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.

Integrative Analysis of Sialylation and Amino Acid Metabolism Signatures Reveals Prognostic Drivers and Immunotherapy Responses in Colorectal Cancer

Yuanzi Zhou
Shanghai University of Traditional Chinese Medicine
Xiao Feng
Nanjing University of Chinese Medicine
Tianyao Yu
Nanjing University of Chinese Medicine
Cunyu Zhang
Nanjing University of Chinese Medicine
Nan Yi
Nanjing University of Chinese Medicine
Linglong Zhang
Macau University of Science and Technology
Xindong Yin
Nanjing University of Chinese Medicine
Su Zhou
zhousu@shaphc.org

Shanghai Public Health Clinical Center

Research Article

Keywords: Colorectal cancer, sialylation, amino acid metabolism, immune infiltration, prognostic model

Posted Date: October 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11307983/v1

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

Additional Declarations: No competing interests reported.

Integrative Analysis of Sialylation and Amino Acid Metabolism Signatures Reveals Prognostic Drivers and Immunotherapy Responses in Colorectal Cancer

Yuanzi Zhou[1†], Xiao Feng[2,3†], Tianyao Yu[2,3†], Cunyu Zhang[2,3], Nan Yi[4], Linglong Zhang[5], Xindong Yin[2,3*], Su Zhou[6*]

[1]Shanghai University of Traditional Chinese Medicine, Shanghai, 201203, China.

[2]Department of General Surgery, Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, Jiangsu 210003, China.

[3]Department of General Surgery, Jiangsu Province Hospital of Chinese Medicine, Nanjing, Jiangsu 210003, China.

[4]Yancheng Hospital of Traditional Chinese Medicine Affiliated to Nanjing University of Chinese Medicine, Yancheng, Jiangsu 224001, China.

[5]School of Pharmacy, Faculty of Medicine & State Key Laboratory of Quality Research in Chinese Medicines, Macau University of Science and Technology, Macau, China.

[6]Department of Interventional Medicine, Shanghai Public Health Clinical Center, Fudan University, Shanghai, 201508, China.

*** Correspondence:**

Su Zhou

zhousu@shaphc.org

Xindong Yin

yinxindong1226@sina.com

Abstract

Background: CRC, or colorectal cancer, exhibits substantial biological diversity, with varied metabolic programs and inconsistent benefit from immunotherapy. Pathways involving sialylation together with metabolism of amino acids shape tumor development and regulation of anti-tumor immunity. However, the prognostic significance of this combined metabolic signature for CRC remains clinically undefined in practice.

Methods: We assembled genes related to sialylation plus those linked to metabolism of amino acids from public databases to construct a comprehensive candidate list. Multi-omics profiling integrated bulk RNA-seq, single-cell transcriptome profiling, and spatially resolved transcriptomic data to examine their functions in CRC. Differential expression testing, survival modeling, and consensus-based clustering were used to define metabolic–epigenetic tumor categories and compare clinical outcomes. From these markers, we

derived a multigene risk score centered on six regulators connected with sialylation and amino-acid metabolic pathways. Its prognostic performance was rigorously evaluated within several external colorectal carcinoma cohorts using multiple survival endpoints. We further examined the key regulator PLOD3 in patient specimens and cultured cells using qRT-PCR, immunoblotting, flow cytometry, colony formation, migration, and invasion assays.

Results: We identified six prognosis-related genes (AOC3, CDO1, PHGDH, MAT2B, ASRGL1, and PLOD3) that function as principal metabolic determinants in CRC. For unsupervised clustering, non-negative matrix factorization separated tumors into six molecular subgroups, among which Cluster 3 showed the most unfavorable survival. Single-cell transcriptomics and spatially resolved transcriptomic analyses demonstrated marked metabolic and immune diversity across the tumor microenvironment. These datasets revealed distinct architectures of immune-cell infiltration and region-specific metabolic activity within tumor tissue. The six-gene risk model displayed high prognostic accuracy, outperforming conventional clinicopathologic variables and previously proposed CRC signatures. Patients assigned to the high-risk group exhibited increased tumor mutational burden, pronounced activation of immune pathways, and perturbed immune responses, indicating exploitable therapeutic opportunities. In experimental validation, PLOD3 expression was markedly elevated in peripheral blood and neoplastic tissue from individuals with CRC. Stable depletion of PLOD3 in colorectal cancer cell lines reduced cell proliferation, migratory capacity, invasive potential, and epithelial–to–mesenchymal plasticity. It simultaneously enhanced apoptotic cell death in colorectal cancer cells under in vitro experimental conditions.

Conclusions: Our investigation emphasizes the central contribution of sialylation and amino-acid–related metabolic programs to colorectal cancer progression and immune modulation. The integrative prognostic framework, together with mechanistic dissection of PLOD3, supports refined risk stratification and informs individualized therapeutic planning for patients with CRC.

Keywords: Colorectal cancer, sialylation, amino acid metabolism, immune infiltration, prognostic model.

Introduction

Colorectal cancer (CRC) remains a major health challenge worldwide and contributes substantially to cancer incidence and death, accounting for 6.1% of global cases and 9.4% of cancer-related mortality[1]. Advances in screening methods, surgical strategies, chemotherapy, and immune checkpoint blockade have improved outcomes, yet patients with advanced disease continue to face limited survival prospects[2, 3]. Increasing evidence

highlights the strong influence of epigenetic mechanisms on gene regulation and the tumor microenvironment in CRC[4]. However, interactions between epigenetic changes and metabolic remodeling, and how these processes shape tumor progression and therapeutic escape, are still insufficiently understood.

Among epigenetic processes, sialylation has emerged as an important regulator in cancer biology[5]. This terminal glycosylation event attaches sialic acid residues to diverse glycoconjugates, thereby generating surface sialoglycans that influence cell communication, immune recognition, and metastatic capacity[5-7]. Abnormal sialylation frequently appears in CRC and correlates with aggressive behavior, reduced drug sensitivity, and unfavorable clinical outcomes[8]. For example, increased ST6Gal-I expression enhances CRC cell survival and protects against apoptosis, while sialylated integrin β4 promotes liver metastasis through Src–Erk pathway activation[9, 10]. Sialylated molecules can also mask tumor antigens and foster immune suppression, weakening anti-tumor immunity[11]. Targeting sialylation has shown promise for restoring T cell activity and improving immunotherapy responses[12]. Yet, the upstream pathways controlling aberrant sialylation in CRC remain largely undefined.

Metabolic reprogramming represents another core feature of malignancy, with amino acid metabolism serving essential roles in cancer growth[13]. Amino acids support protein production and regulate energy turnover, redox stability, epigenetic marks, and immune functions[14-16]. CRC cells display elevated uptake and utilization of amino acids such as tryptophan, glutamine, methionine, and arginine to maintain proliferation, immune evasion, and metastatic spread[17]. Altered arginine and branched-chain amino acid (BCAA) pathways have been linked to CRC development and treatment resistance, and several metabolic enzymes are being explored as biomarkers or therapeutic targets[18].

Emerging studies indicate a bidirectional relationship between amino acid metabolism and sialylation. Amino acid levels can shape glycosyltransferase activity and determine precursor pools required for sialoglycan synthesis[19]. Conversely, dysregulated sialylation may influence transporter expression and metabolic signaling that control cellular nutrient processing. Despite these observations, the coordinated regulation of amino acid metabolism and sialylation in CRC remains insufficiently defined. Clarifying this integration may uncover new biological vulnerabilities and inform targeted therapeutic strategies.

To address these gaps, we applied an integrative multi-omics framework combining bulk transcriptomics, single-cell sequencing, and spatial transcriptomics to comprehensively characterize sialylation patterns and amino acid metabolic programs in CRC. We defined distinct epigenetic–metabolic subgroups, generated a prognostic model, and identified central regulatory molecules and potential therapeutic targets. These findings highlight an

interconnected network between metabolic and epigenetic pathways in CRC and offer a foundation for improved prognostication and personalized treatment strategies.

Materials and methods

Data Collection and Preprocessing

RNA-seq profiles for colorectal cancer were obtained from the TCGA dataset (606 samples with matched clinical information). Expression values were transformed to TPM and log2 - scaled before analysis. External validation used GEO cohorts GSE17537 (n=55), GSE39582 (n=579), GSE12945 (n=62), GSE87211 (n=196), GSE17536 (n=177), GSE38832 (n=122), and GSE41258 (n=182). Differential-expression analyses incorporated GSE22598 and GSE110224 (each comprising 17 tumor–normal pairs) and GSE41328 (10 paired samples). Data preprocessing involved two steps: normalization with normalizeBetweenArrays (limma) and batch correction with Combat (sva).

Single-Cell and Spatial Transcriptomics Data

Single-cell RNA-seq data from GSE132465 (23 CRC tumors) were processed in Seurat (R v4.1.3). Low-quality cells were removed using strict filters. Cells with¿ 20% % mitochondrial genes,> 3% blood-cell signatures, <200 or> 6 000 detected genes or UMI counts outside 200 to 50,000 were excluded. Data were normalized, and 2 000 variable genes were selected using NormalizeData, FindVariableFeatures, and ScaleData. Dimensionality reduction used UMAP and tSNE, followed by Louvain clustering. Cluster-specific genes were detected with FindAllMarkers using $p < 0.05$, $log2FC > 0.25$, and expression frequency > 0.1. Spatial transcriptomics data from GSE225857 (four primary CRC tissues) were processed with SpaceRanger. Data were normalized and corrected using SCTtransform. Cell-type assignments for each spatial spot were predicted using CARD, and results were visualized in Seurat. Gene-set activity in each spot was scored through AUCell.

Cell type annotation and visualization

Cell identities were assigned using canonical lineage markers. Epithelial cells expressed *EPCAM*, *KRT18/19*, and *CDH1*; fibroblasts showed *DCN*, *THY1*, *COL1A1/2*; endothelial cells exhibited *PECAM1*, *CLDN5*, *FLT1*, and *RAMP2*. T-cell subsets were defined by *CD3D/E/G* and *TRAC*, NK cells by *NKG7*, *GNLY*, *NCAM1*, and *KLRD1*. B cells expressed *CD79A*, *IGHM/IGHG3/IGHA2*, and plasma cells were enriched for *JCHAIN*. Mast cells carried *KIT*, *MS4A2*, and *GATA2*. UMAP/t-SNE embeddings and bar-plot/heatmap visualizations summarized transcriptional heterogeneity and cell-type composition.

Gene Set Collection

Sialylation-related genes were obtained from published literature (PMC9623420; Sialylation markers.txt). Amino acid metabolism genes were sourced from KEGG (hsa00001.format.txt). These curated sets were used for downstream correlation and prognostic analyses.

Prognostic Gene Screening and Consensus Clustering

Genes linked to sialylation or amino acid metabolism were identified by correlating their expression with the corresponding pathway gene sets across TCGA and GEO datasets, retaining genes with $|R| > 0.3$ and $p < 0.05$. Differential expression between tumor and normal samples was assessed using $p < 0.05$ and $\log_2FC > 1$. Genes that satisfied both correlation and differential-expression thresholds were designated as candidate prognostic markers. Six prognostic genes *PHGDH*, *AOC3*, *PLOD3*, *MAT2B*, *ASRGL1*, *CDO1* were selected for clustering. Nonnegative matrix factorization (NMF) was applied to TCGA-CRC samples to define metabolic–sialylation subtypes. The cophenetic coefficient guided the selection of the optimal number of clusters, yielding stable risk groups for subsequent analyses.

