration assessment

The degree of immune cell infiltration in patient groups categorized by risk was examined using an analytical tool called IOBR. Specifically, three distinct computational algorithms—ESTIMATE, CIBERSORT, and xCell—were utilized. These algorithms helped to thoroughly assess how abundant and diverse immune cells were within the defined patient categories.

Functional enrichment analysis

Functional enrichment analyses were performed to explore biological implications of selected gene sets. The analyses employed the clusterProfiler R package, referencing both Kyoto Encyclopedia of Genes and Genomes and Gene Ontology databases. Statistical significance was assessed using adjusted p-values (<0.05), corrected with the Benjamini–Hochberg method. Visualization of significantly enriched biological pathways and processes was generated using the ggplot2 R package.

Analysis and comparison of genomic mutations

To examine and compare the genetic mutation profiles between the two different patient groups, an R software package named maftools was employed. Using this software, mutation data were systematically analyzed to measure and contrast the total mutation counts (mutation burden) between patient subsets. Additionally, the relationship between tumor mutation burden (TMB) and patient risk scores was assessed. Lastly, the calculated TMB values were integrated into patient survival analysis to determine their possible effects on clinical outcomes.

Creation of a prognostic model using LASSO—Cox analysis with C3 T_NK cells and age-linked genes

A statistical model was constructed to predict patient prognosis by integrating selected marker genes from the inflammation-rich C3 T_NK cell cluster and genes consistently expressed in both patient age categories. Initially, candidate prognostic genes were identified through univariate Cox regression analysis, applying a significance threshold (p-value < 0.05). Subsequently, gene selection was refined via Least Absolute Shrinkage and Selection Operator (LASSO)-Cox regression using the glmnet R package. Model accuracy and reliability were validated by computing Area Under the Curve (AUC) metrics at 1-, 3-, and 5-year intervals with the timeROC R package.

Clinical sample collection and preparation

Tumor and matched normal tissues were obtained from five colorectal cancer patients with liver metastases. Normal samples were harvested at least 3 cm away from tumor margins. Tissue collection occurred during surgeries conducted at the second hospital of Hebei Medical University from May 2019 to April 2024. Post-excision, tissues were rapidly snap-frozen in liquid nitrogen and stored at −80 °C to preserve RNA integrity for molecular analyses. Ethical approval for this study and patient tissue usage was granted by the Research Ethics Committee of The Second Hospital of Hebei Medical University (2026-R027).

RNA extraction and quantitative real-time PCR (qRT-PCR)

otal RNA was extracted from cells and tissues using TRIzol reagent (Invitrogen, USA) according to instructions. RNA purity and amount were checked using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA). Only RNA with A260/A280 values between 1.8 and 2.0 was used. Next, cDNA was made from 1 μg RNA per sample with the PrimeScript RT Kit (Takara, Japan). LAT expression was measured through qRT-PCR using SYBR Premix Ex Taq (Takara, Japan) on a QuantStudio 5 PCR system (Applied Biosystems, USA). PCR began at 95 °C for 30 s, followed by 40 cycles of 95 °C for 5 s and 60 °C for 30 s. Primer details are in Supplementary Table 1. Gene expression was calculated by the $2^{-\Delta\Delta Ct}$ method, normalized using GAPDH. Each experiment was repeated three times.

Cultivation and validation of cell lines

Cell lines employed in this study were authenticated through short

tandem repeat (STR) profiling. Regular mycoplasma tests were performed to ensure culture purity. Cells were grown in Dulbecco's Modified Eagle Medium (DMEM; Gibco, USA) containing 10% fetal bovine serum (FBS; Gibco, USA) and 1% penicillin-streptomycin solution (Gibco, USA). Cultivation occurred at 37 °C in a humidified incubator with a 5% CO_2 atmosphere. Upon reaching ~80% confluence, total RNA was isolated from cultured cells for subsequent analyses. Quantitative real-time PCR (qRT-PCR) was then conducted to evaluate LAT expression, adhering to the qRT-PCR procedure described for tissue samples. Each experiment, including RNA extraction and qRT-PCR assays, was independently repeated three times to confirm consistency and reproducibility.

Procedure for siRNA transfection

Small interfering RNAs (siRNAs) against LAT and control siRNAs were made by TsingKe Biotechnology. Each siRNA was dissolved in nuclease-free water to a final concentration of 10 μM. Colorectal cancer cells (LS 180 and HT-29) were seeded into 6-well plates (2×10^5 cells/well) and grown overnight at 37 °C with 5% CO_2 to attach properly. For transfection, siRNAs (50 nM) were mixed with Lipofectamine 3000 (5 μL; Invitrogen, USA) in Opti-MEM medium (Gibco, USA) as directed by the manufacturer. The mixture stood at room temperature for 15 min and was then added to the cells. After 6 h, the mixture was replaced with fresh complete DMEM medium. Cells were collected 48 h later, and RNA was extracted using TRIzol reagent as previously described. qRT-PCR measured LAT knockdown, normalizing to GAPDH. LAT expression was reduced by over 70% compared to control ($p < 0.01$, Student's *t*-test). Each experiment was repeated three times for accuracy.

Cell proliferation evaluation using CCK-8 assay

Cell proliferation was evaluated using the Cell Counting Kit-8 (CCK-8; Dojindo, Japan). Cells were seeded into 96-well plates at 3000 cells per well, followed by siRNA transfection as previously described. Proliferation rates were measured at four intervals: 24, 48, 72-, and 96-h post-transfection. At each time point, 10 μL of CCK-8 reagent was added directly to each well. Plates were incubated for 2 h at 37 °C in a humidified incubator with 5% CO_2. Absorbance was measured at 450 nm using a Synergy H1 microplate reader (BioTek, USA). Five technical replicates per condition ensured reliability and reproducibility. Cell proliferation results were presented as fold-change in absorbance relative to baseline (0-hour) measurements obtained immediately after initial cell seeding.

Measurement of cell apoptosis using flow cytometry

Cell apoptosis was assessed at 48 h post-siRNA transfection as previously detailed. Cells were collected and gently washed twice using phosphate-buffered saline (PBS). Apoptosis was determined by utilizing an Annexin V-FITC/Propidium Iodide (PI) apoptosis detection kit (BD Biosciences, USA), strictly adhering to the manufacturer's instructions. Stained cells were immediately analyzed via flow cytometry on a BD FACSVerse instrument. Early apoptotic cells were characterized as Annexin V-positive and PI-negative (Annexin V+/PI−), while late apoptotic cells were identified as Annexin V-positive and PI-positive (Annexin V+/PI+). Flow cytometry data were processed and analyzed using FlowJo software (version 10.0).

Evaluation of cell migration and invasion

Cell migration ability was examined using Transwell chambers (Corning, USA) containing membranes with 8 μm pore size. Cells were collected and resuspended at 5×10^4 cells per chamber in serum-free medium, then seeded into the upper compartment. The lower chambers contained medium supplemented with 10% fetal bovine serum (FBS) as a chemoattractant. Cells were incubated for 24 h at 37 °C to facilitate migration. Following incubation, non-migrated cells on the membrane's upper surface were gently removed using cotton swabs. Migrated cells attached to the underside were fixed with 4% paraformaldehyde and stained using 0.1% crystal violet. Migrated cells were counted microscopically in five random visual fields per Transwell chamber. Cell invasion was assessed similarly, with one modification: membranes were pre-coated with Matrigel (BD Biosciences, USA), diluted 1:8 in DMEM, and incubated at 37 °C for 4 h before cell seeding. Incubation, fixation, staining, and counting procedures for the invasion assay matched those described for migration analysis.

Western blotting analysis

Proteins from cells were collected using RIPA buffer (Beyotime, China) with protease inhibitors (Roche). Protein amounts were measured using a BCA Protein Assay Kit (Pierce, Thermo Fisher Scientific). Samples with equal amounts of protein (30 μg each) were separated using 10% SDS-PAGE gel. Then, proteins were moved onto PVDF membranes (Millipore). Membranes were blocked in 5% skim milk with TBST (0.1% Tween-20) for 1 h at room temperature. They were next incubated overnight at 4 °C with primary antibodies: cleaved Caspase-3 (1:1000, #9664), E-cadherin (1:2000, #3195), Bcl-2 (1:1000, #15,071), Vimentin (1:1000, #5741), and β-actin (1:5000, #4970), all from Cell Signaling Technology. β-actin was used as a control for equal loading. After washing, membranes were incubated with secondary antibodies linked to HRP (1:5000, Cell Signaling Technology) for 1 h at room temperature. Protein bands were detected using ECL solution (Millipore), and their intensities were measured with ImageJ software.

Statistical analysis

Statistical analyses and data visualization were conducted using R software (version 4.1.3). Pearson's correlation analysis assessed relationships between continuous variables. Comparisons of categorical variables across groups employed chi-square tests. Student's *t*-test or Wilcoxon rank-sum test evaluated differences between two independent groups based on data distribution. Survival analyses utilized the survminer R package to identify optimal cut-off points for patient group stratification. Kaplan–Meier survival curves and Cox proportional hazards models were subsequently constructed using the survival R package.

Results

Single-cell RNA sequencing analysis of CRC

After applying rigorous quality control and dimensionality reduction techniques, a total of 18,741 individual cells underwent detailed analysis. These cells were categorized into six major types: epithelial cells, fibroblasts, endothelial cells, T/NK lymphocytes, B lymphocytes, and myeloid cells, based on known markers specific to each cell type (Fig. 1A). Significant differences in the distribution of these cell populations were detected across patient samples, highlighting individual variation (Fig. 1B). The accuracy of cell-type identification was confirmed by assessing the expression patterns of well-established marker genes for each cell group (Fig. 1C).

Single cell CNV analysis

Single-cell genomic changes were examined using InferCNV software, employing normal NKT cells as the reference standard. Distinct CNVs were prominently detected in epithelial cell populations (Fig. 2A).

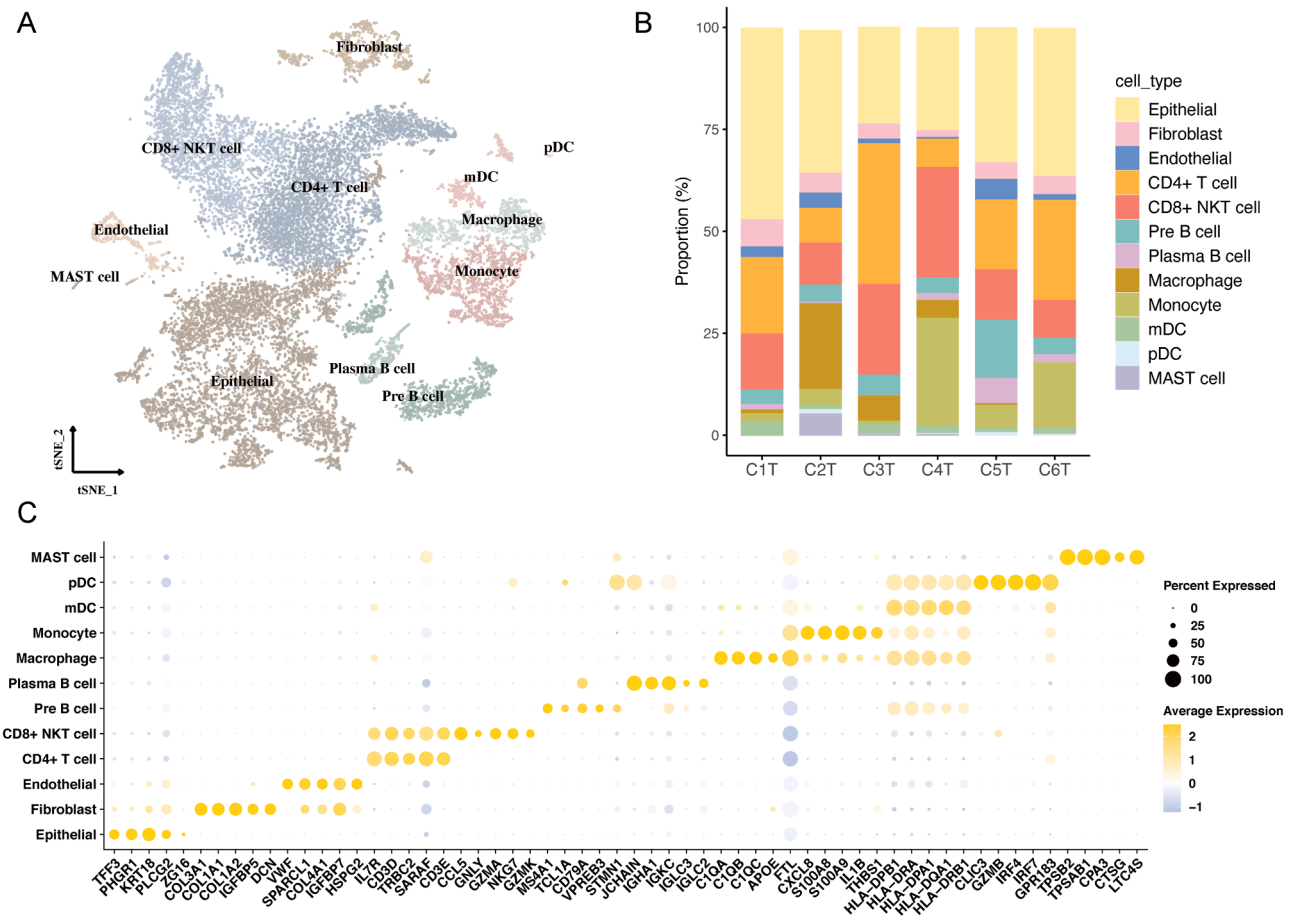


Fig. 1. Cellular diversity in CRC. (A) tSNE plot visualizes diverse single-cell populations in CRC. (B) Bar graphs illustrate patient-specific cellular compositions. (C) Bubble plots highlight marker gene expressions for cell types.

Clustering analysis of T_NK cells

Further classification of T_NK lymphocytes resulted in four separate groups: cluster C0 (CD4+ T cells expressing KLF2), cluster C1 (CD8+ NKT cells expressing CCL4), cluster C2 (CD8+ NKT cells expressing MT2A), and cluster C3 (CD4+ T cells expressing TNFRSF4). The proportions of these clusters differed notably according to patient age and specific cell-cycle stages (Figs. 3A–F). Additionally, we evaluated cluster-specific variations in CNV levels, total UMI counts, and distinct phases of the cell cycle (G2M and S phases) (Figs. 3G–M).

Analysis of intercellular communication

CellChat software was applied to explore signaling interactions between T_NK cells and epithelial cells. Communication patterns involving signaling from T_NK cells to epithelial cells, as well as from epithelial cells back to T_NK cells, are depicted in Figs. 4A and 4B, respectively. Specific ligand-receptor interaction details are further visualized using bubble plots (Figs. 4C–D).

Analysis of cellular differentiation trajectories

Cell differentiation was assessed using CytoTRACE, which indicated that T_NK cells in cluster C3 exhibited lower differentiation potential, whereas clusters C0 and C1 displayed higher differentiation potential (Figs. 5A–B). These findings were further supported by Monocle pseudotime trajectory analysis, which demonstrated that cells from the C3 cluster were predominantly located at the terminal stage of differentiation, while cells from cluster C0 were enriched at the early stage of the differentiation trajectory (Figs. 5C–G). Consistently, the dynamic expression patterns of differentiation-associated marker genes along the pseudotime trajectory further corroborated these observations (Fig. 5H).

Inflammatory response scoring within T_NK cell clusters

We calculated inflammatory response scores across T_NK cell clusters using the GSVA and ssGSEA computational methods. Results indicated notably elevated inflammatory scores for cluster C3 compared to the other clusters (Figs. 6A–D). Functional enrichment analyses revealed pathways associated prominently with T-cell activation and disease progression in cluster C3 relative to the remaining clusters (Figs. 6E–F). Moreover, comparisons among hallmark pathways highlighted additional pathway-level differences between clusters (Figs. 6G–H).

Transcription factor analysis

We analyzed transcription factors (TFs) and grouped them into three clusters (M1, M2, and M3) using the CSI matrix dataset. Fig. 7A illustrates the representative transcription factors, their corresponding binding motifs, and their related cell populations. Furthermore, we computed Regulatory Specificity Scores (RSS) to assess the activity of these transcription factors within each T_NK cell cluster (Fig. 7B).

Prognostic modeling using T_NK cluster C3 and aging markers

We conducted survival analyses using univariate Cox regression on genes specifically associated with the C3 T_NK cluster and those common to Young and old patient groups (Fig. 8A). Subsequently, a