SNV Analysis

Tumor mutation burden (TMB) was derived from TCGA somatic mutation profiles using the maftools package. Differences in TMB between risk groups were evaluated with the Wilcoxon rank-sum test, applying a significance threshold of $p < 0.05$.

Cell–Cell Communication Analysis

Cell–cell communication was assessed using CellChat. The normalized expression matrix was preprocessed with identifyOverExpressedGenes, identifyOverExpressedInteractions, and projectData to detect active ligand–receptor pairs. Communication probabilities were computed using computeCommunProb, filtered with filterCommunication, and mapped to signaling pathways via computeCommunProbPathway. Interaction strengths were summarized using aggregateNet, generating global and pathway-specific communication networks.

Differential expression and enrichment

Differential gene expression between tumor and matched adjacent normal tissues was evaluated with limma. Genes with an adjusted p value < 0.05 and an absolute fold change > 1.2 were considered differentially expressed. Functional enrichment analyses were conducted using clusterProffler. Gene set enrichment analysis (GSEA) incorporated HALLMARK, GOBP, and KEGG collections from MSigDB, and results were visualized with enrichplot.

Prognostic Risk Score Generation

A prognostic model was built by testing 101 machine-learning approaches (e.g., RSF, Elastic Net, Lasso, Ridge, CoxBoost, PLSR-Cox, SuperPC, GBM, survival-SVM). Prognostic genes identified by univariate Cox analysis across eight cohorts, including TCGA-CRC, were used to train models under LOOCV. Performance was validated in seven external datasets using Harrell's C-index.

Immune Response Prediction and IPS Analysis

Immune-therapy response was estimated by calculating risk scores in PRJEB23709 and GSE91061 melanoma cohorts. For TCGA-CRC, immune responsiveness was assessed using TIDE, and IPS values from TCIA were compared across risk groups. Risk scores were correlated with immune-checkpoint gene expression (*HAVCR1, CD28, ICOS, TNFRSF9, IL2RB, CD27*).

Tumor immune infiltration analysis

Immune infiltration in the tumor microenvironment (TME) was quantified with the IOBR framework, which integrates CIBERSORT, TIMER, MCPcounter, and Estimate. Heatmaps were generated to depict immune cell composition in the TME. The Estimate algorithm was further used to compute stromal, immune, and tumor scores, enabling comparison of TME features between different risk groups.

Drug sensitivity analysis

Drug sensitivity was predicted using oncoPredict, and groupwise IC50 differences were tested with the Wilcoxon rank-sum test.

Cell Lines and Culture Conditions

Human CRC cell lines HCT116 (Procell, CL-0096) and Caco-2 (Procell, CL-0050) were maintained under standard culture conditions. HCT116 cells were grown in McCoy's 5A medium (Procell, PM150710) containing 10% FBS and 1% penicillin–streptomycin. Caco-2 cells were cultured in MEM supplemented with NEAA (Procell, PM150410), 20% FBS, and 1% penicillin–streptomycin. All cells were incubated at 37 °C, 5% CO2, and high humidity, screened regularly for mycoplasma, and authenticated by STR profiling.

shRNA-Mediated Silencing

Lentiviral shRNA was used to generate stable knockdown cells. HCT116 or Caco-2 cells (2×10^5) were plated in six-well dishes and allowed to adhere for 24 h. Cells were then infected with concentrated lentivirus (MOI 5^{-10}) plus 8 μg ml $^{-1}$ polybrene and incubated overnight. The viral medium was replaced with fresh complete medium, and cells were cultured for 36–48 h to permit transgene expression. Stable knockdown lines were obtained by puromycin selection (2 μg ml $^{-1}$) for 7–10 days.

Sample collection and ethics

Institutional Review Board approval was obtained from Affiliated Hospital of Nanjing University of Chinese Medicine, and all procedures complied with the Declaration of Helsinki. Written informed consent was secured from every participant. Peripheral blood was collected from six CRC patients and six matched healthy individuals. During surgery, tumor specimens and paired adjacent normal tissues (collected ≥5 cm from the tumor margin) were harvested.

RNA extraction and quantitative RT–PCR

Total RNA was extracted from PBMCs using TRIzol LS, and RNA concentration/purity were assessed with a NanoDrop spectrophotometer. Next, 1 μg of RNA was reverse-transcribed to cDNA using the PrimeScript RT Kit. qRT–PCR was performed in triplicate on a QuantStudio 5 with TB Green Premix Ex Taq II under the following program: 95 °C for 30 s, then 40 cycles of 95 °C for 5 s and 60 °C for 30 s. PLOD3 expression was normalized to GAPDH.

Tissue protein extraction

Tumor and adjacent normal tissues were snap-frozen and stored at −80 ◦C. Approximately 30 mg of tissue was homogenized in RIPA buffer supplemented with protease inhibitors. The lysates were centrifuged at 14,000 × g for 20 min at 4 °C, and the supernatants were collected. Protein concentrations were quantified using a BCA assay.

Western blot analysis

At 90% confluence, cells were rinsed with ice-cold PBS and lysed on ice in RIPA buffer supplemented with protease inhibitors. The lysates were clarified by centrifugation (14,000 × g, 30 min, 4 °C), and protein concentrations were determined using a BCA assay. Equal amounts of protein (25 μg) were resolved by 10% SDS–PAGE and transferred onto PVDF membranes. Membranes were blocked for 1 h in 5% milk/TBST and then incubated overnight at 4 °C with primary antibodies against PLOD3, BAX, BCL2, Vimentin, E-cadherin, N-cadherin, and β-actin. After washing, membranes were incubated with an HRP-conjugated secondary antibody for 1 h. Signals were developed using ECL, captured with a ChemiDoc MP system, and band intensities were quantified in Image Lab.

Data Collection and Preprocessing

Apoptosis was assessed using the Annexin V-FITC/PI kit (Beyotime). Parental and PLOD3-knockdown HCT116 and Caco-2 cells (2×10^5/well) were seeded in 6-well plates and cultured for 48 h to reach 70–80% confluence. Fresh Annexin V-FITC/PI working solution was prepared on ice. After medium removal, 100 μL of staining solution was added and incubated for 15 min at 37 °C in the dark. Cells were gently detached with EDTA-free

trypsin and resuspended in 400 µL cold PBS. Samples were analyzed within 1 h using a BD FACSCanto II, collecting 10,000 events/sample.

Colony Formation Assay

Clonogenic growth was evaluated using a colony formation assay. Parental and PLOD3-knockdown HCT116 and Caco-2 cells were collected in the logarithmic phase and seeded at 500 cells per well in 6-well plates containing 2 mL complete medium. Plates were gently agitated for uniform distribution and maintained at 37 °C, 5% CO2 for 10–14 days, with medium renewed every 3–4 days. Colonies were ffxed in 4% PFA for 15 min, stained with 0.5% crystal violet (in 25% methanol) for 20 min, rinsed thoroughly, and air-dried overnight.

Wound Healing Assay

Cell migration was assessed with a wound-healing assay. Parental and PLOD3- knockdown HCT116 and Caco-2 cells were seeded in 6-well plates (5×10^5 cells/well) and grown to full confluence, then serum-starved for 12 h in medium containing 0.5% FBS. A linear scratch was made with a sterile 200 µL tip, debris was removed with two PBS washes, and wells were replenished with 2 mL of medium containing 0.5% FBS. Images were captured at 0 h and again at 24 h using an inverted microscope (Olympus IX73). Wound closure was quantified in ImageJ. Three independent experiments were performed, each with triplicate wells.

Transwell Migration and Invasion Assays

Migration and invasion were measured using 8.0 µm Transwell chambers. Invasion inserts were coated with Matrigel (100 µL, 1 mg/mL; 4–6 h at 37 °C). Parental and PLOD3-knockdown HCT116 (2×10^4 cells) and Caco-2 (1×10^4 cells) were seeded in serum-free medium into the upper chamber; the lower chamber contained medium with 20% FBS. After 24 h (HCT116) or 48 h (Caco-2), non-migrated cells were removed, and traversed cells were fixed (4% PFA), stained (0.1% crystal violet), imaged, and counted in five random fields. Experiments were run in triplicate and repeated three times.

Statistical Analysis

Statistical analyses were performed in R 4.1.3. Correlations between continuous variables were assessed using Pearson or Spearman methods. Chi-squared tests were applied to categorical data, and continuous variables were compared using Wilcoxon rank-sum or t-tests. Survival analyses were conducted using Cox regression and Kaplan–Meier approaches within the survival package.

Results

Sialoglycan and Amino Acid Metabolism Signatures Identify Prognostic Drivers in Colorectal Cancer

We first examined the correlation between 105 sialylation-related genes and 255 genes associated with amino acid metabolism. A total of 360 genes with strong correlations were identified (Fig.2 A). We conducted differential expression analyses across four independent cohorts (TCGA, GSE110224, GSE22598, and GSE41328) to investigate their potential relevance in colorectal cancer (CRC). Among these 360 genes, 270 showed significant differential expression between tumor and adjacent normal tissues in at least one dataset (Figure 2 B). Subsequent correlation analysis of these 270 genes revealed distinct expression patterns (Figure 2 D). Univariate Cox regression identified six genes—*AOC3*, *PLOD3*, *CDO1*, *PHGDH*, *MAT2B*, and *ASRGL1*—that were significantly associated with patient prognosis (Figure 2 C), suggesting their potential roles as prognostic biomarkers.

Metabolic Signature-Based Classification Reveals Prognostically Distinct Tumor Subtypes

We utilized non-negative matrix factorization (NMF) clustering to assess the prognostic significance of the six identified genes. Based on the consensus matrix (Figure 3 A), six clusters were identified as optimal. Kaplan–Meier survival analysis revealed significant overall survival differences across the six clusters, with Cluster 1 (C1) having the best prognosis and Cluster 3 (C3) the worst. Clinical feature comparisons indicated significant differences in survival status across subtypes (Figure 3 B), and NMF-based subtypes were found to be associated with TCGA immune subtypes and molecular classifications (Figure 3 C). We conducted a differential gene expression analysis between C1 and C3 to highlight molecular differences (Figure 3 D). Analysis of differentially expressed genes indicated significant roles in immune responses, metabolism, and vascular development (Figure 3 E). Furthermore, we assessed correlations between ssGSEA enrichment scores of biological pathways and the expression levels of the six prognostic genes, highlighting their functional relevance (Figure 3 F).

Single cell / spatial analysis of the tumor ecosystem

To better characterize how sialylation shapes the tumor microenvironment (TME), we analysed single cell RNA seq data and delineated seven major cell populations: myeloid cells, B cells, NK cells, plasma cells, epithelial cells, endothelial cells, and fibroblasts. Each cell was scored using the six gene prognostic signature, with epithelial and TNK (T and NK) populations showing the highest scores. Cells were stratified into high and low score groups using AUCell thresholds, followed by differential expression and functional enrichment analyses to dissect biological differences (Figure 4A). In parallel, spatial transcriptomics

revealed distinct spatial arrangements of epithelial, immune, and stromal compartments. Epithelial regions displayed high functional scores, consistent with the single cell findings (Figure 4B). Functional scores correlated positively with immune infiltration but inversely with epithelial and stromal content. Enrichment of immunometabolic pathways in high score regions further supported their immunometabolic activity (Figure 4C).

A Prognostic Model Built on Metabolic Drivers Robustly Predicts Patient Outcomes

To develop a reliable prognostic tool, we evaluated 101 modelling strategies built from the six prognostic genes. The TCGA cohort served as the training set, and seven independent GEO datasets were used for external validation. Among all candidate models, the Random Survival Forest (RSF) combined with Ridge regression achieved the highest mean C index across the test datasets (Figure 5 A). Time dependent ROC analyses showed stable predictive performance at 1 , 3 , and 5 year time points in all eight datasets (Figure 5 B). Consistent C index values indicated robust model behaviour, and survival analyses demonstrated that higher risk scores were strongly associated with worse prognosis across all cohorts.

Our Model Outperforms Established Predictors Across Multiple Cohorts

Risk curves and principal component analysis (PCA) in all eight datasets revealed a clear separation between high and low risk groups (Figure 6 A–B). In most datasets, our risk score achieved a higher C index than conventional clinical indicators (Figure 6 C). We further benchmarked the model against 22 prognostic signatures published in the past two years. Although the signature did not outperform all comparators in the TCGA cohort, it showed superior or comparable performance in most external cohorts (Figure 6 D), supporting its robustness and generalizability.

A Nomogram Integrating Clinical and Molecular Features Enhances Risk Stratification

We combined the molecular risk score with standard clinical parameters and evaluated their prognostic value using univariate and multivariate Cox regression (Figure 7 A). On this basis, we constructed a nomogram to estimate individual survival probabilities (Figure 7 B). Decision curve analysis (DCA) indicated that both the nomogram and the risk score provided greater net clinical benefit than traditional clinical variables alone (Figure 7 C). Calibration curves for 1 , 3 , and 5 year survival showed close agreement between predicted and observed outcomes (Figure 7 D). Survival analyses based on nomogram derived scores confirmed that patients with higher scores experienced significantly poorer survival (Figure 7 E).

High-Risk Patients Show Immune Activation and Increased Tumor Mutation Burden

Analysis of risk score distributions across the six NMF defined subtypes revealed marked heterogeneity, with Cluster 3 exhibiting the highest risk (Figure 8 A). Correlation analyses between the risk score and hallmark gene sets highlighted pathways predominantly related to immune regulation and metabolism. Tumor mutation burden (TMB), calculated from somatic mutation profiles, differed significantly between risk groups (Figure 8 A). High risk patients also displayed altered tumor microenvironment features, reffected by immune, stromal, ESTIMATE, and tumor purity scores. Immune infiltration deconvolution using CIBERSORT, MCP counter, and TIMER

Immune Landscape and Drug Sensitivity Suggest Distinct Therapeutic Vulnerabilities

Risk scores were positively associated with the expression of multiple immune checkpoint genes. TIDE analysis suggested that patients in the high risk group had a lower probability of responding to immunotherapy, and non responders tended to have higher risk scores. Conversely, the low risk group showed higher IPS values (Figure 9 A). Analyses of two melanoma immunotherapy cohorts (PRJEB23709 and GSE91061) further indicated that elevated risk scores correlated with poorer survival and reduced benefit from immune checkpoint blockade (Figure 9 B). Drug sensitivity modelling predicted that agents such as bortezomib, dactinomycin, daporinad, and dinaciclib would be more effective in the low risk group (Figure 9 C), pointing to distinct therapeutic opportunities.

Cell-Cell Communication Patterns Reveal Immune Escape Programs in High-Risk Tumor Cells

To explore differences in intercellular communication related to prognostic stratification, we computed single-cell risk scores and categorized cells into high- and low-risk groups based on the median score. CellChat analysis revealed distinct communication patterns between these groups (Figure 10 A). High-risk epithelial, endothelial, and fibroblast cells showed increased signaling via the MDK–NCL pathway, whereas decreased MIF–(CD74+CXCR4) signaling was observed (Figure 10 B). These results indicate that impaired cell–cell communication might play a role in the unfavorable prognosis of high-risk cellular groups.

PLOD3 Is Upregulated In Colorectal Cancer And Promotes Tumor Progression In Vitro

Our multi-omics analysis identified PLOD3 as a potential prognostic driver in CRC. To validate its clinical relevance, we first analyzed PLOD3 expression in patient samples. Quantitative RT-PCR revealed that PLOD3 mRNA levels were significantly elevated in peripheral blood mononuclear cells from CRC patients compared to healthy controls (Figure 11 A). Concordantly, western blot analysis demonstrated substantial upregulation of PLOD3 protein in tumor tissues relative to matched adjacent normal mucosa from the same

patients (Figure 11 B), confirming its dysregulation in CRC. To investigate the functional consequences of PLOD3 dysregulation, stable PLOD3 knockdown was achieved in Caco-2 and HCT116 CRC cell lines using two independent shRNAs (Figure 11 C). Depletion of PLOD3 triggered marked molecular changes characteristic of tumor suppression: increased expression of the pro-apoptotic protein BAX alongside decreased anti-apoptotic Bcl-2, and elevated E-cadherin accompanied by reduced Ncadherin and Vimentin—a pattern consistent with attenuated epithelial-mesenchymal 11transition (Figure 11 D). Functionally, PLOD3 knockdown significantly enhanced apoptosis in both cell lines (Figure 11 E) and severely compromised their clonogenic capacity (Figure 11 F). In motility assays, PLOD3-deffcient cells exhibited markedly impaired wound closure, reduced Transwell migration, and diminished Matrigel invasion (Figure 12).

Discussion

Colorectal cancer (CRC) is characterized by significant biological heterogeneity and a limited response to immunotherapy, underscoring the need for novel biomarkers that integrate multiple aspects of tumor biology. Our study identified a six-gene signature that reflects the interplay between sialylation and amino acid metabolism, effectively stratifying CRC patients into prognostically and immunologically distinct subtypes. This integrative metabolic framework provides novel insights into how metabolic reprogramming in CRC contributes to tumor progression and immune modulation, with direct implications for prognosis and therapeutic response.

Emerging evidence highlights a critical link between altered amino acid metabolism and aberrant protein glycosylation, particularly sialylation. For instance, PHGDH, a key enzyme in the serine biosynthesis pathway, has been shown to interact with phosphofructokinase and regulate the hexosamine-sialic acid biosynthetic axis. Loss of this interaction leads to an increased production of sialic acid precursors, resulting in elevated sialylation of glycoproteins such as integrin αvβ3, which in turn promotes tumor cell motility and invasion[20]. Our study corroborates this metabolic-glycomic axis in CRC by identifying *PHGDH*, *MAT2B*, and *ASRGL1* as core risk genes, which are co-enriched in high-risk subtypes exhibiting both elevated sialylation and amino acid metabolism signatures. These genes likely contribute to the biosynthesis of nucleotide sugars, which act as essential donors for sialyltransferase-mediated glycosylation, thereby enhancing tumor invasiveness and immune evasion.

Furthermore, reprogrammed amino acid metabolism in CRC is not merely a passive reflection of proliferative demands; it actively shapes tumor biology through its influence on oncogenic signaling, epigenetic regulation, and immune modulation[21]. Enzymes such as

PHGDH and *MAT2B* serve as molecular switches, linking nutrient utilization to tumor-promoting pathways[22, 23]. *PHGDH*, as a rate-limiting enzyme in serine biosynthesis, not only fuels nucleotide and redox metabolism but also regulates glycosylation precursor flux, which has been implicated in promoting metastatic traits21. Likewise, *MAT2B*, by modulating the methionine cycle, affects the availability of S-adenosylmethionine, a critical methyl donor that drives methylation-dependent gene regulation. This, in turn, contributes to the activation of the MEK/ERK pathway, a key signaling cascade implicated in CRC growth and survival[24].

Beyond cellular proliferation, amino acid metabolism intersects with cellular stress responses and immune modulation. *PLOD3*, traditionally known for its role in collagen biosynthesis, has recently emerged as a modulator of autophagy and immune signaling. It drives tumor progression through the TNF-α/NF-κB pathway, with emerging evidence indicating its role in regulating T cell activity within the metastatic microenvironment[25]. These findings underscore the multifaceted role of amino acid metabolism in CRC, revealing its impact not only on intrinsic tumor properties but also on the tumor's ability to influence immune cell dynamics.

Our study further solidifies these findings by identifying *PHGDH, MAT2B, PLOD3, CDO1, ASRGL1*, and *AOC3* as key players in CRC prognosis. These genes are involved in various metabolic pathways, including redox regulation (*CDO1*), glutamine metabolism (*ASRGL1*), and immune cell trafficking (*AOC3*). The high co-enrichment of these genes in high-risk subtypes suggests that metabolic circuits may influence not only tumor-intrinsic characteristics but also the immune contexture of CRC. For instance, *AOC3*, which is involved in the metabolism of polyamines, has been implicated in regulating immune cell trafficking, further linking metabolic alterations with immune evasion and resistance to therapy.

In summary, our findings establish a conceptual framework in which amino acid metabolism serves as a central hub that integrates tumor growth, immune modulation, and therapeutic resistance in CRC. The six-gene signature identified in this study not only provides prognostic value but also offers potential therapeutic targets for metabolic intervention in CRC. By targeting metabolic pathways that influence both tumor growth and immune evasion, we can enhance the efficacy of existing therapies and develop new strategies to overcome the challenges of immunotherapy resistance in CRC.

Ethics statement

Ethics approval and consent to participate This study was approved by the Ethics Committee of Jiangsu Province Hospital of Chinese Medicine (Approval No. 2022NL-

KS086). All procedures were performed in accordance with the Declaration of Helsinki, and written informed consent was obtained from all participants. Consent for publication Informed consent was obtained from all individual participants for the publication of their anonymized clinical data and any associated images included in this study.

Trial registration: Not applicable. This study was not a clinical trial because no participants were prospectively assigned to any health-related intervention.

Data availability statement

All data used in this study are publicly available. TCGA colorectal cancer data were obtained from the NCI Genomic Data Commons, including TCGA-COAD (https://portal.gdc.cancer.gov/projects/TCGA-COAD) and TCGA-READ (https://portal.gdc.cancer.gov/projects/TCGA-READ). GEO datasets included GSE17537, GSE39582, GSE12945, GSE87211, GSE17536, GSE38832, GSE41258, GSE22598, GSE110224, GSE41328, GSE132465, GSE225857, and GSE91061; each dataset can be accessed through the NCBI GEO database by its accession number (https://www.ncbi.nlm.nih.gov/geo/).

The single-cell RNA-sequencing data were obtained from GEO accession GSE132465 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE132465), and the spatial transcriptomics data were obtained from GEO accession GSE225857 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE225857).

The melanoma immunotherapy cohort PRJEB23709 was obtained from the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena/browser/view/PRJEB23709), and the additional immunotherapy-related dataset GSE91061 was obtained from GEO (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE91061).