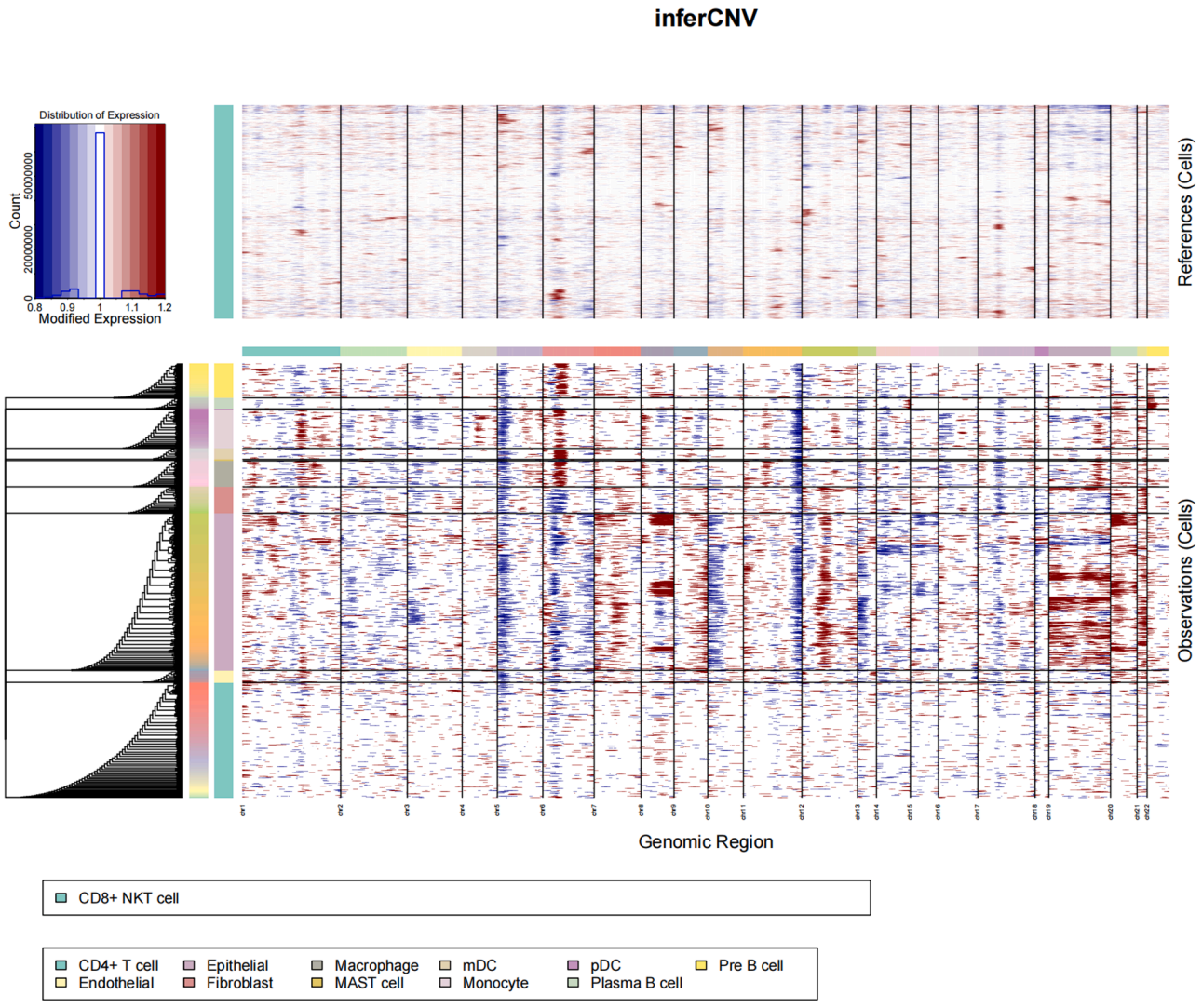


Fig. 2. Genomic variation via single-cell CNV. (A) CNV heatmap shows genomic changes using NKT cells as baseline.

prognostic model was constructed by applying LASSO—Cox regression analysis (Fig. 8B). The selected genes, their regression coefficients, and corresponding expression patterns were clearly presented (Figs. 8C–D). Predictive performance was evaluated through survival probability analyses and ROC curves, which confirmed the robustness of the prognostic model (Figs. 8E–F). Additionally, we examined the individual prognostic impact of each gene included in the model (Fig. 8G).

Analysis of immune infiltration and tumor mutation burden (TMB)

Expression patterns of prognostic genes in CRC samples were analyzed, and immune infiltration was predicted through ESTIMATE, CIBERSORT, and Xcell computational tools. These predictions are presented as heatmaps in Fig. 9A. Correlation analyses between prognostic gene expression, patient risk scores, and immune checkpoint markers revealed clear differences between patient subgroups (Figs. 9B–C). Significant variations in immune cell proportions identified by CIBERSORT were illustrated using bar charts and box plots (Figs. 9D–E). Additionally, relationships among various immune cell subpopulations were explored and visualized (Figs. 9F–G). Lastly, differences in tumor mutation burden (TMB) were assessed across groups, with correlations between TMB, risk scores, and patient survival clearly demonstrated (Figs. 9H–J).

Functional validation of LAT in CRC

The functional importance of LAT, identified as a significant prognostic marker, was verified experimentally in CRC. Initially, we compared LAT mRNA expression levels between CRC tumor tissues and adjacent normal samples, observing substantially elevated levels within tumors (Fig. 10A). Further analysis of various CRC cell lines identified COLO 201 and SW620 cells as those with the highest LAT expression; these cell lines were therefore selected for subsequent experiments (Fig. 10B). Treatment of these cells with siRNA targeting LAT resulted in a notable reduction of LAT expression (Fig. 10C). Cell proliferation assays (CCK-8 method) revealed a marked decrease in growth rates for cells where LAT expression was inhibited, suggesting that LAT positively regulates CRC cell proliferation (Figs. 10D–E). Additionally, apoptosis assessments performed using flow cytometry showed significantly elevated apoptosis rates in LAT-depleted COLO 201 cells compared to control cells (Figs. 10F–G). Migration and invasion potential were evaluated through transwell assays, demonstrating reduced cellular migratory and invasive capabilities following LAT knockdown, thus supporting LAT's role in promoting aggressive cellular behavior in CRC (Figs. 10H–I).

Finally, western blot analysis validated these functional observations. Following LAT knockdown, expression levels of pro-apoptotic markers (cleaved caspase-3) and epithelial marker proteins (E-

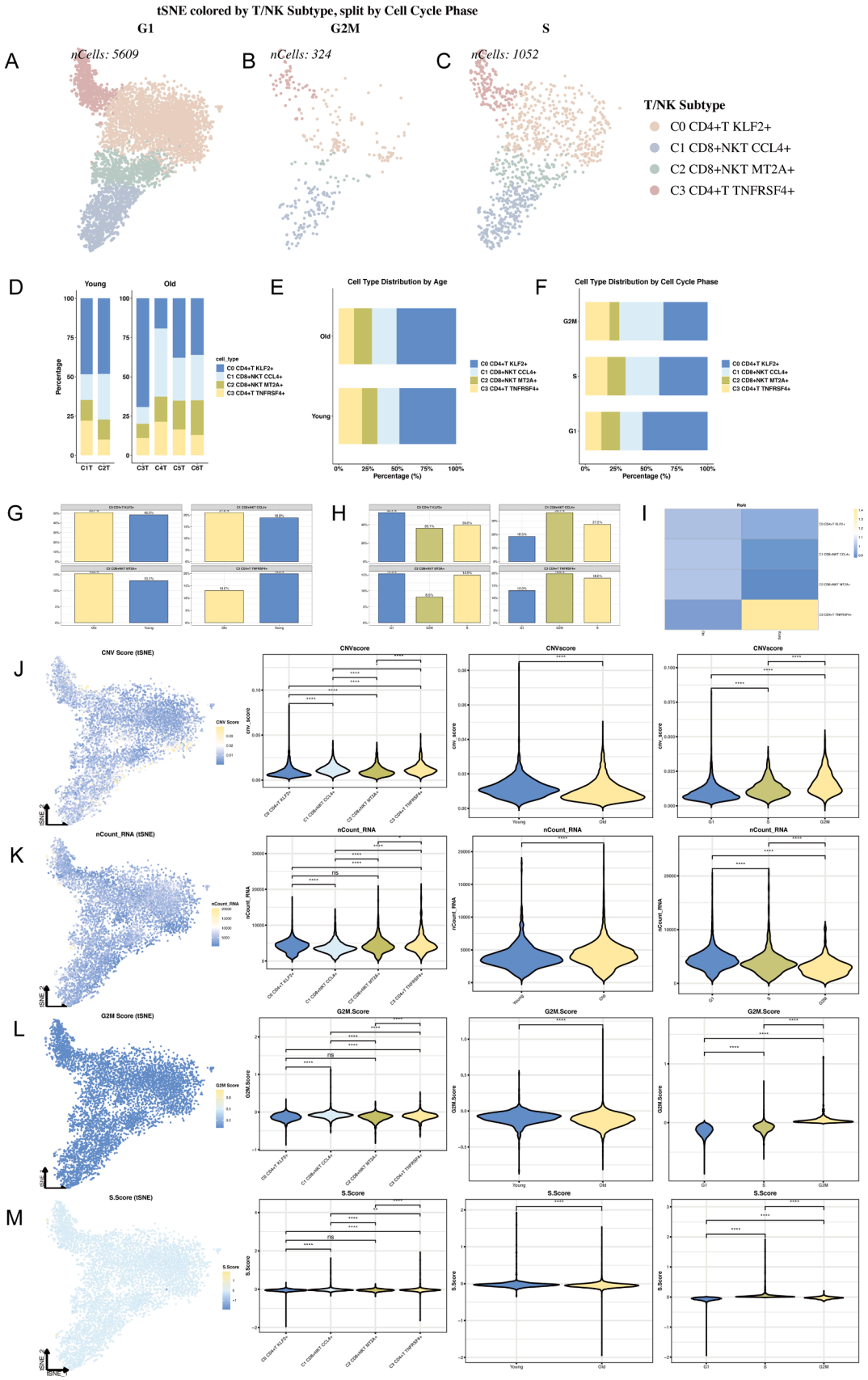


Fig. 3. Characterization of T_NK cell subtypes. (A-C) tSNE plots depict T_NK cell clusters by patient, age, and cycle. (D-H) Bar charts show T_NK cluster proportions across variables. (I) Ro/e ratios indicate subtype tissue preferences. (J-M) tSNE and violin plots display CNV, RNA counts, and cycle phases.

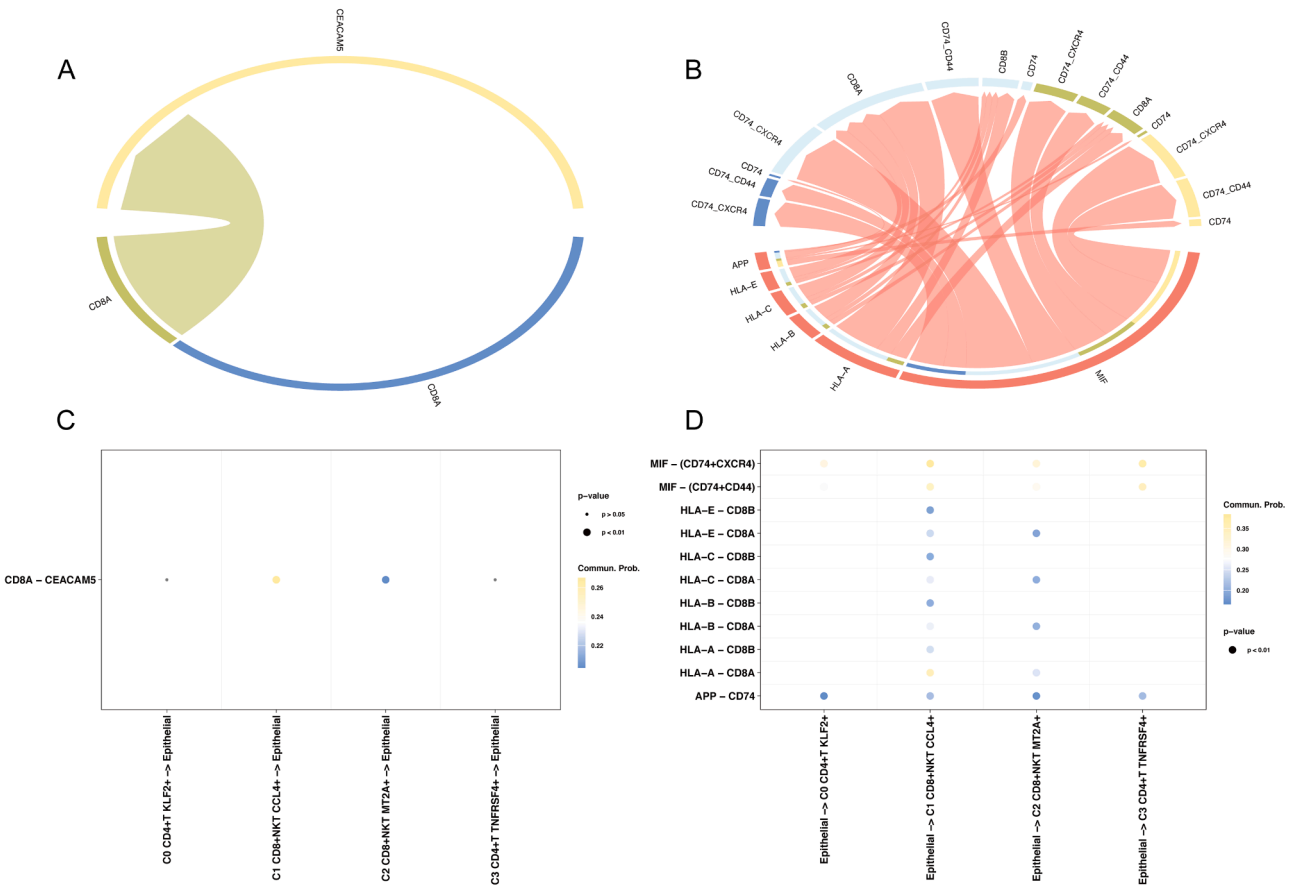


Fig. 4. Interactions between T_NK and epithelial cells. (A, B) Chord diagrams illustrate ligand-receptor interactions. (C, D) Bubble plots summarize key molecular interactions.

cadherin) increased, while anti-apoptotic proteins (Bcl-2) and mesenchymal markers (Vimentin) decreased notably (Figs. 10J–K).

Discussion

The current study offers a detailed single-cell characterization of CRC, providing essential insights into the diverse cellular populations, their differentiation pathways, and intercellular interactions within the tumor microenvironment [25,26]. Utilizing cutting-edge analytical methods such as single-cell RNA sequencing (scRNA-seq), CytoTRACE differentiation analysis, pseudotime trajectory analysis, CNV assessment, and transcription factor profiling, this study significantly enhances our understanding of CRC complexity and identifies potential targets for therapy [27,28]. Here, we discuss our primary findings in relation to existing literature, emphasize the study's strengths and limitations, and outline future research directions. Our analysis confirmed six major cellular populations within CRC tissues: epithelial, fibroblast, endothelial, T/NK, B lymphocytes, and myeloid cells, corroborating earlier single-cell investigations that similarly described CRC's diverse cellular composition [29,30]. For example, Li and colleagues (2017) reported comparable cellular diversity in CRC using scRNA-seq [31,32]. However, our research further advances the understanding of the CRC immune landscape through more detailed subtype annotations of immune cells, achieved via automated *Sc*-Type classification, highlighting novel immune cellular subpopulations not previously emphasized. The CNV analysis highlighted significant genomic instability, particularly within epithelial cells. This aligns closely with established knowledge regarding CRC's genomic heterogeneity, previously described by Peng et al. (2019), who reported similar genomic instability patterns in CRC epithelial populations [33,34]. Our unique methodology of employing NKT cells as the reference group during CNV evaluation enhanced the specificity of detecting malignant genomic aberrations, contributing valuable insights into the clonal dynamics of CRC [35,36]. We conducted in-depth differentiation analyses of T_NK cells using CytoTRACE and monocle pseudotime algorithms, identifying the C3 cluster as the most differentiated subtype, situated early in pseudotime. This finding is supported by analogous studies, such as Guo et al. (2018), which observed similar differentiation states in T cells from other cancer types [37,38]. The identification of cluster C3 as highly differentiated yet positioned at the pseudotime initiation point suggests its crucial role in priming antitumor immune responses, a hypothesis warranting further experimental validation. Inflammatory response profiling, assessed by GSVA and ssGSEA, demonstrated that T_NK subtype C3 exhibited significantly heightened inflammatory activity. This aligns with existing evidence indicating that highly differentiated immune cells typically participate actively in inflammatory signaling (Thommen and Schumacher, 2018) [39,40]. These findings underscore the functional diversity among T_NK subtypes and emphasize distinct inflammatory roles within the CRC tumor microenvironment. Transcription factor (TF) analysis further categorized these regulatory proteins into three distinct groups (M1, M2, and M3), each associated with specific T_NK cell subsets. These results are consistent with previous research by Zhang et al. (2020), who reported similar transcription factor clustering across various malignancies [41,42]. However, our findings uniquely link specific TF clusters to T_NK cell subsets, enhancing comprehension of the transcriptional regulation underpinning immune cell functionality in CRC. Moreover, we successfully constructed a prognostic model based on markers from C3 T_NK cells and genes differentially expressed with aging, employing robust LASSO+Cox regression methodology. Similar statistical approaches have demonstrated predictive reliability in cancer