The sialylation-related gene set was derived from the published study available in PubMed Central, PMID/PMCID: PMC9623420 (https://pmc.ncbi.nlm.nih.gov/articles/PMC9623420/). Amino-acid-metabolism pathway annotations were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG), including the human pathway resource used for the analysis (https://www.genome.jp/pathway/hsa00001). Gene-set enrichment analyses used the Molecular Signatures Database (MSigDB), specifically the Hallmark, Gene Ontology Biological Process, and KEGG-related collections. The MSigDB collection index is available at https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp.

All datasets listed above are publicly available from the indicated repositories. The processed data generated in this study are provided in the Supplementary Materials.

Conflict of Interest

The authors have declared that no competing interests exist.

Author Contributions

Author contribution YZ, XF and TY contributed equally to this work and share first authorship. LZ designed the methodology and was responsible for project administration. YZ performed the bioinformatics analysis and wrote the original draft. XF conducted the experimental validation and wrote the original draft. TY was responsible for visualization and image editing. CZ and NY performed the experiments and validation. XY acquired the funding. XY and SZ supervised the study and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Funding This work was supported by Shanghai Municipal Science and Technology Commission. Project Title: Application Demonstration Project of Polyvinyl Alcohol Embolization Microspheres — Subproject: Application and Comparative Evaluation of the Drug-Loading Performance of Polyvinyl Alcohol Embolization Microspheres. Grant number is 24SF1905505.

References

1. Hossain MS, Karuniawati H, Jairoun AA, Urbi Z, Ooi J, John A, et al. Colorectal Cancer: A Review of Carcinogenesis, Global Epidemiology, Current Challenges, Risk Factors, Preventive and Treatment Strategies. Cancers (Basel). 2022; 14.
2. Anand U, Dey A, Chandel AKS, Sanyal R, Mishra A, Pandey DK, et al. Cancer chemotherapy and beyond: Current status, drug candidates, associated risks and progress in targeted therapeutics. Genes Dis. 2023; 10: 1367-1401.
3. Xie YH, Chen YX, Fang JY. Comprehensive review of targeted therapy for colorectal cancer. Signal Transduct Target Ther. 2020; 5: 22.
4. Sadida HQ, Abdulla A, Marzooqi SA, Hashem S, Macha MA, Akil ASA, et al. Epigenetic modifications: Key players in cancer heterogeneity and drug resistance. Transl Oncol. 2024; 39: 101821.
5. Zhu W, Zhou Y, Guo L, Feng S. Biological function of sialic acid and sialylation in human health and disease. Cell Death Discov. 2024; 10: 415.
6. Liang S, Tang Q, Guo X, Li Z, Guo Y, Chang J, et al. Mutant glycosidases for labeling sialoglycans with high specificity and affinity. Nat Commun. 2025; 16: 1427.

7. Silva Z, Soares CO, Barbosa M, Palma AS, Marcelo F, Videira PA. The role of sialoglycans in modulating dendritic cell function and tumour immunity. Semin Immunol. 2024; 74-75: 101900.
8. Pietrobono S, Stecca B. Aberrant Sialylation in Cancer: Biomarker and Potential Target for Therapeutic Intervention? Cancers (Basel). 2021; 13.
9. Chen J, Zhang S, Huang X, Wang Q, Xu W, Huang J, et al. Sialylated IgG-activated integrin β4-Src-Erk axis stabilizes c-Myc in a p300 lysine acetyltransferase-dependent manner to promote colorectal cancer liver metastasis. Neoplasia. 2025; 61: 101140.
10. Cui H, Yang S, Jiang Y, Li C, Zhao Y, Shi Y, et al. The glycosyltransferase ST6Gal-I is enriched in cancer stem-like cells in colorectal carcinoma and contributes to their chemo-resistance. Clin Transl Oncol. 2018; 20: 1175-1184.
11. van de Wall S, Santegoets KCM, van Houtum EJH, Büll C, Adema GJ. Sialoglycans and Siglecs Can Shape the Tumor Immune Microenvironment. Trends Immunol. 2020; 41: 274-285.
12. Egan H, Treacy O, Lynch K, Leonard NA, O'Malley G, Reidy E, et al. Targeting stromal cell sialylation reverses T cell-mediated immunosuppression in the tumor microenvironment. Cell Rep. 2023; 42: 112475.
13. Liu X, Ren B, Ren J, Gu M, You L, Zhao Y. The significant role of amino acid metabolic reprogramming in cancer. Cell Commun Signal. 2024; 22: 380.
14. Yang L, Chu Z, Liu M, Zou Q, Li J, Liu Q, et al. Amino acid metabolism in immune cells: essential regulators of the effector functions, and promising opportunities to enhance cancer immunotherapy. J Hematol Oncol. 2023; 16: 59.
15. Encarnación-Rosado J, Sohn ASW, Biancur DE, Lin EY, Osorio-Vasquez V, Rodrick T, et al. Targeting pancreatic cancer metabolic dependencies through glutamine antagonism. Nat Cancer. 2024; 5: 85-99.
16. Lee JM, Hammarén HM, Savitski MM, Baek SH. Control of protein stability by post-translational modifications. Nat Commun. 2023; 14: 201.
17. Do LK, Lee HM, Ha YS, Lee CH, Kim J. Amino acids in cancer: Understanding metabolic plasticity and divergence for better therapeutic approaches. Cell Rep. 2025; 44: 115529.
18. Jing Y, Kobayashi M, Shoulkamy MI, Zhou M, Thi Vu H, Arakawa H, et al. Lysine-arginine imbalance overcomes therapeutic tolerance governed by the transcription factor E3-lysosome axis in glioblastoma. Nat Commun. 2025; 16: 2876.
19. McArthur JB, Chen X. Glycosyltransferase engineering for carbohydrate synthesis. Biochem Soc Trans. 2016; 44: 129-142.
20. Rossi M, Altea-Manzano P, Demicco M, Doglioni G, Bornes L, Fukano M, et al. PHGDH heterogeneity potentiates cancer cell dissemination and metastasis. Nature. 2022; 605: 747-753.
21. Chen J, Cui L, Lu S, Xu S. Amino acid metabolism in tumor biology and therapy. Cell Death Dis. 2024; 15: 42.

22. Wan X, Zeng W, Fan H, Wang C, Han S, Sun Z, et al. MAT2B regulates the protein level of MAT2A to preserve RNA N6-methyladenosine. Cell Death Dis. 2024; 15: 714.
23. Annibali D, Fendt SM. Nuclear PHGDH protects cancer cells from nutrient stress. Nat Metab. 2021; 3: 1284-1285.
24. Panmanee J, Bradley-Clarke J, Mato JM, O'Neill PM, Antonyuk SV, Hasnain SS. Control and regulation of S-Adenosylmethionine biosynthesis by the regulatory β subunit and quinolone-based compounds. Febs j. 2019; 286: 2135-2154.
25. Ding M, Wang C, Hu J, She J, Shi R, Liu Y, et al. PLOD3 facilitated T cell activation in the colorectal tumor microenvironment and liver metastasis by the TNF-α/ NF-κB pathway. J Transl Med. 2024; 22: 30.

Supplementary Material

Original Data for Western Blotting All original, uncropped, and full-length membranes for western blot analysis presented in the main ffgures are provided in this Supplementary Information (Supplementary Figures S1–S6). Molecular weight markers are indicated, and the regions cropped for main ffgures are clearly delineated.

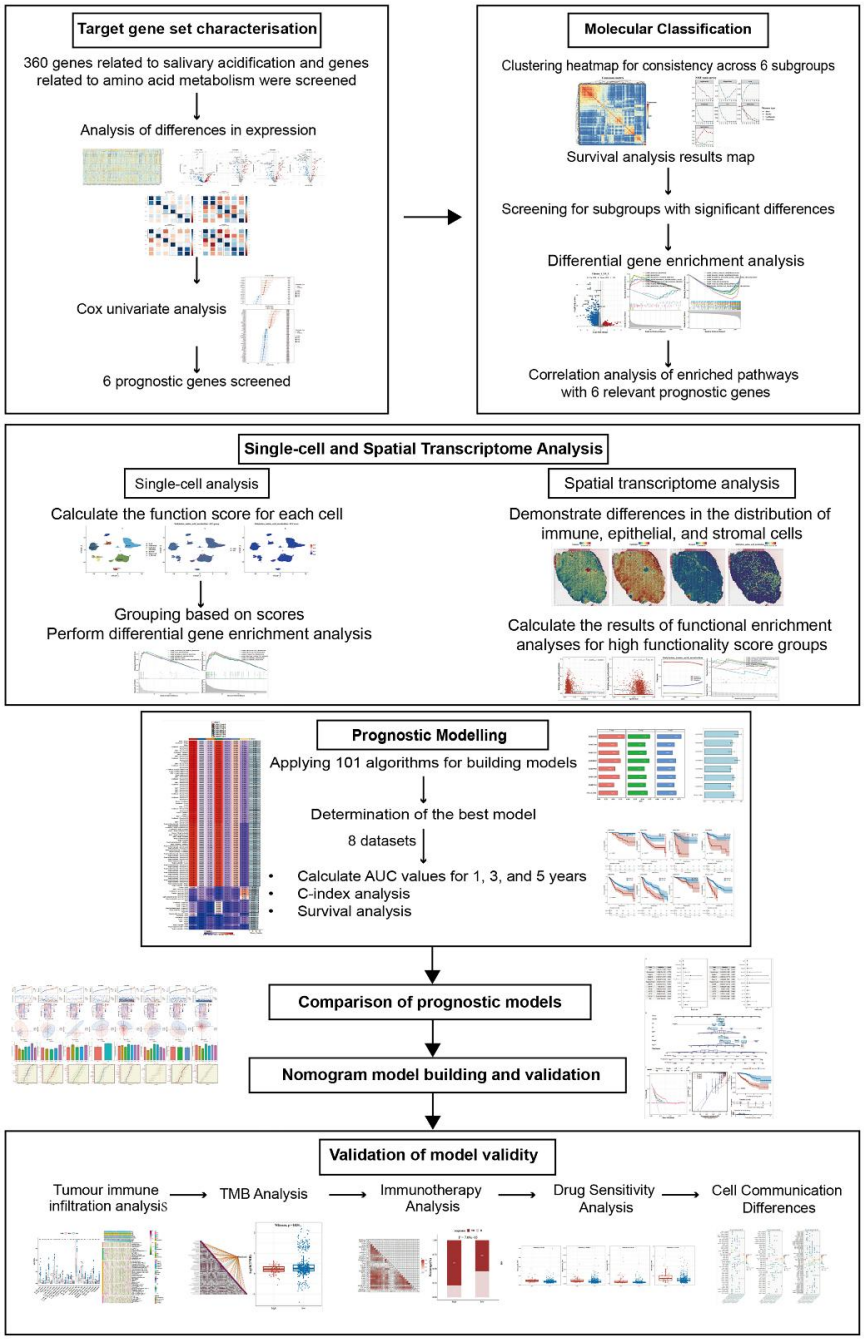
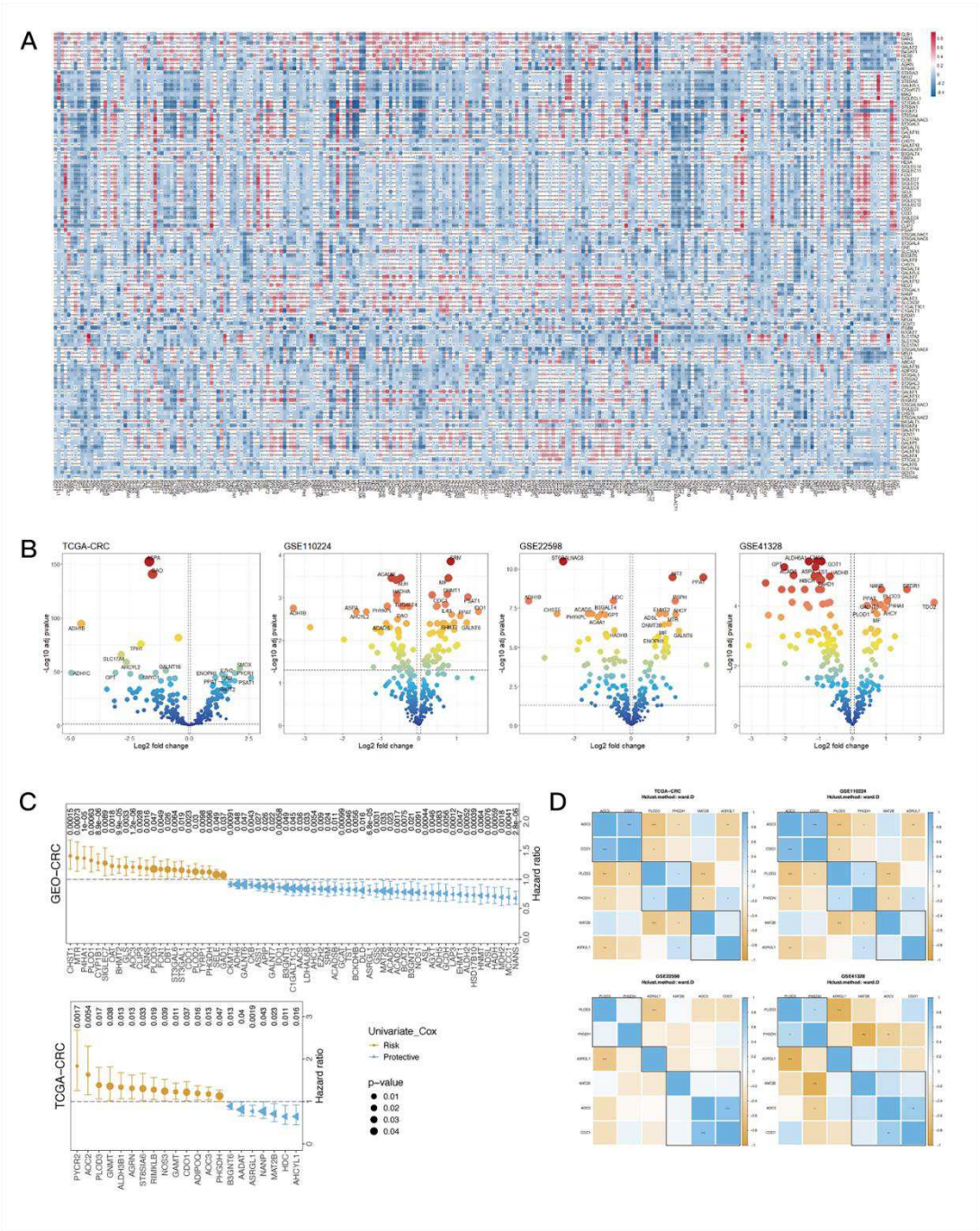