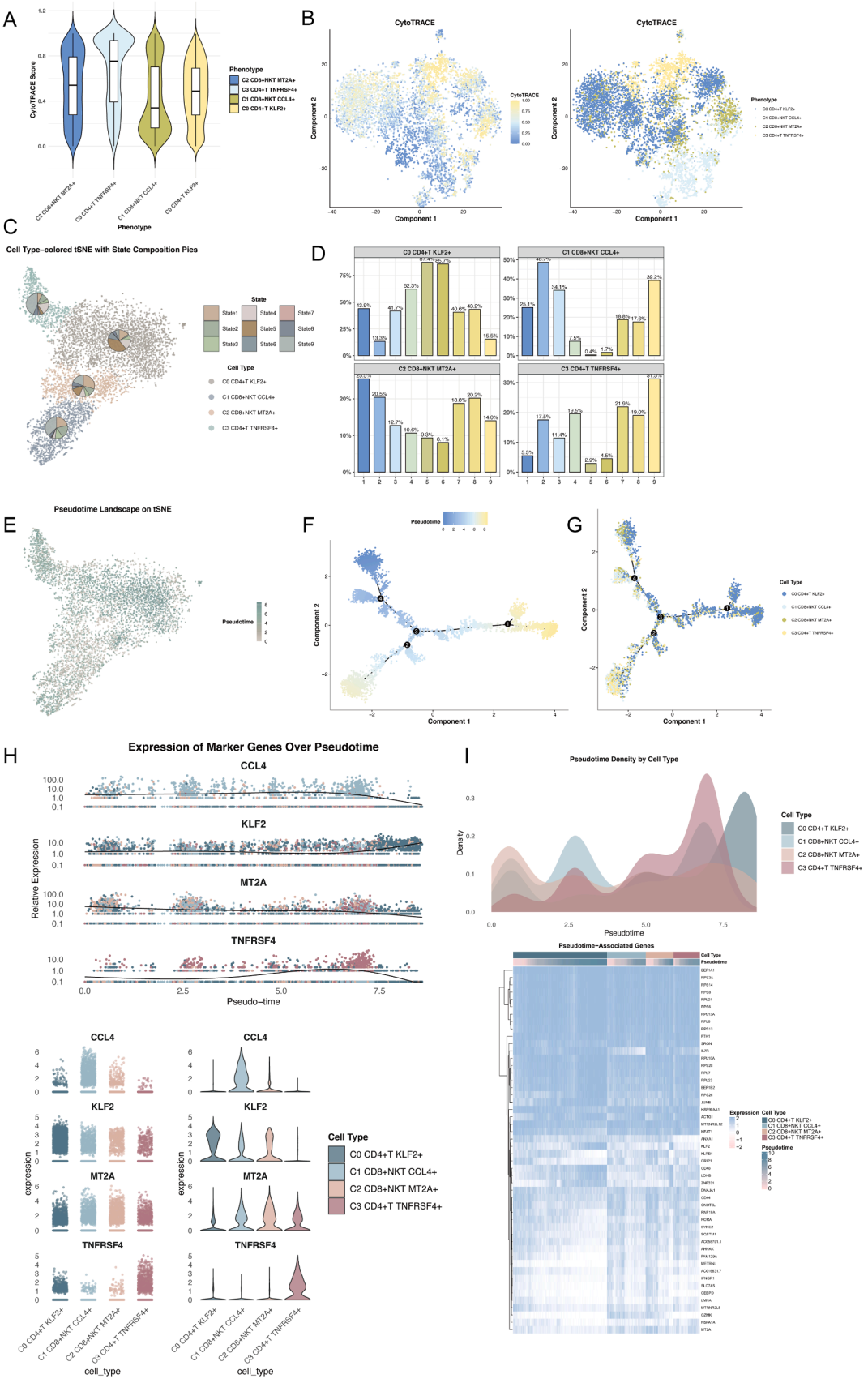


Fig. 5. Stemness and differentiation trajectories. (A, B) CytoTRACE identifies stemness variations across T_NK cells. (C-G) Monocle trajectory shows C1 T_NK cells least differentiated. (H) Heatmap displays marker genes along differentiation trajectories.

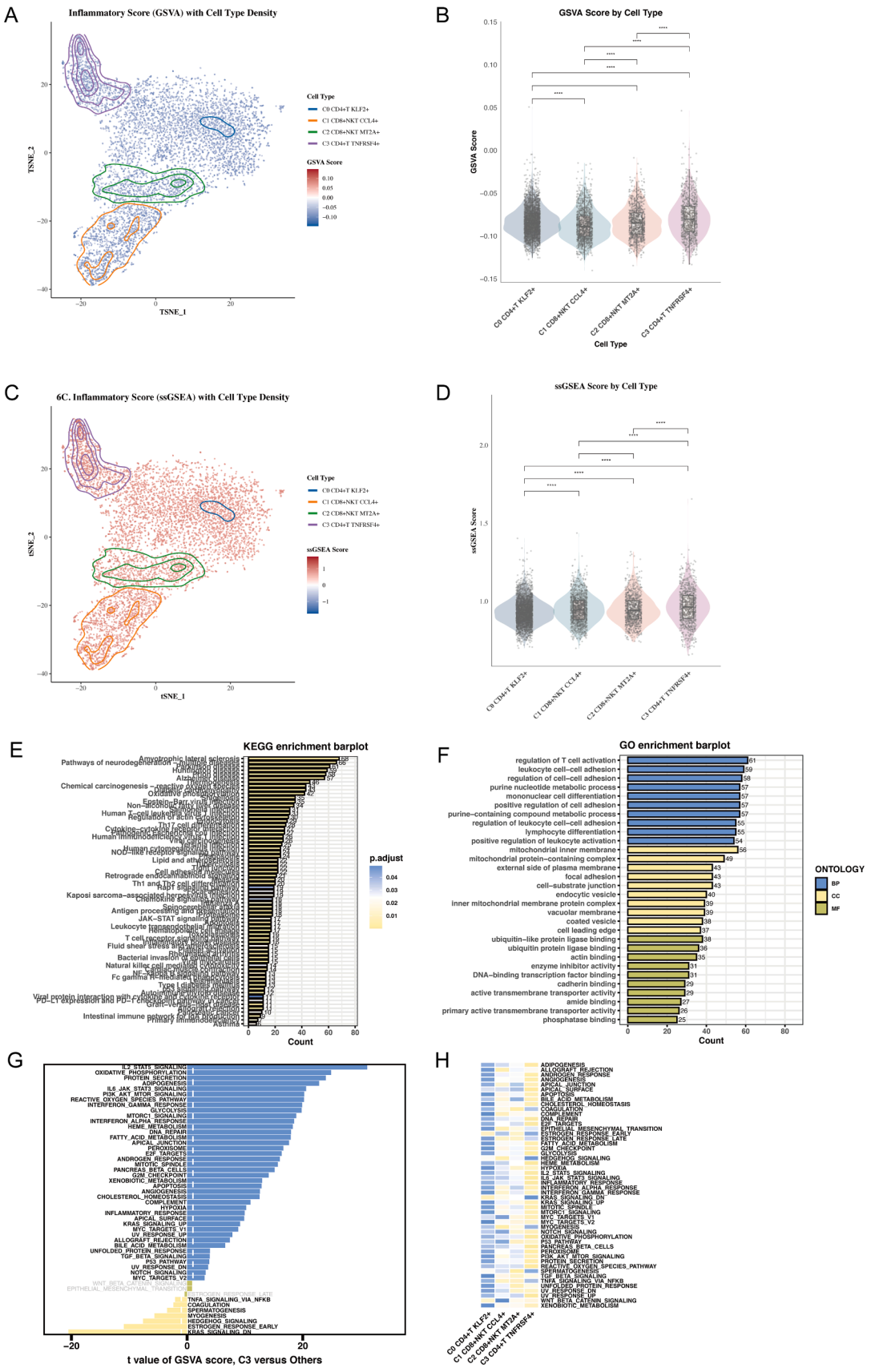


Fig. 6. Inflammatory profiles in T_NK cells. (A, B) GSVA scores highlight immune response differences. (C, D) ssGSEA analysis supports GSVA findings. (E-H) KEGG, GO, GSVA, and GSEA highlight C3 subtype inflammation.

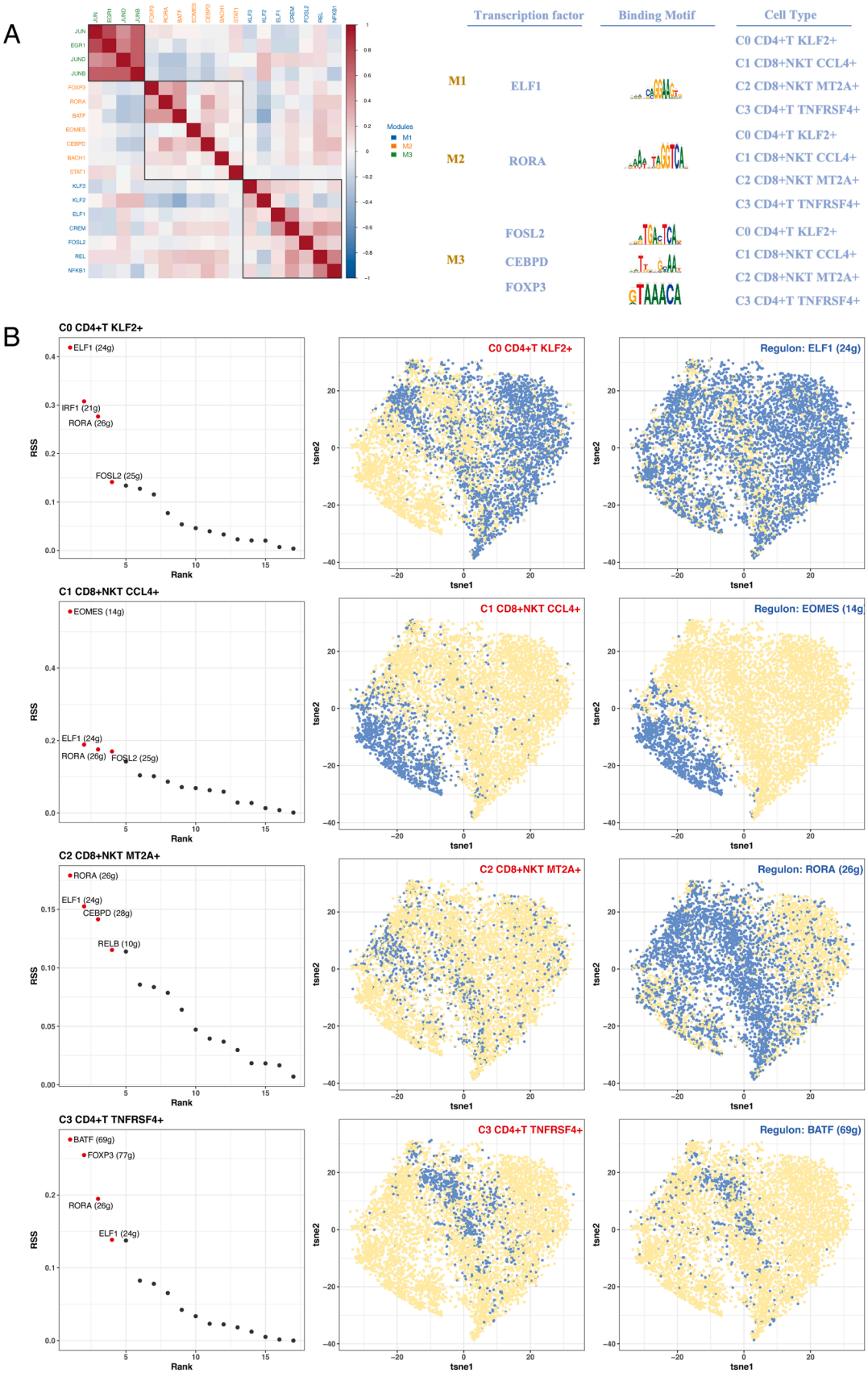


Fig. 7. Transcription factor regulation in T_NK cells. (A) Mapping identifies transcription factors and binding motifs. (B) RSS rankings and tSNE plots show regulatory factor activities.

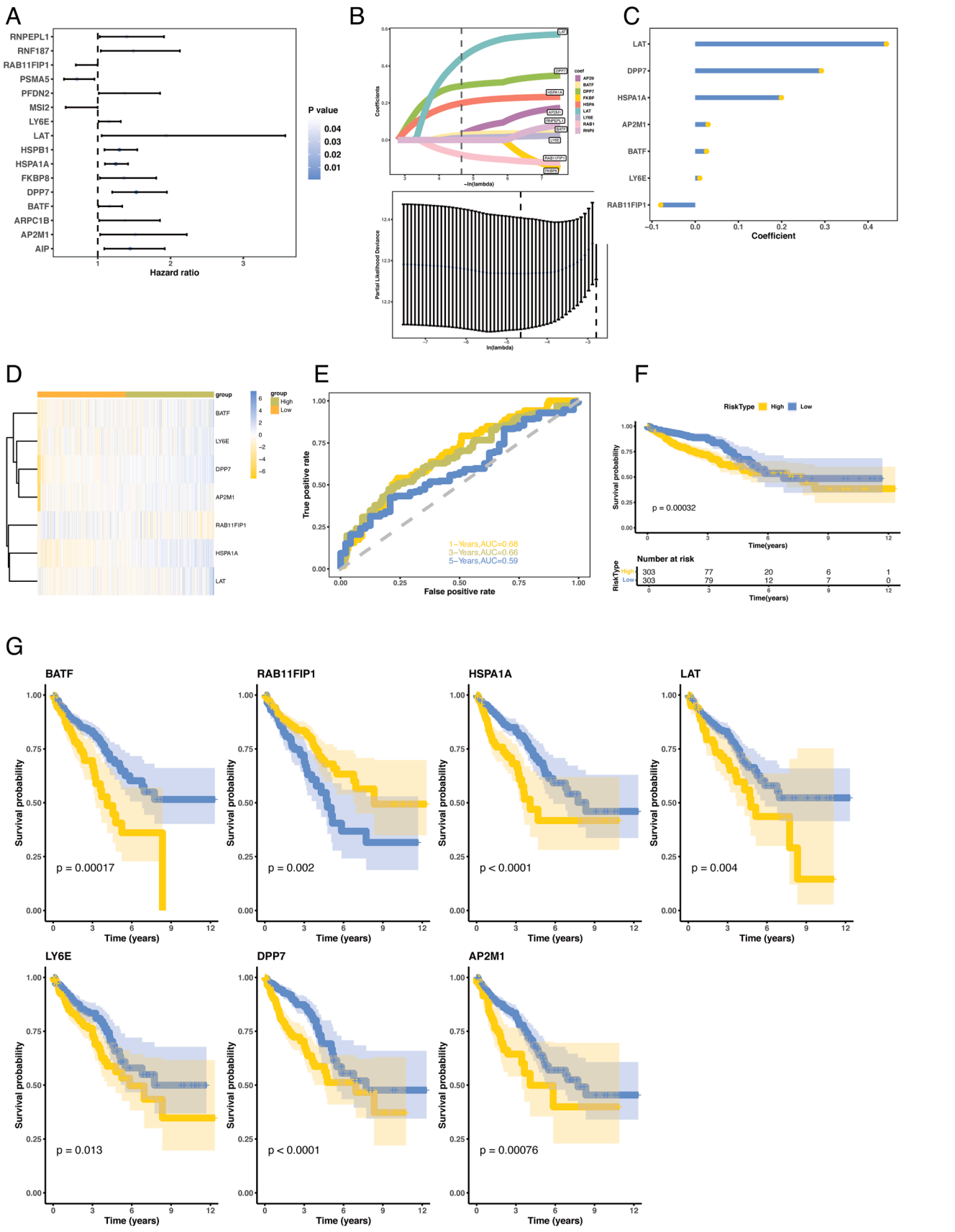


Fig. 8. Prognostic markers in CRC and aging. (A) Forest plot of Cox regression on T_NK and aging genes. (B) LASSO identifies seven robust prognostic markers. (C, D) Dot and heatmap clarify gene contributions and expressions. (E) ROC curves confirm prognostic accuracy over time. (F, G) Kaplan-Meier and gene survival analyses show risk distinctions.