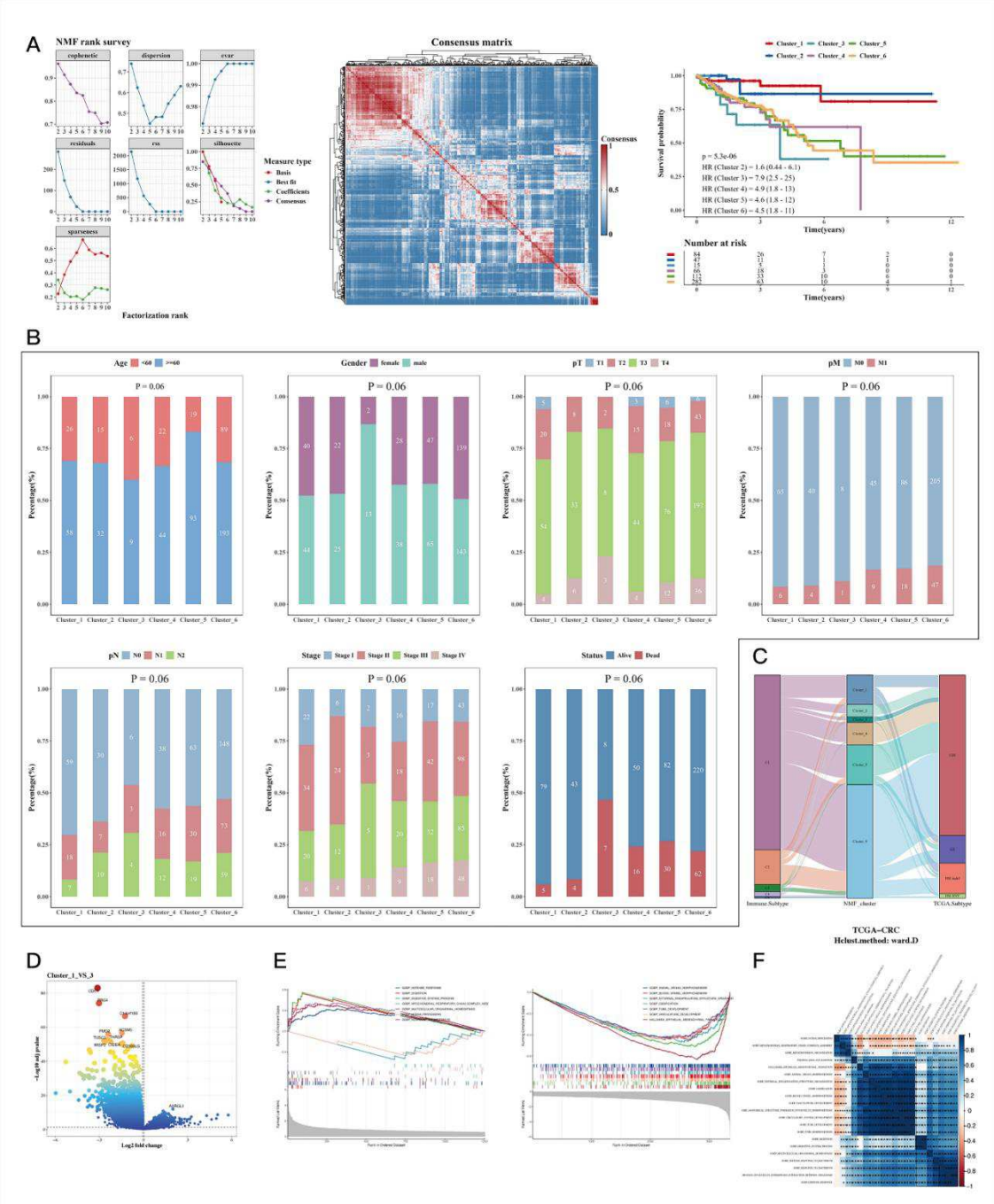
Fig. 1 Experimental workflow of the study. Their potential relevance in colorectal cancer (CRC). Among these 360 genes, 270 showed significant differential expression between tumor and adjacent normal tissues in at least one dataset. Subsequent correlation analysis of these 270 genes revealed distinct expression patterns. Univariate Cox regression identified six genes—AOC3, PLOD3, CDO1, PHGDH, MAT2B, and ASRGL1—that were significantly associated with patient prognosis, suggesting their potential roles as prognostic biomarkers.



581

582 **Fig. 2** Characterization of sialylation and amino acid metabolism gene sets. A. Heatmap
583 showing significant correlations between sialylation-related and amino acid metabolism–
584 related genes. B. Volcano plots of differentially expressed genes between tumor and
585 adjacent tissues across four datasets (TCGA and three GEO cohorts). C. Forest plot of
586 univariate Cox regression for prognostic genes in the combined cohort. D. Expression
587 correlation heatmap of differentially expressed genes across all datasets.

588



589

590 **Fig. 3** Molecular subtyping and functional analysis. A. NMF clustering of patients based on
591 six prognostic genes identifies six molecular subtypes with distinct survival patterns. B.
592 Distribution of clinical features across subtypes. C. Sankey diagram showing relationships
593 between NMF subtypes, TCGA immune subtypes, and TCGA molecular subtypes. D.
594 Differential gene expression between subtypes C1 and C3. E. GSEA results of upregulated
595 and downregulated genes between C1 and C3. F. Correlation heatmap between enriched
596 pathways and ssGSEA scores of the six prognostic genes.

597

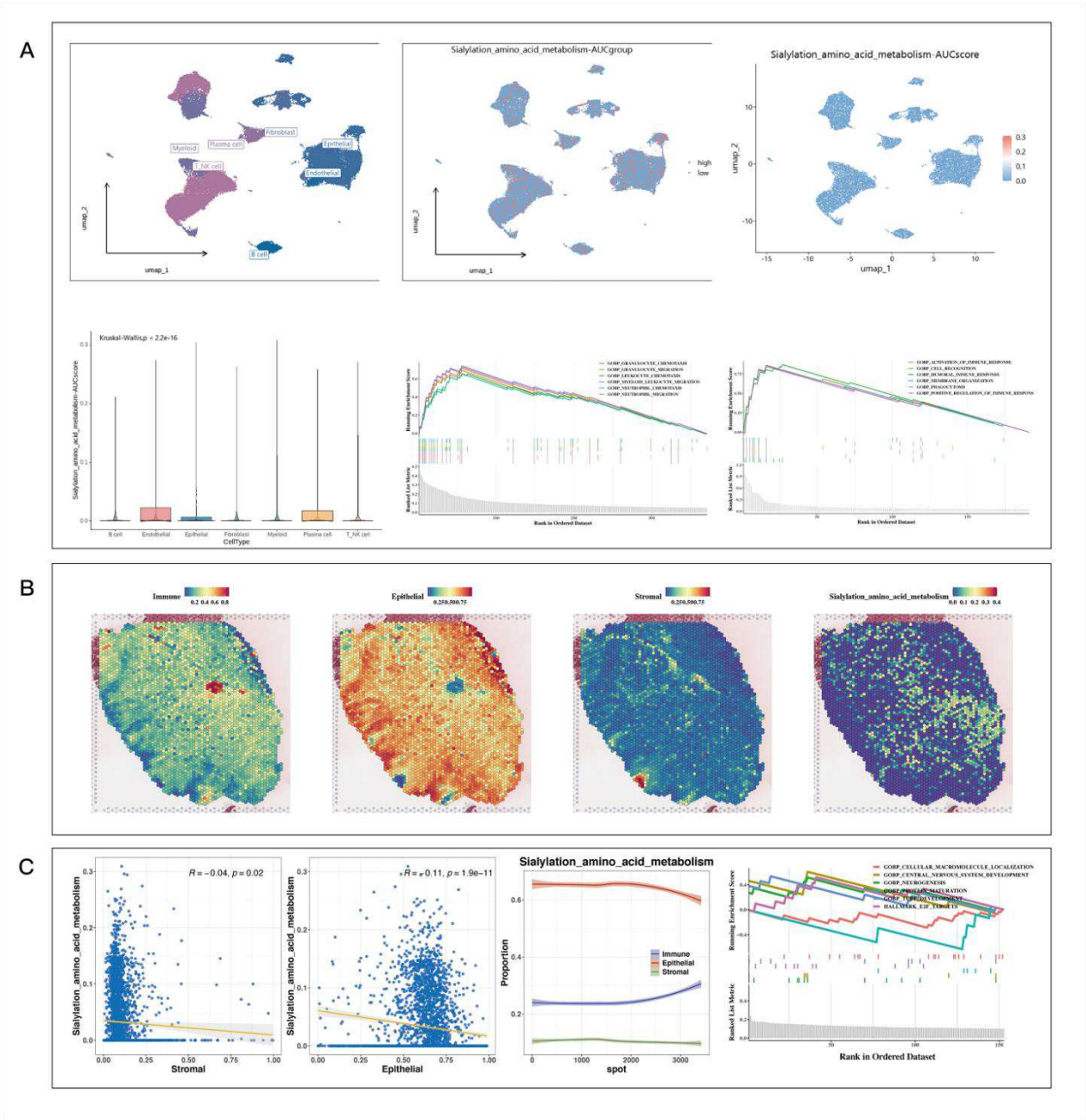


Fig. 4 Single-cell and spatial transcriptomic validation. A. Cell-type annotation, AUCell scoring, and functional enrichment analysis of high- vs. low-scoring cells in scRNA-seq data. B. Spatial transcriptomic maps showing cell-type distributions and regional enrichment of functional scores. C. Correlations between functional score and epithelial, stromal, or immune features; GSEA of highscore regions.

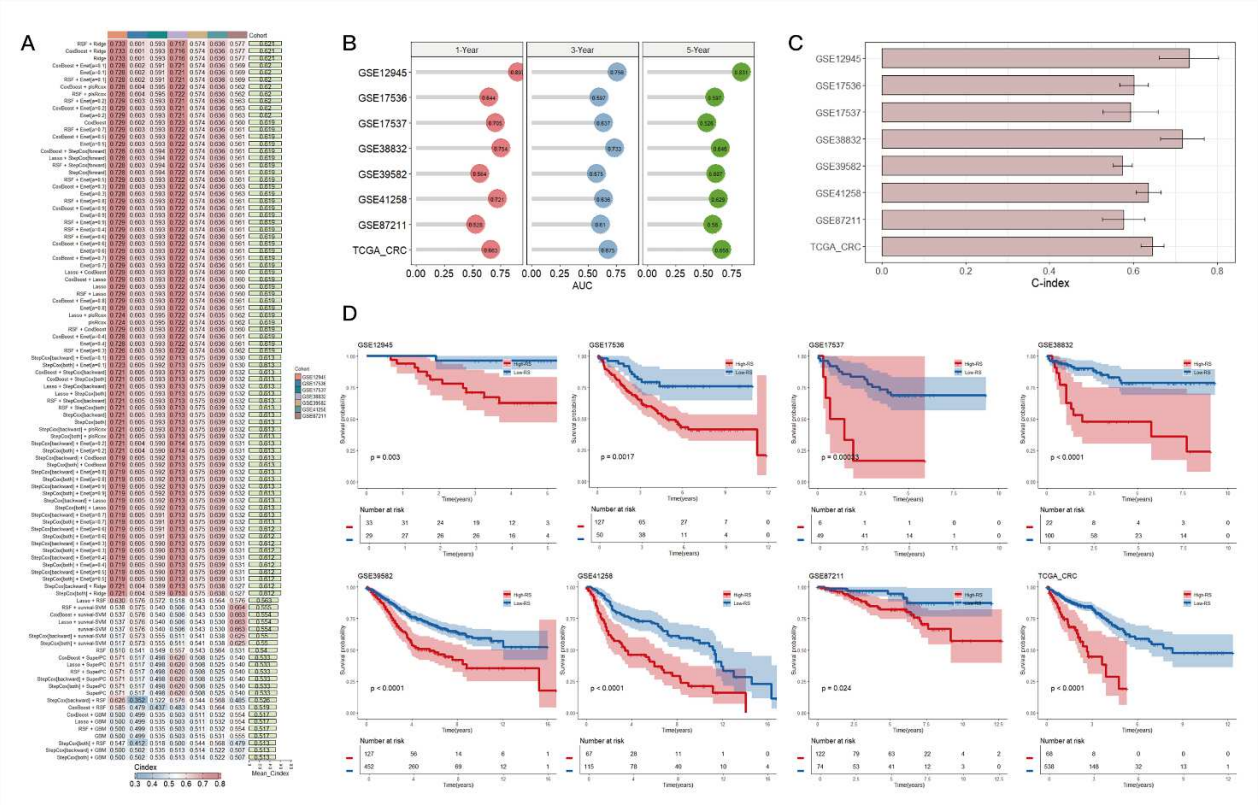


Fig. 5 Construction of the prognostic model. A. C-index heatmap of 101 algorithm combinations across seven validation datasets. B. Time-dependent AUCs for predicting 1-, 3-, and 5-year survival. C. C-index of the ffnal model across eight datasets. D. Kaplan–Meier curves comparing survival between high- and low-risk groups.

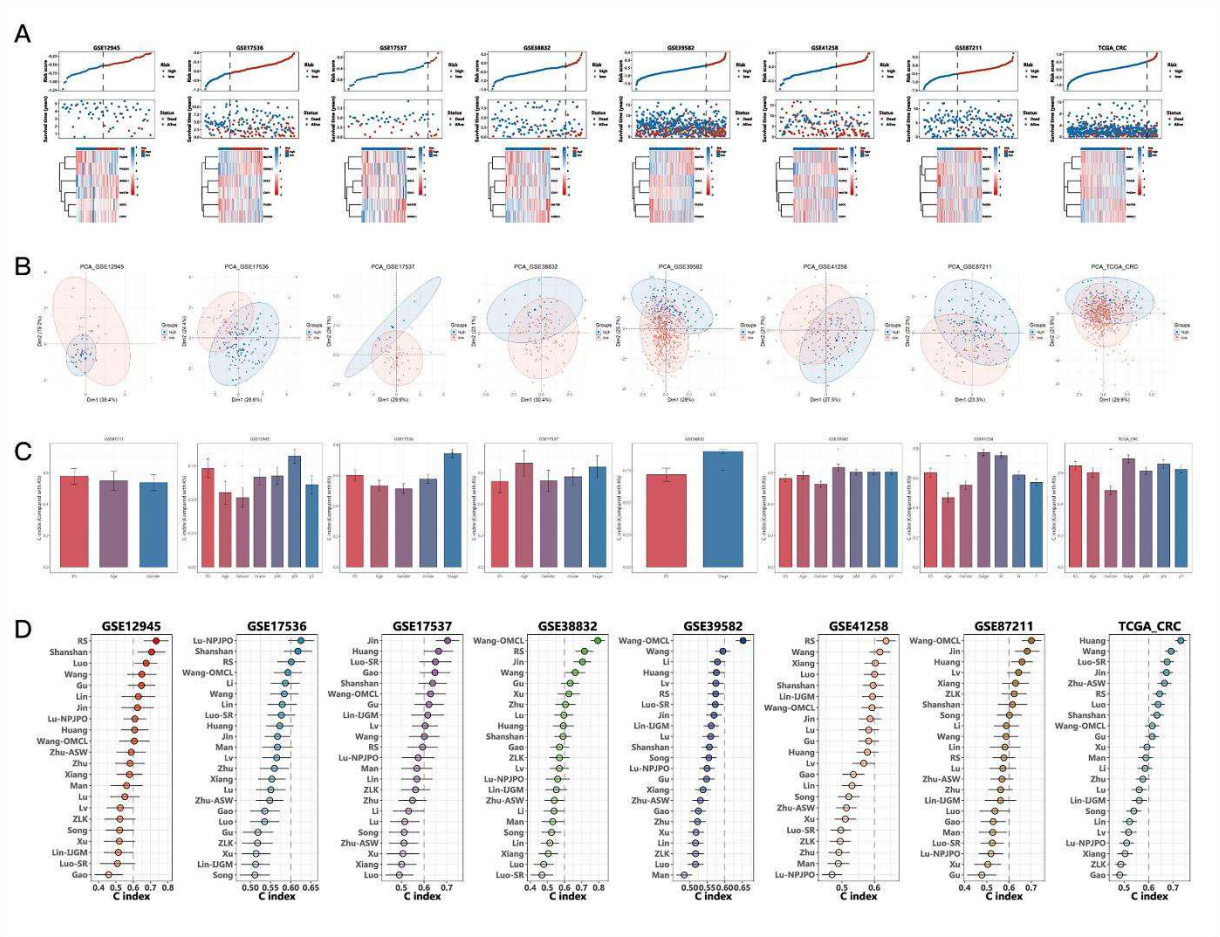


Fig. 6 Comparative performance of the risk model. A. Risk score distributions across datasets. B. PCA showing separation between risk groups. C. C-index comparison of the risk score with clinical features. D. C-index comparison with 22 published prognostic models across eight datasets.