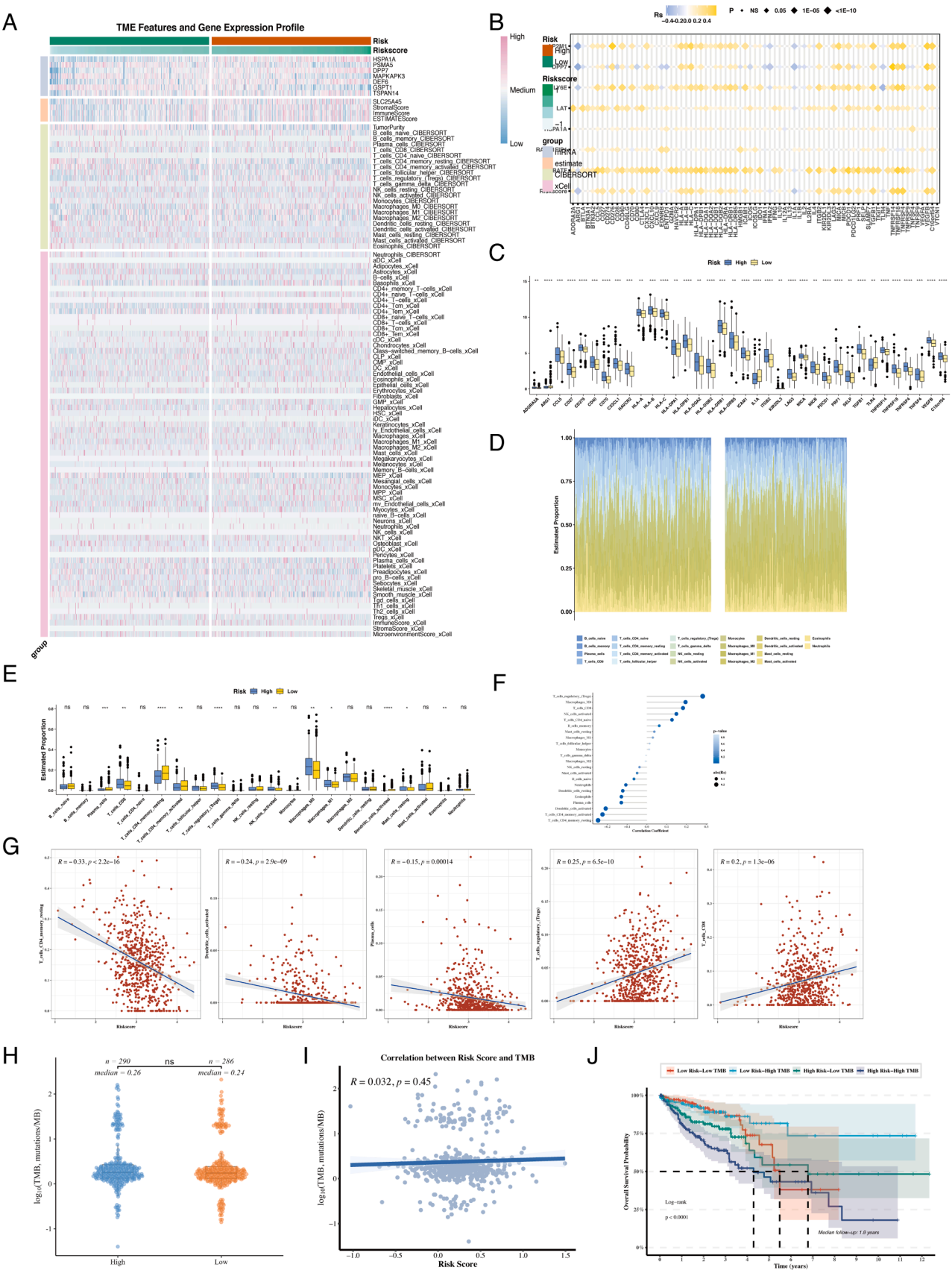


Fig. 9. Immune landscape and mutation burden. (A) Heatmap combines ESTIMATE, CIBERSORT, and Xcell results. (B) Bubble plot correlates immune genes, risk, and prognosis. (C-E) Comparisons depict immune checkpoints across risk groups. (F, G) Correlation plots link immune cells to risk scores. (H-J) TMB analyses differentiate mutation burden by risk.

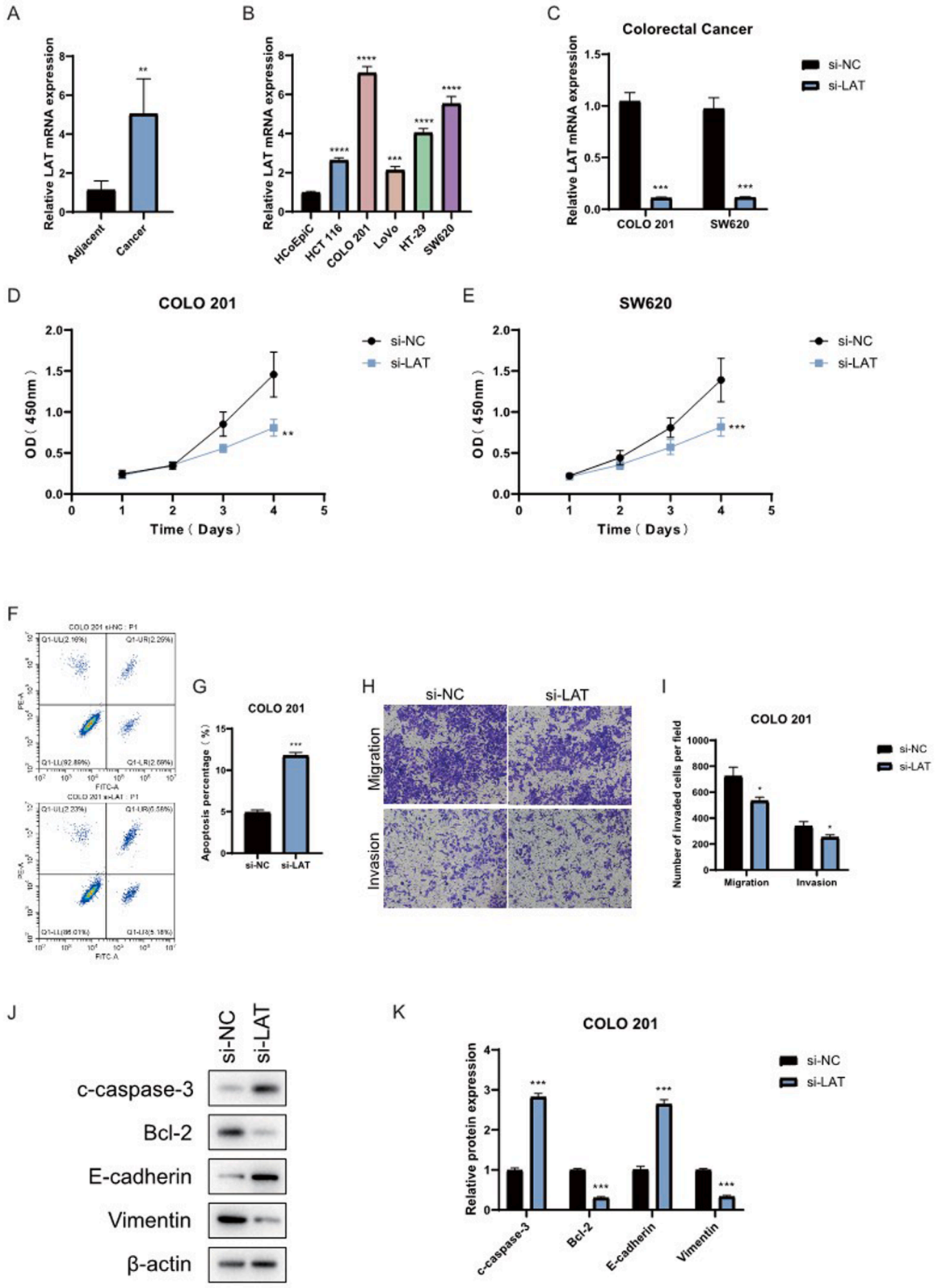


Fig. 10. Functional role of LAT in CRC. (A) mRNA expression differs significantly between CRC and normal tissues. (B) LAT shows altered expression in CRC cell lines versus FHC. (C) qRT-PCR validates effective LAT knockdown. (D-E) CCK-8 assays demonstrate reduced proliferation post-knockdown. (F-G) Flow cytometry shows increased apoptosis after LAT depletion. (H-I) Migration and invasion reduced in LAT-silenced cells. (J-K) Western blots highlight apoptosis and EMT marker changes.

prognostics (Tibshirani, 1997) [43,44]. Our model accurately stratifies CRC patients into different risk profiles based on survival analyses and ROC curve validations. Such a model has substantial implications for personalized medicine, particularly in tailoring immunotherapeutic approaches.

Our comprehensive examination also revealed notable heterogeneity in immune infiltration and TMB. Significant associations between risk scores, immune checkpoint expression, and immune cell composition reinforce the concept that immune contexture is pivotal to CRC biology and patient prognosis [45]. The complex interplay between immune infiltration and genomic characteristics, including TMB, highlights important biomarkers that could predict therapeutic responses, particularly immunotherapy efficacy. We experimentally assessed the functional relevance of LAT, which was identified as a candidate prognostic marker. Consistent with its elevated expression in CRC tissues, LAT knockdown in CRC cells led to reduced cell proliferation, increased apoptosis, and decreased migratory and invasive behaviors. Despite these robust findings, this study has limitations. Primarily, the sample size for clinical validation and functional assays was modest, indicating a need for larger, independent cohorts to validate our prognostic model and LAT's biological role comprehensively. Additionally, integrating spatial transcriptomics could further delineate the precise cellular interactions within the CRC microenvironment, an aspect currently beyond our analysis scope. Future research should aim at broader integrative analyses encompassing multi-omics approaches, including proteomics, metabolomics, and spatial transcriptomics, to unravel CRC's complexity fully. Such comprehensive studies could significantly contribute to discovering novel therapeutic targets and enhancing patient stratification for tailored treatments. In conclusion, this study substantially advances the current understanding of CRC biology by characterizing cellular heterogeneity, differentiation states, inflammatory roles, and regulatory networks within the tumor microenvironment. Our findings provide novel insights with significant implications for improving personalized therapeutic strategies and identifying promising avenues for future CRC research.

Ethical statement

The study was approved by the Ethics Committee of The Second Hospital of Hebei Medical University (Approval number: 2026-R027). Written informed consent was obtained from a legally authorized representative(s) for anonymized patient information to be published in this article.

CRediT authorship contribution statement

Zhenyu Chi: Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Rui Kong:** Data curation, Software, Visualization, Writing – original draft. **Song Wang:** Writing – review & editing, Writing – original draft, Resources, Project administration. **Shaofan Qiu:** Project administration, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known comp financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work was supported by the Chongqing Municipal Science and Technology-Health Joint Research Project (No. 2022MSXM088) and the Medical Science Research Project of Hebei (20220104).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tranon.2026.102761.

Data availability

The single-cell RNA-seq and bulk RNA-seq data analyzed in this study were obtained from public repositories, including GEO and TCGA, with detailed accession numbers provided in the Methods section. All code used for data analysis is available from the corresponding author upon reasonable request.

References

[1] Abedini, F., et al., The prognostic value of PD-1/PD-L1 expression on tumor cells and tumor-infiltrating immune cells in patients with colorectal cancer: a systematic review and meta-analysis protocol. 2021.
[2] B. Al-Zubeidy, et al., Spatial transcriptomics and functional mapping of the suppressive TME in primary breast tumor and breast-to-liver metastases, J. Clin. Oncol. 43 (16_suppl) (2025) e13119. -e13119.
[3] K.M. Alkhalifah, et al., The Use of Hyaluronic Acid in the Treatment of Unilateral Vocal Fold Paralysis: A Systematic Review and Meta-Analysis, Cureus, 2025.
[4] C. Avirup, et al., TMIC-52. decoding metabolic interactions in glioblastoma tumor microenvironment through comprehensive spatial transcriptomic and proteomic analysis, Neuro-Oncol. (Supplement_8) (2024) p. Supplement_8.
[5] A.M. Baker, et al., Abstract 5934: tracking T cell clonal dynamics across time and space in metastatic colorectal cancer, Cancer Res. 83 (7–Sup) (2023) 3.
[6] M.S. Beth Childs, et al., Single-cell insights into cellular heterogeneity and immune dynamics in cutaneous T-cell lymphoma, Blood 144 (Sup1) (2024) 6268.
[7] R. Chen, L. Zou, Combined analysis of single-cell sequencing and bulk transcriptome sequencing reveals new mechanisms for non-healing diabetic foot ulcers, PLoS One 19 (7) (2024).
[8] S.G. Danielli, et al., Abstract A052 project hope: a spatiotemporal single-cell landscape of high-grade gliomas in children, adolescents and young adults, Cancer Res. 84 (17–Sup) (2024) 8.
[9] H.M.U. Farooqi, Single-cell transcriptomic approaches for decoding non-coding RNA mechanisms in colorectal cancer, Noncoding RNA (2025) 11.
[10] H.R. Feng, et al., Unveiling major histocompatibility complex-mediated pan-cancer immune features by integrated single-cell and bulk RNA sequencing, Cancer Lett. 597 (2024) 14.
[11] D. Guo, et al., Single-cell transcriptomic analysis reveals the landscape of epithelial-mesenchymal transition molecular heterogeneity in esophageal squamous cell carcinoma, Cancer Lett. (2024) 587.
[12] H. He, S. Chen, Y. Yu, Zhecai Qian, Youwen Dong, Yaping Song, Yuting Zhong, Caiming Sun, Xiaojuan Cao, Qiqi Li, Shiyao Huang, Weihan Li, Wenxin Zhuang, Mingzhu Yang, Jinxian Wang, Xianming Wang, Jiaqian Wu, Dongfang Wang, Hongyang Wen, Wen, Comprehensive single-cell analysis deciphered microenvironmental dynamics and immune regulator olfactomedin 4 in pathogenesis of gallbladder cancer, Gut: J. Br. Soc. Gastroenterol. 73 (9) (2024) 1529–1542.
[13] E. Horeth, et al., Transcriptomic and single-cell analysis reveals regulatory networks and cellular heterogeneity in mouse primary sjgren's syndrome salivary glands, Front. Immunol. (2021).
[14] J. Huang, et al., Single-cell RNA sequencing in autoimmune diseases: new insights and challenges, Pharmacol. Ther. (2025) 267.
[15] L. Huang, X. Wang. Yin, Single-cell profiling reveals sustained immune infiltration, surveillance, and tumor heterogeneity in infiltrative basal cell carcinoma, J. Invest. Dermatol. 143 (11) (2023) 2283–2294.
[16] H. Jia, et al., Spatial and single-cell transcriptomic analysis reveals fibroblasts dependent immune environment in colorectal cancer, Biofactors 51 (2) (2025).
[17] N. Koutsoubis, et al., Privacy-preserving federated learning and uncertainty quantification in medical imaging, Radiol.: Artif. Intell. (2025).
[18] X. Li, et al., Mapping glioma progression: single-cell RNA sequencing illuminates cell–cell interactions and immune response variability, Discov. Oncol. 16 (1) (2025) 1–16.
[19] T. Liu, et al., Elucidating the role of SLC4A7 in glioma prognosis: a comprehensive approach combining bioinformatics, single-cell analysis, and tissue validation, J. Med. Sci. (1011-4564) 43 (5) (2023).
[20] X. Liu, et al., Single-cell RNA-seq of the stromal vascular fraction of adipose tissue reveals lineage-specific changes in cancer-related lymphedema, bioRxiv (2020).
[21] T. Lyu, et al., Single-cell RNA sequencing reveals the tumor heterogeneity and immunosuppressive microenvironment in urothelial carcinoma, Cancer Sci. 116 (3) (2025).
[22] Lyu, X., et al., Has the digital transformation promoted energy-saving-biased technological progress in China's manufacturing sector? 2025.
[23] K. Mf, et al., Single cell transcriptomic analysis reveals cellular diversity of murine esophageal epithelium and age-associated mitochondrial dysfunction, bioRxiv (2021).
[24] T. Peng, A comprehensive analysis of gene expression and the immune landscape in gastric cancer through single-cell and multi-omics approaches, Discov. Oncol. 15 (1) (2024) 1–15.

[25] C. Wang, et al., Single-cell RNA sequencing reveals the heterogeneity of MYH11+ tumour-associated fibroblasts between left-sided and right-sided colorectal cancer, J. Cell. Mol. Med. 28 (18) (2024) 11.
[26] Wang, P., et al., Chlamydia trachomatis infection and risk of ovarian cancer: a systematic review and meta-analysis. 2025.
[27] S.Z. Wu, A. Swarbrick, Single-cell advances in stromal-leukocyte interactions in cancer, Immunol. Rev. 302 (1) (2021) 286–298.
[28] S.Z. Wu, A. Swarbrick, Single cell advances in stromal leukocyte interactions in cancer, Immunological reviews 302 (1) (2021) 286–298.
[29] D. Xie, et al., Integrative analysis of bulk and single-cell RNA sequencing reveals sphingolipid metabolism and immune landscape in clear cell renal cell carcinoma, Environ. Toxicol. 39 (12) (2024).
[30] Xu, P., et al., scSiameseClu: a siamese clustering framework for interpreting single-cell RNA sequencing data. 2025.
[31] Q.Q. Xu, et al., Single cell-pair proteomics for decoding immune-cancer cell interactions, Adv. Sci. 12 (11) (2025).
[32] H. Yan, et al., Comprehensive analysis of partial methylation domains in colorectal cancer based on single-cell methylation profiles, Brief. Bioinform. (2021).
[33] X. Yang, X. Wang, J. Yang, Single-cell analysis reveals cellular heterogeneity, gene expression profiles, and pathway dynamics in acne vulgaris, Arch. Dermatol. Res. 317 (1) (2025).
[34] Yu, Z., et al., Single cell RNA sequencing of infantile hemangioma reveals cellular heterogeneity. 2020.
[35] H. Yuan, et al., Graph neural networks for knowledge graph construction: research progress, agricultural development potential, and future directions, Smart Agric. 7 (2) (2025) 41–56.
[36] G. Zhang, et al., Combined single cell and spatial transcriptome analysis reveals cellular heterogeneity of hedgehog pathway in gastric cancer, Genes Immun. 25 (6) (2024).
[37] H. Zhang, et al., Deciphering the molecular heterogeneity of soft tissue sarcoma by integrating multiomics and single cell sequence, Int. J. Biochem. Cell Biol. (2025) 185.
[38] Q. Zhang, et al., Integration of single-cell RNA sequencing and bulk RNA transcriptome sequencing reveals a heterogeneous immune landscape and pivotal cell subpopulations associated with colorectal cancer prognosis, Front. Immunol. (2023).
[39] Z. Zhang, et al., Abstract 6883: investigation of tumor-immune interactions in pediatric sarcomas using single-cell multi-omics, Cancer Res. 84 (6–Sup) (2024) 5.
[40] Z. Zhang, Z. Chen, Exploring the causal links between inflammation-related genes and atherosclerosis through mendelian randomization analysis, Med. (Baltim.) 104 (23) (2025) e42584–e42588.
[41] J. Zhao, et al., Single-cell RNA sequencing reveals the heterogeneity of liver-resident immune cells in human, Cell Discov. 6 (1) (2020) 22.
[42] N. Zhao, J. Rosen, Breast cancer heterogeneity through the lens of single-cell analysis and spatial pathologies, Semin. Cancer Biol. (2021).
[43] L. Zhilong, et al., The network structural entropy for single-cell RNA sequencing data during skin aging, Brief. Bioinform. (1) (2025) 1.
[44] L. Zhou, et al., Unveiling Leydig cell heterogeneity and its role in male infertility: a single-cell transcriptomic study of human testicular tissue, Reprod. Biol. 25 (1) (2025).
[45] J. Zhu, et al., Integrative analysis of semaphorins family genes in colorectal cancer: implications for prognosis and immunotherapy, Front. Immunol. (2025).