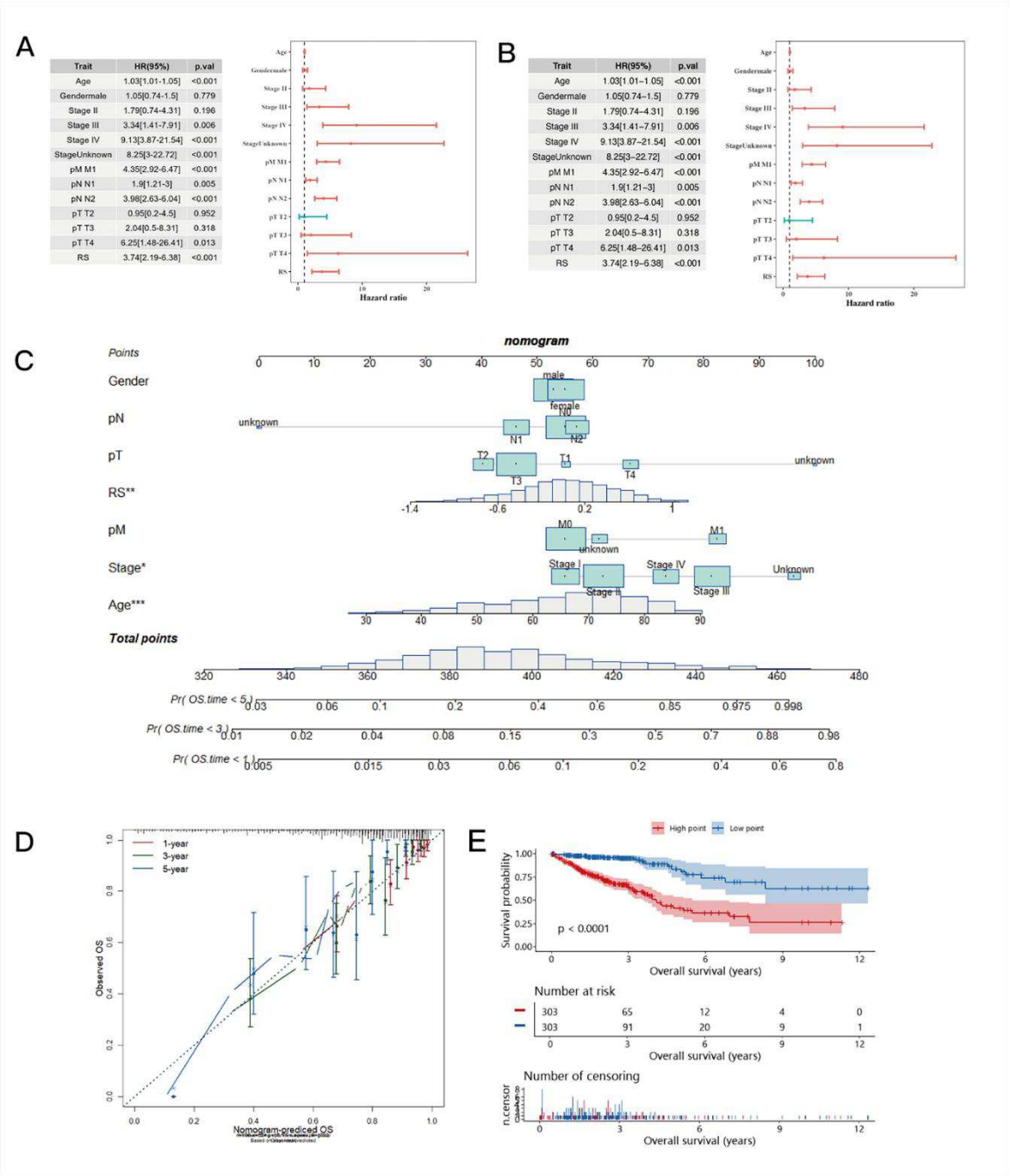
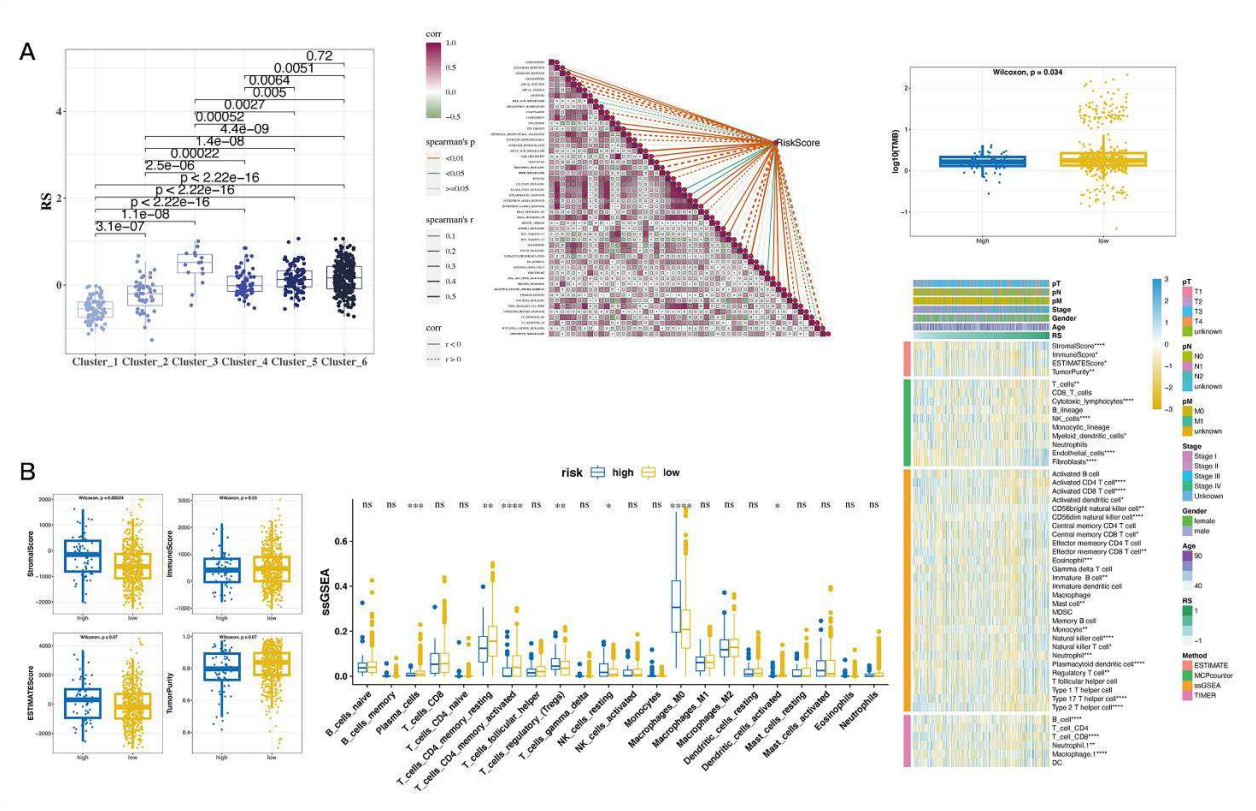


Fig. 7 Nomogram construction and evaluation. A. Univariate and multivariate Cox analyses of the risk score and clinical variables. B. Nomogram integrating the risk score and clinical factors. C. Decision curve analysis comparing models. D. Calibration plots for 1-, 3-, and 5-year survival predictions. E. Survival analysis stratiffed by total nomogram score.



623

624 **Fig. 8** Immune inffltration and tumor mutation burden. A. Risk score distribution across NMF
625 subtypes; correlations with hallmark gene sets; TMB differences between risk groups. B.
626 Immune, stromal, and ESTIMATE scores, as well as tumor purity; immune cell inffltration
627 proffles from CIBERSORT, MCP-counter, and TIMER.

628

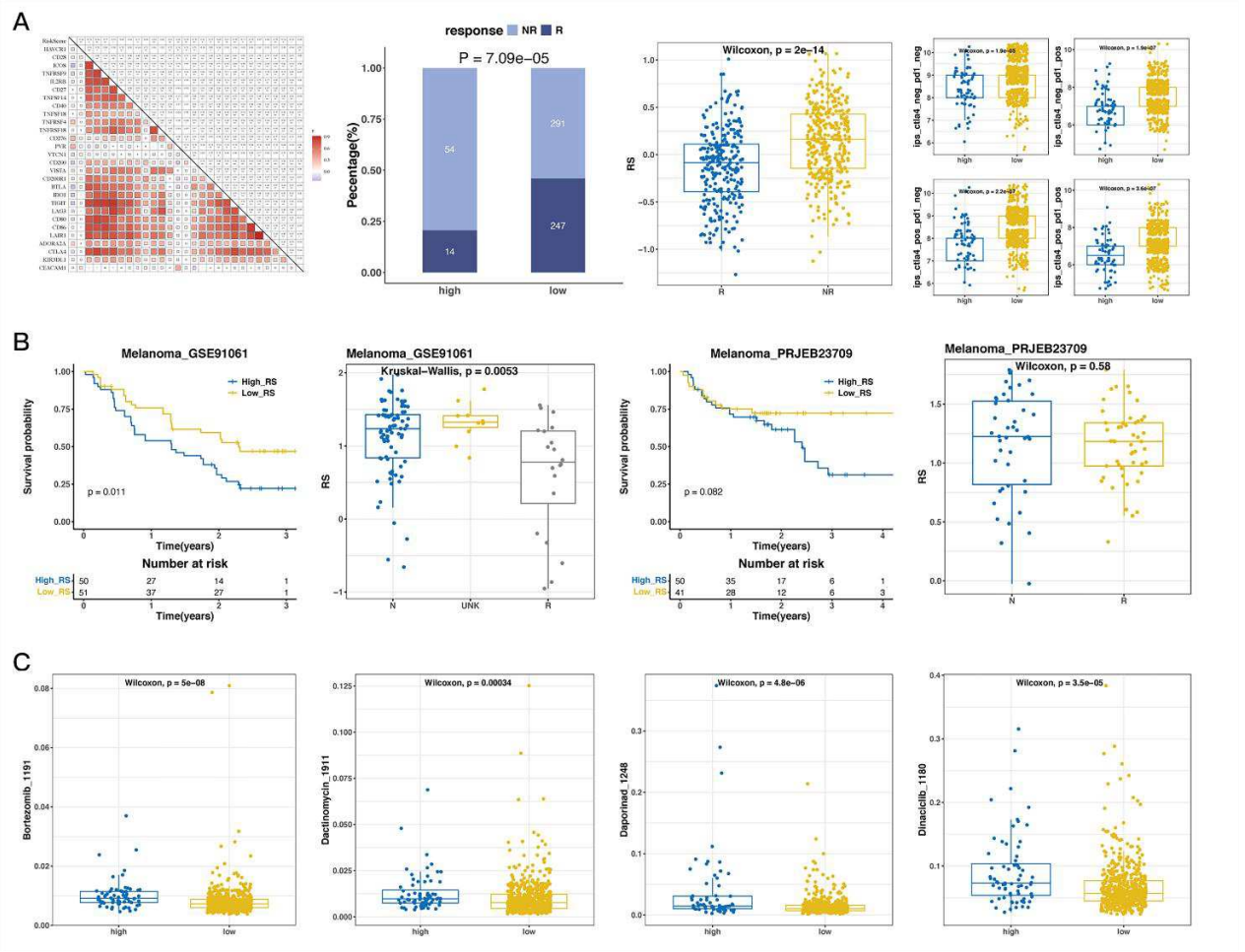


Fig. 9 Immunotherapy prediction and drug sensitivity. A. Correlations between the risk score and immune checkpoint genes; TIDE and IPS analysis in TCGA cohort. B. Survival analysis and risk score distributions in PRJEB23709 and GSE91061 immunotherapy cohorts. C. Differential drug sensitivity (IC50) between risk groups for selected compounds.

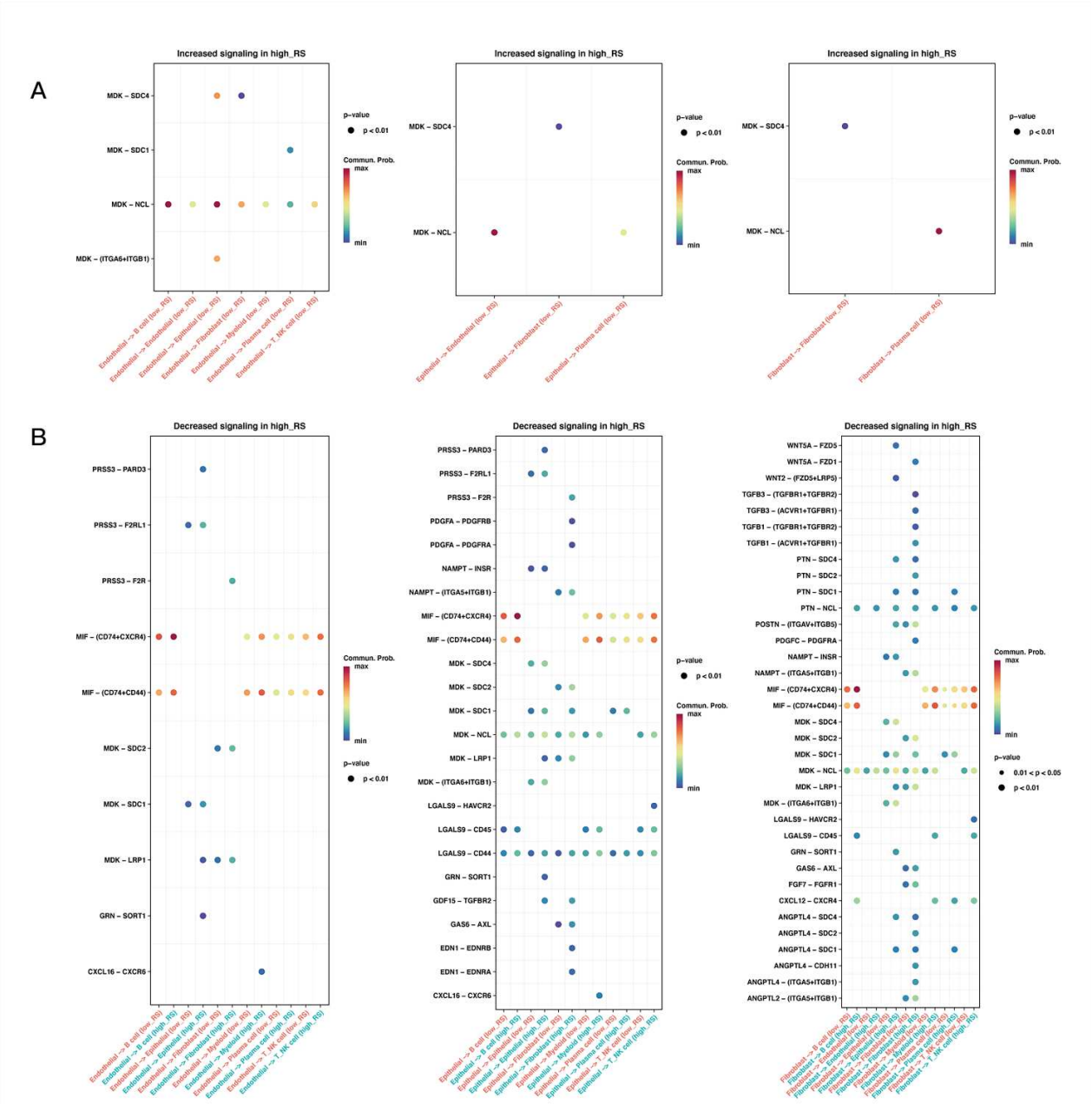


Fig. 10 Cell-cell communication differences at single-cell resolution. A. Increased signaling interactions in endothelial, epithelial, and ffbroblast cells in high-risk cells. B. Reduced signaling pathways in corresponding cell types in high-risk group.

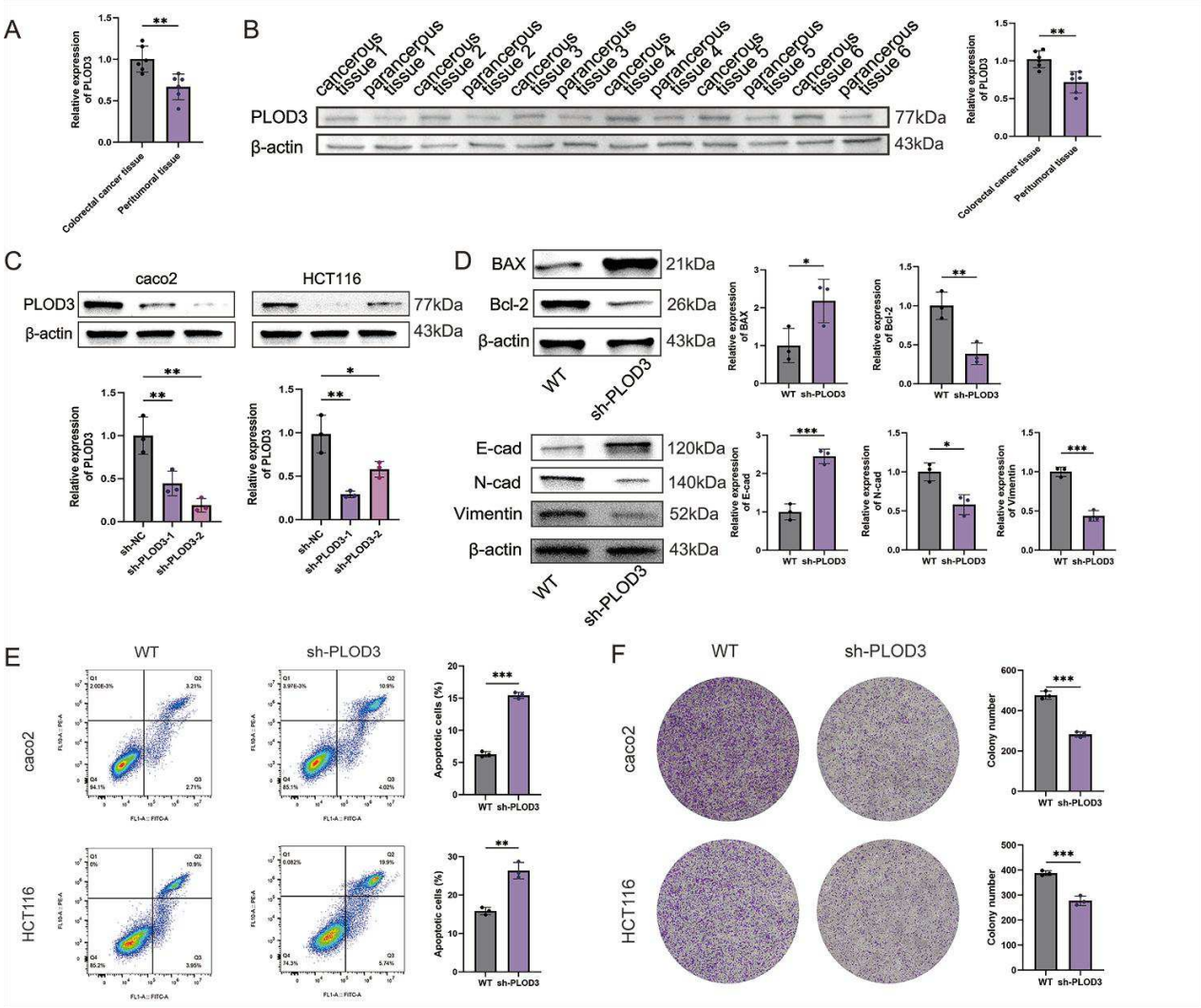
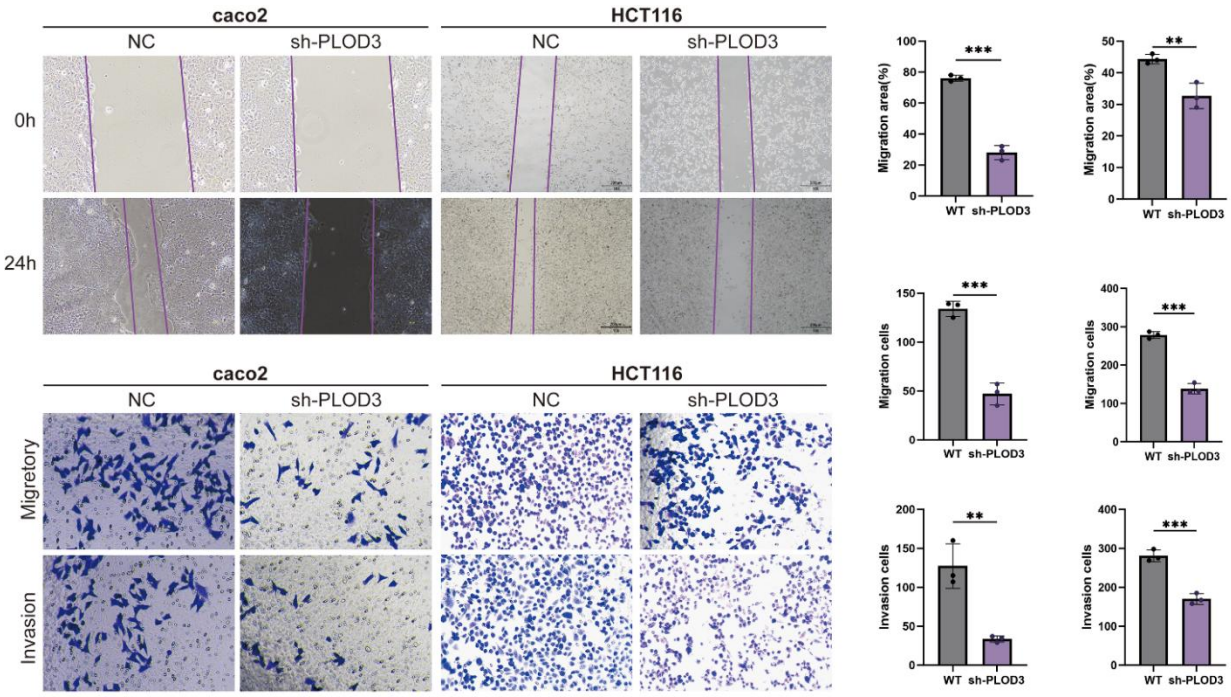


Fig. 11 PLOD3 is upregulated in colorectal cancer and promotes tumor cell survival and clonogenicity. A. PLOD3 mRNA expression levels in PBMCs from CRC patients (n=6) and healthy controls (n=6) as determined by qRT-PCR. B. Representative western blot and quantitative analysis of PLOD3 protein expression in paired CRC tumor and adjacent normal tissues from six patients. C. Validation of stable PLOD3 knockdown in Caco-2 and HCT116 cells transduced with two independent shRNAs (shPLOD3-1, shPLOD3-2) or a non-targeting control (shNC). Protein levels were assessed by western blot. D. Western blot analysis of apoptosis-related proteins (BAX, Bcl-2) and epithelial-mesenchymal transition (EMT) markers (E-cadherin, N-cadherin, Vimentin) in control and PLOD3-knockdown Caco-2 and HCT116 cells. E. Apoptosis analysis by Annexin V-FITC/PI staining and ffow cytometry in control and PLOD3-knockdown cells. F. Representative images and quantiffcation of colony formation assays in control and PLOD3-knockdown cells.



654

655 **Fig. 12** PLOD3 knockdown impairs the migratory and invasive capacities of colorectal
656 cancer cells. Representative images and quantitative analysis of wound healing assays
657 performed in control and PLOD3-knockdown Caco-2 and HCT116 cells; Transwell migration
658 assays. Representative images and quantiffcation of migrated control and PLOD3-
659 knockdown cells after 24 h (HCT116) or 48 h (Caco-2). Migrated cells were stained with
660 crystal violet; Matrigel invasion assays. Representative images and quantiffcation of
661 invaded control and PLOD3-knockdown cells after 24 h (HCT116) or 48 h (Caco-2).
662 Invaded cells were stained with crystal violet.

663

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf