Am J Transl Res 2026;18(3):2674-2692
www.ajtr.org /ISSN:1943-8141/AJTR0171016

Original Article

Metabolic reprogramming-associated genomic instability drives colorectal cancer progression via the UBXN1-NF-κB axis

Long Qian[1,2*], Ziyi Li[1,2*], Tangtang Yang[1,2*], Shasha Xia[3*], Lei Jin[1,2], Chengyu Zhu[1,2], Wenshan Jing[1,2], Yue Wang[1,2], Yun Ye[5], Yi Shen[1,2], Lixiang Li[2,4], Hui Peng[1,2,3], Qingsheng Yu[1,2]

[1]Department of General Surgery, The First Affiliated Hospital of Anhui University of Chinese Medicine, Hefei 230031, Anhui, P. R. China; [2]Institute of Chinese Medicine Surgery, Anhui Academy of Chinese Medicine, Hefei 230031, Anhui, P. R. China; [3]Institute of Data Space, Hefei Comprehensive National Science Center, Hefei 230000, Anhui, P. R. China; [4]Department of Hepatobiliary and Pancreatic Surgery, Lujiang County People's Hospital, Anhui Province, Hefei 231500, Anhui, P. R. China; [5]Department of Traditional Chinese Medicine, Shexian Guilin Town Health Center, Huangshan 245000, Anhui, P. R. China. [*]Equal contributors.

Received December 25, 2025; Accepted March 25, 2026; Epub March 25, 2026; Published March 30, 2026

Abstract: Colorectal cancer (CRC) is a common malignancy in clinical practice, and its treatment is greatly challenged by tumor heterogeneity. The most prominent features of CRC heterogeneity are differences in metabolic states and genomic instability, which ultimately lead to unfavorable clinical outcomes. Based on this, the present study aimed to investigate the association between metabolic reprogramming and copy number variation (CNV) in CRC using single-cell datasets. By integrating four publicly available single-cell RNA sequencing datasets, a comprehensive single-cell atlas of CRC was constructed. Subsequently, epithelial cells were specifically analyzed, and consensus non-negative matrix factorization (cNMF) was applied to identify six gene expression programs, covering functional modules such as cell cycle, metabolism, inflammatory stress, and immune interaction. Genomic instability was assessed using the inference of copy number variations (InferCNV) analytical tool, which identified malignant epithelial cells characterized by large-scale CNVs. Meanwhile, metabolic pathway activity at the single-cell level was evaluated using the area under the curve cell (AUCell) method, and predictive performance was further assessed using machine learning algorithms. The results demonstrated that metabolic features could effectively predict the malignant state defined by CNVs, achieving an area under the curve (AUC) of 0.985, with protein metabolism and TP53-related pathways contributing most significantly. Further integrative analysis identified 13 metabolism-related genes associated with clinical prognosis, among which UBXN1 was identified as a central node in the protein-protein interaction network. Functional analysis of UBXN1 revealed that it suppresses the NF-κB signaling pathway, thereby regulating the malignant phenotype of CRC cells. In conclusion, this study systematically elucidates the critical link between metabolic features and genomic instability in CRC, suggesting that UBXN1 may serve as a potential therapeutic target.

Keywords: CRC, UBXN1, NF-κB signal pathway, single-cell data analysis

Introduction

Colorectal cancer (CRC) is currently one of the most common malignancies worldwide, with persistently high incidence and mortality rates each year [1, 2]. The standard therapeutic strategies for CRC include surgical resection, chemotherapy, and targeted immunotherapy. However, once the disease progresses to an advanced stage, the prognosis remains extremely poor [3]. A major underlying reason is tumor heterogeneity, which poses a significant challenge to current treatment approaches. This heterogeneity is manifested at multiple levels, including both genetic and phenotypic variations [4, 5]. Therefore, an in-depth understanding of tumor heterogeneity is essential for overcoming the therapeutic challenges associated with CRC. Genomic instability represents a key feature of heterogeneity in CRC, leading to

https://doi.org/10.62347/IWGB8367

large-scale copy number variations (CNVs). Differentially expressed genes driven by these genomic alterations play critical roles in the initiation and progression of CRC [6, 7]. In addition, chromosomal amplifications and deletions are pivotal events in cancer development, as they can activate oncogenes and inactivate tumor suppressor genes, thereby promoting the formation of malignant phenotypes [8, 9]. Beyond genomic alterations, phenotypic variation is another crucial component of tumor heterogeneity [10]. These phenotypic characteristics include functional processes such as metabolic reprogramming, inflammatory signaling, and stress responses. Together with genomic instability, they contribute to tumor heterogeneity and enhance malignant behaviors, ultimately resulting in therapeutic resistance and clinical challenges [11, 12]. Metabolic reprogramming refers to a global alteration of metabolic networks in tumor cells in response to environmental and cellular stress [13, 14]. As a hallmark of cancer, it enables cancer cells to rapidly generate energy and biosynthetic precursors, while also regulating redox balance, macromolecule synthesis, immune evasion, and resistance to therapy [15-18]. In CRC, dysregulated glucose, lipid, and protein metabolism is linked to disease progression and poor outcomes [19-22]. However, the relationship between metabolic pathway activity and genomic instability at the single-cell level is not well understood. Single-cell RNA sequencing (scRNA-seq) provides high-resolution analysis of tumor heterogeneity, allowing precise identification of malignant and non-malignant cell populations in the tumor microenvironment [23-25]. Recent CRC scRNA-seq studies have found diverse epithelial cell states and functional programs, but systematic integration of metabolic pathway activity with CNV-derived malignancy indices across large patient cohorts is limited [26, 27]. In this study, we integrated multiple high-quality public scRNA-seq datasets to create a comprehensive CRC single-cell atlas. By employing consensus non-negative matrix factorization (cNMF) for gene program identification, inference of copy number variations (inferCNV) for malignancy assessment, area under the curve cell (AUCell) for evaluating metabolic pathway activity, and machine learning modeling, we conducted a comprehensive analysis of the relationship between metabolic reprogramming and genomic instability. We further found prognostic metabolic driver genes and experimentally confirmed the oncogenic role of UBXN1 by regulating the NF-κB signaling pathway. So, our findings show a framework linking metabolism, genomic instability, and malignant progression in CRC.

Materials and methods

Data acquisition and preprocessing of single-cell RNA-seq datasets

From the public repositories, like Gene Expression Omnibus (GEO) and Array Express, we obtained the scRNA-seq data for CRC and matched non-malignant tissues. The raw count matrices, along with their corresponding metadata - including sample type, patient ID, sequencing platform, and library preparation protocol - were obtained through repository-specific tools or R/Bioconductor packages. Only datasets that have clear clinical/pathological annotations and enough cell numbers were included, according to the current best practices for scRNA-seq data reuse.

Quality control of single-cell transcriptomes

Initial quality control (QC) was conducted at both cell and gene levels. Cells of low quality or those undergoing cell death were eliminated based on criteria including extremely low counts of detected genes or unique molecular identifiers (UMIs), as well as abnormally high proportions of mitochondrial transcripts. Genes present in only a minor proportion of cells were excluded to minimize sparsity and noise [28]. QC thresholds were established based on the distribution of each metric within individual datasets, adhering to established guidelines for scRNA-seq analysis, thereby retaing rare yet biologically significant cell populations.

Identification and removal of doublets

To minimize artifacts caused by doublets (droplets encapsulating multiple cells), computational doublet detection was performed prior to downstream analysis [29]. For each sample, potential doublets were independently identified using either Scrublet or DoubletFinder. These methods produce simulated doublets by analyzing observed transcriptomes and then score real cells by comparing them to synthetic profiles. Cells that exceed an empirically deter-

mined threshold (according to bimodal score distribution) were removed as doublets. For each batch, doublet detection was done separately to prevent distortions due to sample composition differences.

Integration of multiple samples and batch effect correction

After performing QC and removing the doublet, the expression matrices from individual samples were normalized and integrated to account for batch effects caused by patient, tissue, or technical variations [30]. Integration was done using Seurat's anchor-based framework, which identifies biologically similar cells across samples ("anchors") and harmonizes datasets in a shared low-dimensional space. The integrated object supported joint dimensionality reduction, visualization, and clustering. This approach allowed direct comparison of cell states across different patients and conditions. It also preserved true biological differences.

Unsupervised clustering and cell-type annotation

The integrated expression matrix underwent principal component analysis (PCA), followed by the construction of a shared nearest neighbor (SNN) graph using the top principal components. Using graph-based community detection methods, cells were partitioned into transcriptionally homogeneous clusters and subsequently characterized with feature markers. The annotated clusters were then compared against relevant entries in the literature and public atlases. Further annotation of cellular subpopulations was performed to distinguish and classify the associated subtypes. We used automated annotation tools to support these labels. We also referred to published CRC single-cell datasets to improve annotation accuracy.

Consensus non-negative matrix factorization to infer gene expression programs

We used cNMF to analyze recurrent transcriptional programs in both malignant and non-malignant cells. We applied this method to the normalized count matrix. This method decomposes the data into gene expression programs and their usage in each cell. We repeated the analysis multiple times and combined the results to obtain stable and reproducible components. We defined each program by its top-weighted genes. We then performed pathway and functional annotation analyses. This approach helped distinguish cell identity programs from context-dependent activities, such as proliferation, stress response, and metabolic rewiring in CRC.

Inference of copy number variation from scRNA-seq using inferCNV

We used the *inferCNV* package to infer large-scale CNVs from scRNA-seq data. This approach allowed us to distinguish malignant epithelial cells from normal cells. It also enabled the assessment of intratumoral heterogeneity. We divided epithelial cells into two populations: potential "tumor" cells and "reference" cells.

Gene expression values were arranged according to chromosomal coordinates, smoothed within genomic windows, and then compared to reference cells. This comparison facilitated the inference of chromosomal gains and losses, with optional refinement through a hidden Markov model [31]. Cells displaying widespread CNV patterns characteristic of malignancy were identified as tumor cells, whereas epithelial cells without CNV alterations were deemed non-malignant.

Quantification of gene set activity at single-cell resolution using AUCell

The AUCell algorithm was employed to assess pathway and gene-signature activity within individual cells, assigning scores based on the enrichment of predefined gene sets relative to ranked gene expression levels. AUCell first constructs per-cell gene rankings based on the expression matrix, and then calculates an area under the curve (AUC) score that quantifies the proportion of genes from a specific set among the top-ranked genes in each cell. The AUC distributions identified various cell populations that exhibit active oncogenic, metabolic, or immune-related traits, allowing for the comparison between malignant and non-malignant cell compartments.

Acquisition of bulk transcriptomic and clinical data from TCGA

Bulk RNA-seq and clinical data for CRC were acquired from The Cancer Genome Atlas (TCGA)

Table 1. Primers for quantitative real-time PCR analysis

Name	Coding	Anticoding
IL-8	GAGTGATTGAGAGTGGACCACACT	AGACAGAGCTCTCTTCCATCAGAAA
IL-1β	CCACAGACCTTCCAGGAGAATG	GTGCAGTTCAGTGATCGTACAGG
TNF-α	CTCTTCTGCCTGCTGCACTTTG	ATGGGCTACAGGCTTGTCACTC
UBXN1	AAGAGTTGTCAGCAGCACGA	CCCACACTGCCACCATACTT
GADPH	TGCACCACCAACTGCTTAG	AGAGGCAGGGATGATGTTC

Table 2. The sequences of siRNAs

Target	Sequences
siRNA-control	S: UUCUCCGAACGUGUCACGUTT AS: ACGUGACACGUUCGGAGAATT
siRNA-UBXN1	S: GCUGACGGCUCUUGAGAGUTT AS: ACUCUCAAGAGCCGUCAGCTT
OE-UBXN1	F: ATGGCGGAGCTGACGGCTCT R: TCACAAAGCCCCTTCTCTTTGCAT

through the Genomic Data Commons (GDC) portal. We obtained Level 3/4 harmonized gene expression data (either HTSeq counts or FPKM) for colon adenocarcinoma (TCGA-COAD) and rectal adenocarcinoma (TCGA-READ), along with the corresponding clinical and survival information, via GDC data transfer tools or dedicated R packages. We merged and re-annotated the samples to create a unified CRC cohort.

Construction of protein-protein interaction networks and identification of hub genes

Protein-protein interaction (PPI) networks were constructed using the STRING database. The target gene was submitted to the database with the species specified as "Homo sapiens", and only interactions with high confidence were retained. The resulting network data were then imported into Cytoscape for visualization and topological analysis, with key topological properties such as node degree, betweenness centrality, and clustering coefficient calculated. Finally, based on higher centrality or node degree metrics, potential hub genes were identified for further investigation.

Cell culture

The CRC cell lines HCT16 and SW480, which are confirmed to be mycoplasma-free, were cultivated in high-glucose DMEM with serum. Cells were kept at 37°C in a 5% CO_2 incubator, and the medium was refreshed every 48 hours. After reaching 80% of the confluence, cells were subcultured for experimental purposes.

Reagents

The design and synthesis of all siRNAs and plasmids were carried out by OBiO Biotechnology Co., Ltd. (Shanghai, China). Antibodies p-NF-κB (ab8629), NF-κB (ab16502), GADPH (ab18-1602), and UBXN1 (ab151723) were acquired from Abcam (Cambridge, UK). RT-PCR primers were purchased from RiboBio (Guangzhou, China). Transwell inserts (354480) and Matrigel were obtained from Corning Inc., and the apoptosis detection kit (BL110A) was purchased from BioSharp Biotechnology. Thermo Fisher Scientific purchased Lipofectamine™ 20 (168019). **Tables 1** and **2** contain detailed information about primers for RT-qPCR and transfection.

Transfection

Cells were seeded in culture plates or dishes and grown to 50-70% confluency at the time of transfection. For the transfection mixture, 1-4 μg DNA was diluted in Opti-MEM in a sterile 1.5-ml tube, while Lipofectamine 2000 was separately diluted in Opti-MEM (typically 1:50). The diluted solutions were combined, gently mixed, and incubated at room temperature for 5-10 min to allow formation of DNA-Lipofectamine complexes. The culture medium was replaced with fresh Opti-MEM, and the complexes were added to the cells. Cells were incubated at 37°C in a CO_2 incubator for 4-6 h. The transfection mixture was then removed and replaced with complete medium to allow recovery and normal growth.

RNA extraction and RT-qPCR

Total RNA was extracted using TRIzol reagent (Invitrogen). One microgram of total RNA was reverse transcribed into cDNA using the HiScript III First-Strand cDNA Synthesis Kit according to the manufacturer's instructions. Quantitative PCR was performed with ChamQ SYBR qPCR Master Mix on a LightCycler 480 systemt. Gene expression levels were normalized to *GAPDH* expression in the same reaction.

Western blot

Cells were lysed with RIPA buffer (Beyotime Biotechnology, China). Protein concentration

was determined using a BCA assay kit (Beijing Biological Technology). A total of 50 μg of protein was separated on 10% SDS-PAGE and transferred to a nitrocellulose membrane (EMD Millipore). Membranes were blocked with 5% BSA and washed three times with PBS, followed by incubation with primary antibodies overnight at 4°C. After washing with PBS, membranes were incubated with H&L secondary antibodies for 1 h at room temperature. Protein bands were detected using a Thermo Fisher ECL substrate kit. Band intensity was semi-quantified using Image-Pro Plus software (v6.0; Media Cybernetics).

Clonogenic assay

Cells were seed at low density and then allowed to attach overnight. After the treatment, cells were incubated for a long time to enable colony formation. Colonies were fixed with paraformaldehyde, stained with crystal violet, and counted manually to evaluate clonogenic survival.

Transwell

Cells were seed in Transwell inserts with serum-free medium. Adding medium with chemoattractants to the lower chamber induced migration. After the incubation process, non-migrated cells were removed, and then migrated cells were fixed and stained with crystal violet. For invasion assays, inserts were pre-coated with Matrigel, and other procedures were kept the same. Migrated cells were analyzed using a microscope.

Cell apoptosis assay

Cells were collected in EP tubes, then washed with PBS, and resuspended in incubation buffer. They were stained with 5 μL Annexin V-FITC/PI using apoptosis detection kit (Biosharp, 159045, China) for 15 minutes. After three washes, cells were analyzed with flow cytometry. FlowJo software was utilized to compute the apoptotic rate.

Results

Construction of CRC single-cell atlas

The study analyzed four single-cell transcriptomic cohorts from the GEO database, which is a widely used database to characterize the cellular composition of CRC tumor microenvironment: GSE132465, GSE14735, GSE178318, and GSE178341. The datasets comprise primary tumors, liver metases, and CRC samples gathered under varied treatment conditions. These datasets have facilitated the development of CRC single-cell atlases and the exploration of tumor microenvironment remodeling. During quality control, the datas removed low-quality cells, cells with unusually high mitochondrial gene expression, and doublets. After filtering, we retained 499,079 high-quality cells and 26,745 genes for further analysis.

Because the datasets were created with various sequencing platforms, sample loading methods, and preprocessing procedures, direct merging might cause a lot of batch effects that affect clustering accuracy and downstream analyses. To tackle this problem, we employed a multi-sample integration framework rooted in scVI, which utilizes conditional variational autoencoders (conditional VAE) to analyze and regress batch effects in the latent space while maintaing the biological variation present. The integrated data demonstrated a strong mixing and separation effect in a single UMAP space, allowing the identification of several transcriptionally distinct clusters via Leiden clustering (**Figure 1A**). In the integrated embedding space, classical marker genes like EPCAM, KRT8/19, CD3D, CD3E, CD8A, LYZ, LST1, S100A8/S100A9, MS4A1, MZB1, LUM, and DCN were analyzed via UMAP expression patterns (**Figure 1B-E**) and a marker-cell type dot plot (**Figure 1F**) to guide biological annotation of each cluster. Cells were finally categorized into seven main groups: T cells, Natural Killer cells (NK cells), B cells, plasma cells, myeloid cells, fibroblasts/cancer-associated fibroblasts (fibroblast/CAF), endothelial cells, and epithelial cells (**Figure 1G**). In the integrated CRC single-cell atlas, T/NK cells and myeloid cells are the main immune populations, flowed by tumor epithelial cells. The distribution of these seven cell types remained consistent across all datasets (**Figure 1H**), closely aligning with lineage distributions observed in recent single-cell and spatial transcriptomics studies of CRC [32]. These findings show that the CRC tumor microenvironment contains diverse immune cell populations. They also reveal a highly heterogeneous population of malignant epithelial cells.

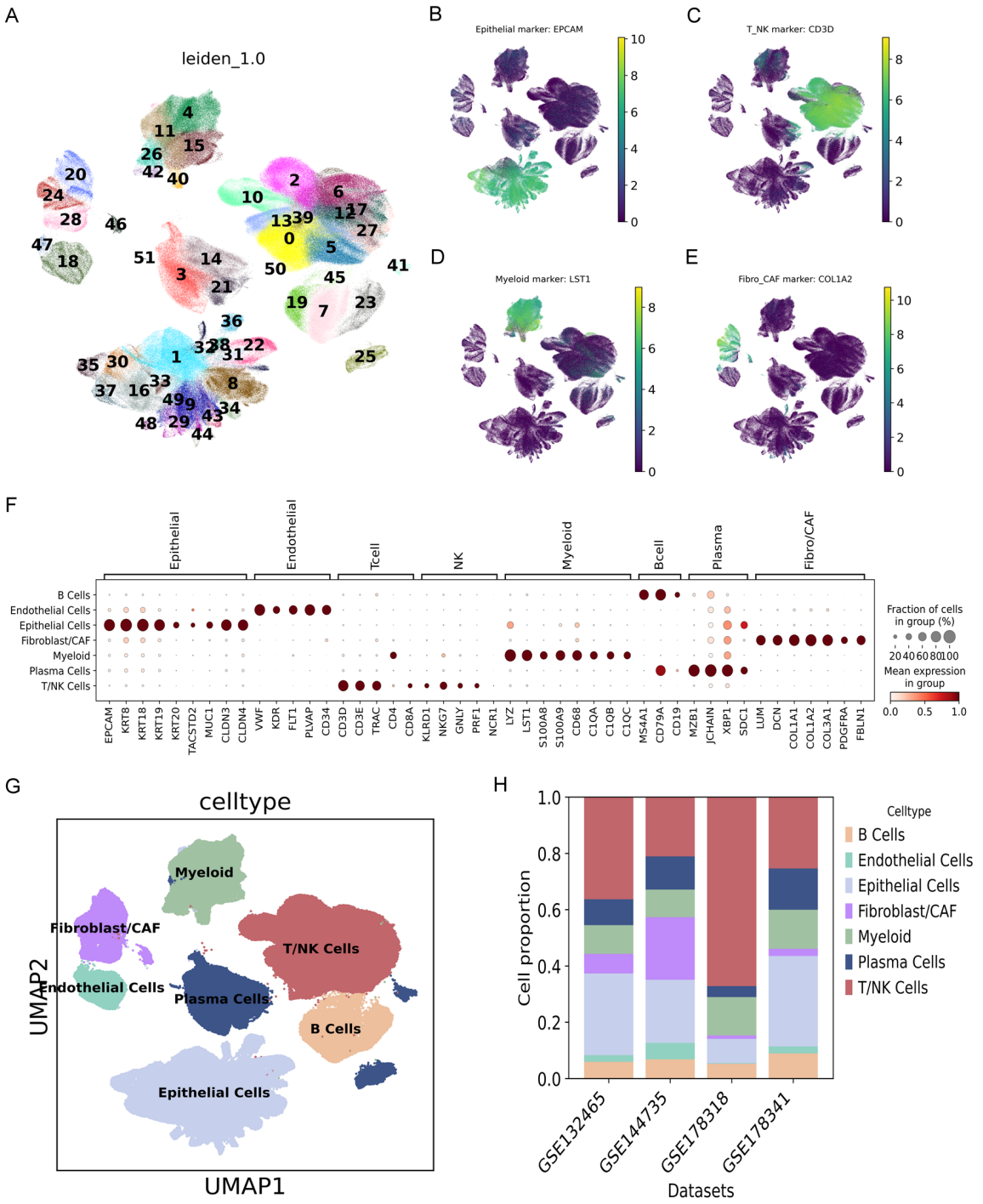


Figure 1. Construction method of the single-cell atlas of colorectal cancer (CRC). (A) Uniform Manifold Approximation and Projection (UMAP) distribution of integrated single-cell transcriptomic datasets from all four colorectal cancer studies, colored by unsupervised Leiden clustering. (B-E) Expression distribution of canonical lineage markers in UMAP space. CD8A and CD3D are highly expressed in the T/NK cell population (B, E), EPCAM is highly expressed in epithelial/tumor epithelial cells (C), and differential co-expression network (DCN) is highly expressed in fibroblasts/cancer-associated fibroblasts (D). (F) Dot plot of classical marker genes for seven major cell types, including B cells, endothelial cells, epithelial cells, fibroblasts/CAF, myeloid cells, plasma cells, and T/NK cells. The size of the dots represents the proportion of cells expressing the gene within that cell type, and the color intensity indicates the average expression level. (G) UMAP colored and annotated by finalized cell types, showing well-separated clusters including T/NK cells, B cells, plasma cells, myeloid cells, epithelial cells, fibroblasts/CAF, and endothelial cells. (H) The stacked bar chart shows cell composition across GEO datasets (GSE132465, GSE144735, GSE178318, and

GSE178341). Each bar represents the proportion of major cell types within a dataset. The chart indicates that the overall cellular composition remains consistent across different cohorts, with T/NK cells, myeloid cells, and epithelial cells constituting the predominant populations.

Construction of gene expression programs in epithelial cells

In colorectal cancer, when excluding the influence of stromal cells, epithelial cells are considered the primary cellular source for the development of malignant phenotypes. Therefore, the gene expression profiles of epithelial cells currently serve as the main focus for tumor subtype classification and prognosis-related studies [33]. Previous single-cell studies have shown that colorectal cancer epithelial cells can exhibit multiple functional states, including stem/progenitor-like states, secretory states, and others, with each state being associated with specific patterns of genomic instability. In this study, we extracted all epithelial cells from the integrated CRC single-cell atlas and performed preliminary processing using a unified normalization and batch correction framework (**Figure 2A**). Next, we used the cNMF algorithm to decompose epithelial transcriptional profiles and identify stable gene expression programs (GEPs). cNMF performs repeated NMF decomposition followed by consensus clustering, generating a robust gene-program matrix and cell-program "usage" while minimizing the effects of random initialization. It is widely used to identify reproducible transcriptional programs in tumor cells.

At the optimal K value, six major epithelial gene expression programs were identified (**Figure 2B**, **2C**). These include a "cell cycle module" enriched in cell cycle-related genes, such as proliferation markers and DNA replication genes (**Figure 2D**); a "metabolic module" enriched in glycolysis, oxidative phosphorylation, and amino acid metabolism pathways (**Figure 2E**); an "inflammatory stress module" enriched in NF-κB signaling, inflammatory mediators, and stress-response genes; and an "immune interaction module" containing genes related to antigen processing and presentation, interferon response, and immune checkpoints (**Figure 2F**). Similar associations between "gene expression programs and clinical outcomes" have been reported in CRC and other tumors, suggesting that tumor epithelial cells exhibit reproducible functional states [34]. These findings indicate that the transcriptional heterogeneity of CRC epithelial cells can be organized into a limited set of functional gene expression programs. This framework provides a basis for integrating genomic CNV and metabolic pathway activity to construct a "malignancy-metabolic characteristics" model in future studies.

Malignancy assessment and machine learning modeling based on metabolic pathways

CNV is a significant feature of malignant tumors, and it serves as a standard method for differentiating malignant from normal epithelial cells. Previous single-cell CRC studies with tools like inferCNV have shown that malignant epithelial cells usually have amplifications or deletions in chromosomal regions like 5q, 17p, and 18q, which shows a lot of genomic instability [35]. In this study, epithelial cells were integrated with B cells from the tumor microenvironment into a unified expression matrix, using B cells as a relatively genomically stable reference population to infer CNV in CRC epithelial cells. Compared to B cells, epithelial cells had significantly higher genome-wide CNV scores (**Figure 3A**), with strong chromosomal amplifications, suggesting greater malignancy and genomic instability (**Figure 3B**, **3C**).

Metabolic reprogramming is a key feature of malignant tumors. It supports the high proliferative requirements of cancer cells and is closely linked to DNA damage repair, redox homeostasis, and immune evasion [36]. To comprehensively evaluate the relationship between metabolic processes and epithelial cell malignancy, this study compiled 301 metabolism-related signaling pathways from public databases, such as Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome. Then, the AUCell algorithm was employed to compute pathway activity scores for individual epithelial cells. AUCell cells rank genes, and evaluates the enrichment of a given gene set compared to highly expressed genes. It is commonly employed for the assessment of single-cell pathway activity and is particularly well-suited for the quantitative analysis of metabolic and

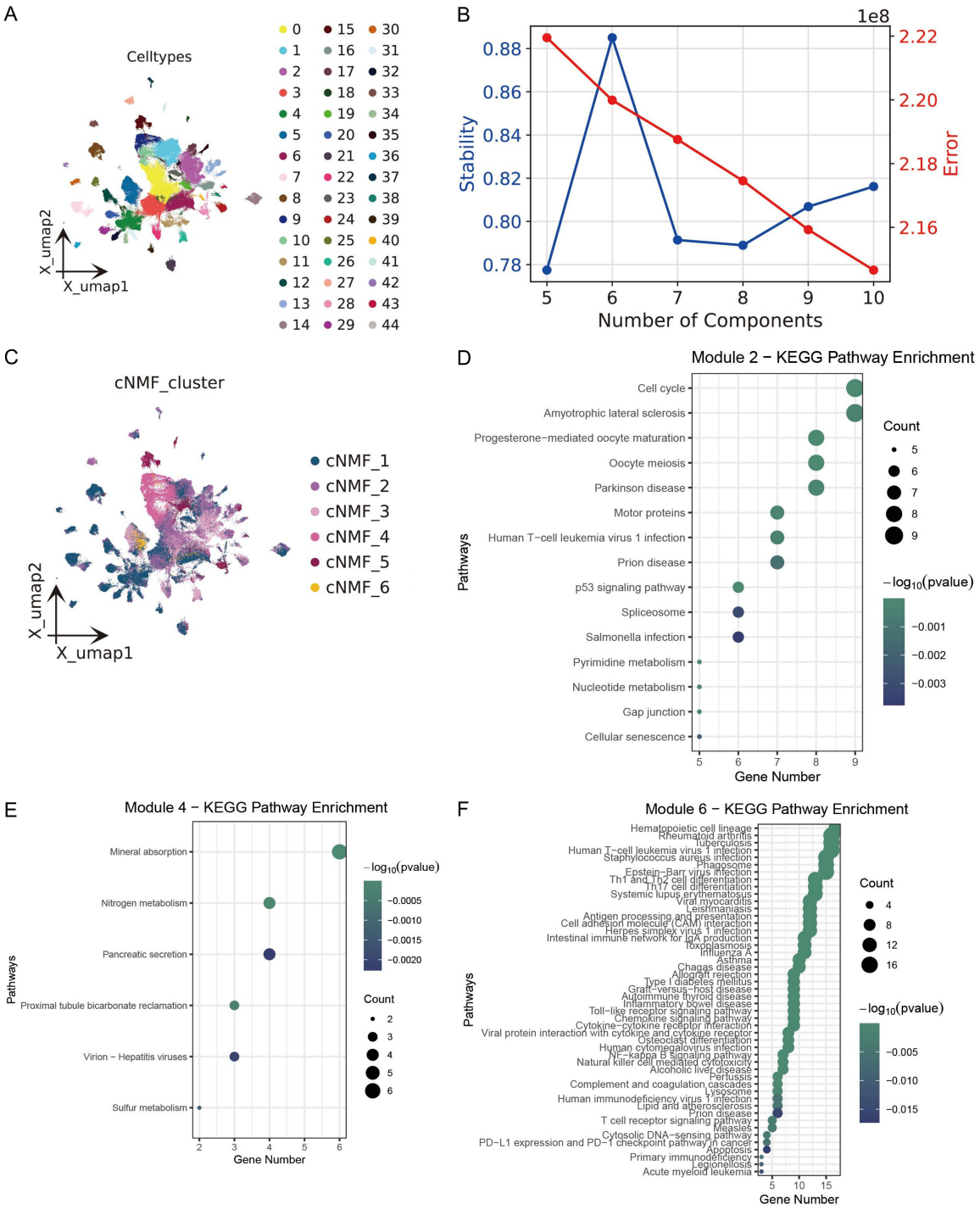
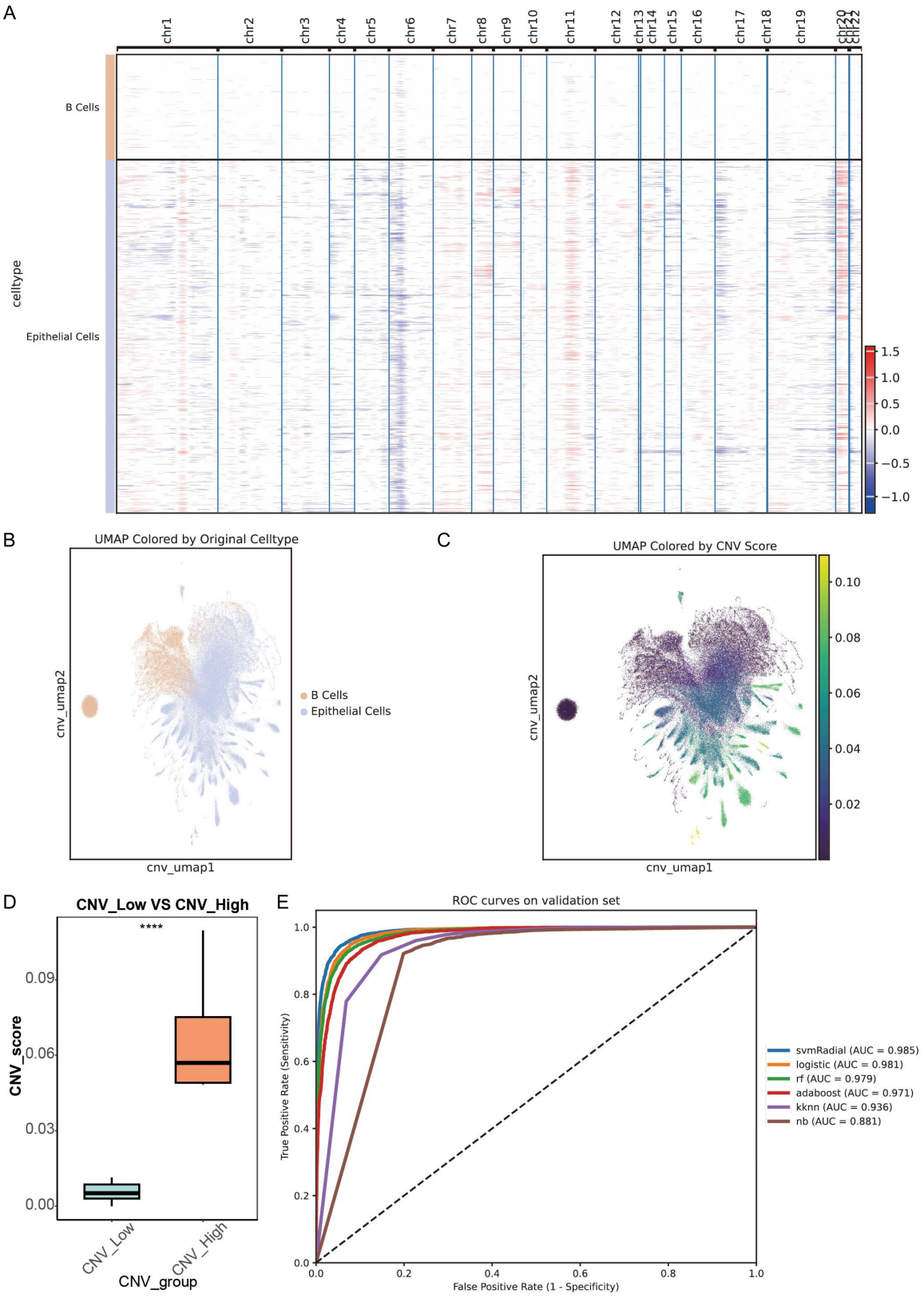


Figure 2. Gene expression programs and functional annotation of epithelial cells in colorectal cancer. A. UMAP plot showing the distribution of integrated epithelial cells. B. Selection process of the number of consensus non-negative matrix factorization (cNMF) components. The blue line represents program stability calculated from multiple decompositions, and the red line represents reconstruction error. K = 6 was selected as the optimal number of gene expression programs. C. The same epithelial cell population colored according to the dominant cNMF program. Six gene expression programs exhibit distinct patterns across different epithelial subpopulations. D. Bubble plot showing the Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis results for Module 2. Module 2 is significantly enriched in pathways related to the cell cycle, p53 signaling pathway, and RNA processing. E. KEGG enrichment analysis of Module 4. Module 4 is enriched in pathways such as mineral absorption, nitrogen metabolism, pancreatic secretion, and sulfur metabolism. F. KEGG enrichment analysis of Module 6. Module 6 is enriched in immune- and inflammation-related pathways, including T cell receptor signaling, cytokine-receptor interaction, and antigen processing and presentation.


A
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
B Cells
celltype
Epithelial Cells
1.5
1.0
0.5
0.0
−0.5
−1.0
B
UMAP Colored by Original Celltype
cnv_umap2
cnv_umap1
B Cells
Epithelial Cells
C
UMAP Colored by CNV Score
cnv_umap2
cnv_umap1
0.10
0.08
0.06
0.04
0.02
D
CNV_Low VS CNV_High
****
CNV_score
0.09
0.06
0.03
0.00
CNV_Low
CNV_High
CNV_group
E
ROC curves on validation set
True Positive Rate (Sensitivity)
False Positive Rate (1 - Specificity)
0.0
0.2
0.4
0.6
0.8
1.0
svmRadial (AUC = 0.985)
logistic (AUC = 0.981)
rf (AUC = 0.979)
adaboost (AUC = 0.971)
kknn (AUC = 0.936)
nb (AUC = 0.881)

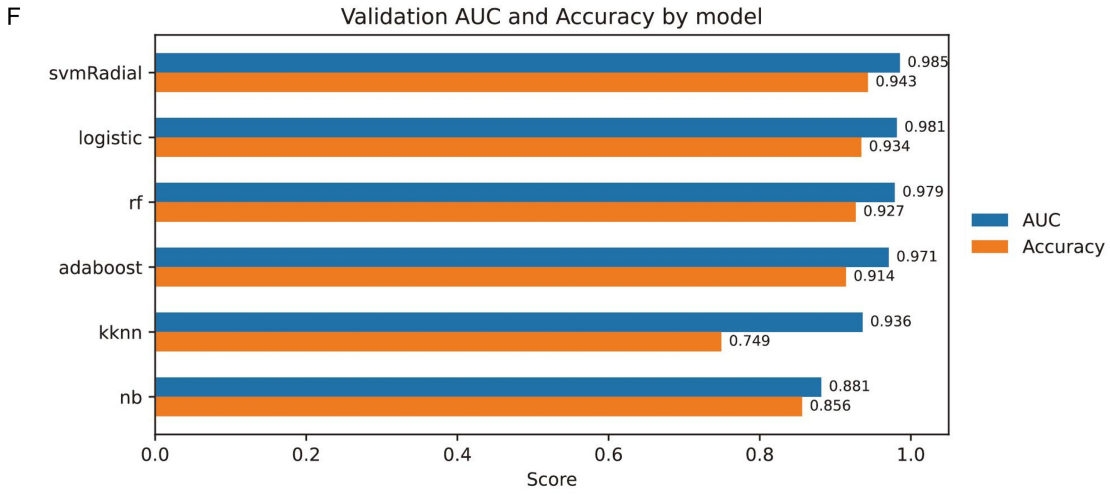


Figure 3. Evaluation of malignant potential in colorectal cancer epithelial cells based on copy number variation and metabolic pathway-based machine learning analysis. A. Heatmap of chromosomal copy number variations (CNVs), with B cells as the reference. The upper panel shows clustered B cells, and the lower panel shows clustered epithelial cells from CRC samples. B. UMAP plot of dimensionally reduced CNV data, colored according to the original cell types. C. Calibrated UMAP plot based on continuous CNV scores. D. Comparison between CNV_Low and CNV_High groups based on CNV scores (****$P < 0.0001$). E. Receiver operating characteristic (ROC) curves of machine learning models for predicting CNV_High status, including support vector machine with radial basis function kernel (svmRadial), logistic regression, random forest, AdaBoost, k-nearest neighbors (kknn), and naïve Bayes. F. ROC AUC values and accuracy of different models for predicting CNV_High status.

immune pathways. Using the extent of patient-level CNV amplification as a measure of malignancy, the CRC patients were categorized into two groups: those with high CNV amplification and those with low CNV amplification (**Figure 3D**). The activity of 301 metabolic pathways was summarized at the patient level. A variety of machine learning models, like Random Forest, AdaBoost, Support Vector Machine, Logistic Regression, K-Nearest Neighbors, and Naive Bayes, were developed based on these metabolic pathway features to predict the status of CNV. Model evaluation results indicate a strong predictive capability, with the model performing at the highest receiver operating characteristic (ROC) AUC of 0.985 (**Figure 3E**). These results indicate that the metabolic profile of CRC epithelial cells is strongly linked to their genomic instability and malignancy (**Figure 3F**). This finding reinforces the established relationship between metabolic reprogramming and genomic instability.

Screening of key metabolic pathways and candidate driver genes

In the machine learning modeling described above, the study further compared the predictive contributions of different metabolic pathways. The results showed that the 'Reactome Pathways Metabolism of Proteins' and 'TP53 Regulates Metabolic Genes' were the most important for distinguishing CNV_high from CNV_low patients (**Table 3**). The former includes key processes such as protein translation, folding, modification, and degradation, whereas the latter reflects transcriptional regulation of metabolic genes (related to glucose metabolism, lipid metabolism, and redox balance) by the classical tumor suppressor TP53. The activity scores of both pathways were significantly higher in CNV_high patients, suggesting persistent activation of protein metabolism and p53-mediated metabolic regulation in CRC with high genomic instability (**Figure 4A**, **4B**). High-CNV and low-CNV groups were stratified to identify genes associated with copy number variation status. Among these, the top five metabolism-related pathways based on functional enrichment were selected. Genes within these five pathways were extracted, and their activity was quantified by calculating the AUC. In addition, Pearson and Spearman correlation analyses were performed to evaluate the associations between gene expression and CNV scores. Through these analyses, a total of 220 candidate genes were identified, all of which exhibited an AUC greater than 0.75 and strong

Table 3. Candidate data table pathway summary

Feature	auc	auc_abs_dev	rf_importance	coef	coef_abs	auc_for_high	high_score_towards
Metabolism of Proteins_aucell	0.799935693	0.299935693	0.032305856	1.612579857	1.612579857	0.799935693	CNV_High
Metabolism_aucell	0.771073928	0.271073928	0.030490545	-0.651865622	0.651865622	0.771073928	CNV_High
TP53 Regulates Metabolic Genes_aucell	0.766447048	0.266447048	0.041920806	0.822582581	0.822582581	0.766447048	CNV_High
Metabolism of Nitric Oxide NOS3 Activation and Regulation_aucell	0.704170584	0.204170584	0.015699638	0.60691555	0.60691555	0.704170584	CNV_High

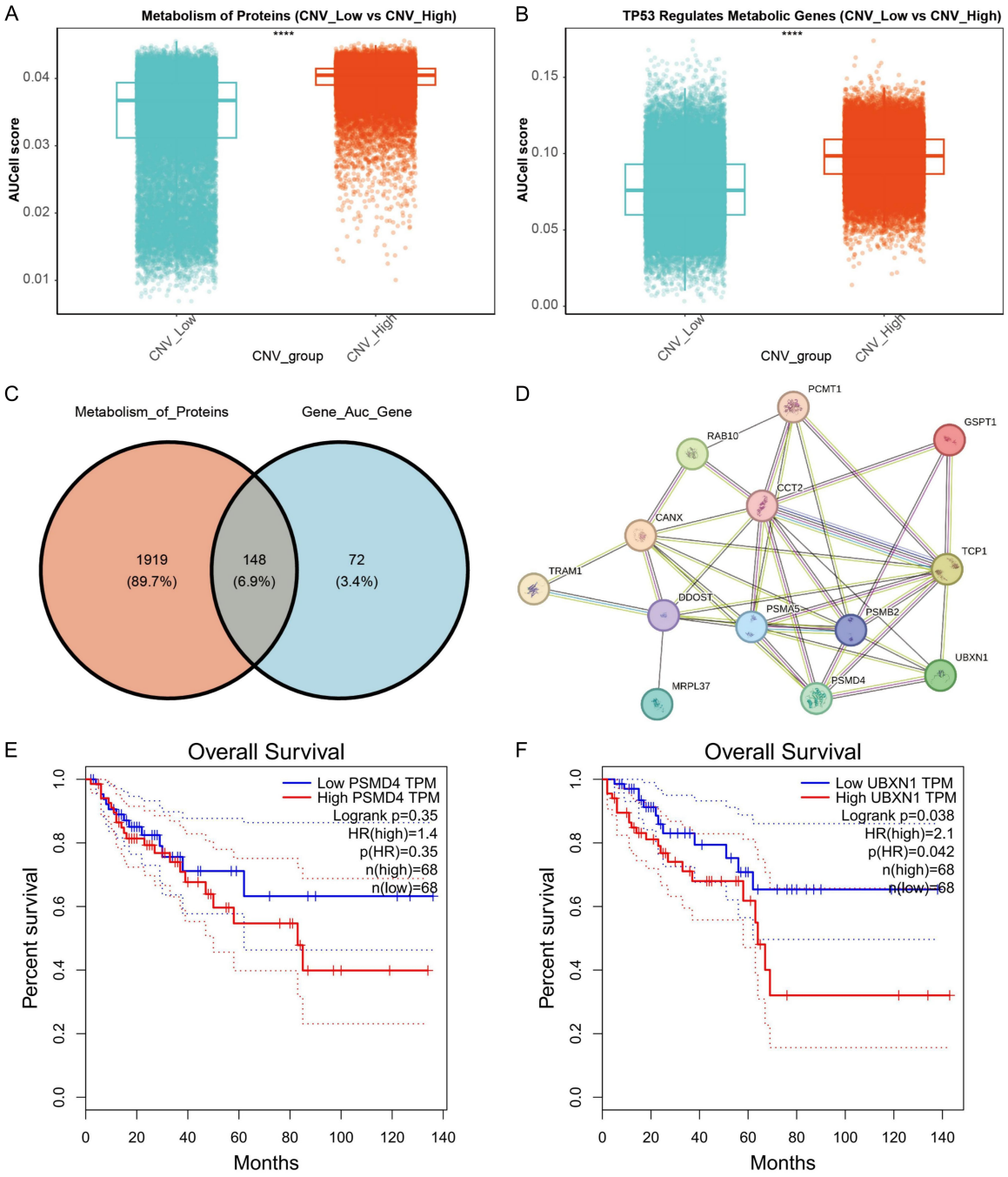


Figure 4. Metabolic programs linked to CNV status and identification of candidate prognostic genes in protein metabolism. (A) Scatter plot of gene set scores in epithelial cells from the CNV_Low and CNV_High groups, based on AUCell analysis of the Reactome "Metabolism of Proteins" pathway (****$P < 0.0001$). (B) Scatter plot of gene set scores in epithelial cells from the CNV_Low and CNV_High groups, based on the "TP53-regulated metabolic genes" pathway (****$P < 0.0001$). (C) Venn diagram showing the overlap between all genes in the "Metabolism of Proteins" pathway (n = 1,919) and the candidate gene set (n = 220) identified through machine learning and CNV score correlation analysis. (D) Protein-protein interaction (PPI) network of 13 key genes selected from 148 genes based on prognostic data from the TCGA COAD/READ cohort. The genes in the network include PSMD4, UBXN1, DDOST, and PSMA5, among others. (E, F). Kaplan-Meier curves of overall survival for patients in the TCGA COAD/READ cohort segmented byUBXN1 (F) and PSMD4 (E) expression levels (median grouping, each group n = 68), set 95% confidence interval, showing significant differences in overall survival between high and low expression groups (UBXN1 log-rank P = 0.038; PSMD4 log-rank P = 0.35), further supporting the close link between protein metabolism-related genes and CRC prognosis.

correlations with CNV scores as indicated by both Pearson and Spearman coefficients. These 220 genes were therefore defined as key genes associated with genomic instability in CRC. Subsequently, the intersection between these 220 genes and genes related to the protein metabolism pathway was determined, yielding 148 overlapping genes (**Figure 4C**).

Functional enrichment analysis of the 148 genes revealed that they were primarily associated with pathways related to the ubiquitin-proteasome system, protein processing and degradation, and endoplasmic reticulum stress. These pathways are highly consistent with the biological characteristics of tumor cells, which adapt to rapid proliferation and external stress by enhancing protein turnover. To assess their clinical relevance, we combined the colon cancer (COAD) and rectal cancer (READ) cohorts from TCGA. We analyzed the expression of the 148 genes in relation to overall survival in CRC patients. This analysis identified 13 genes that were significantly associated with prognosis (**Table 4**). A PPI network constructed from these 13 genes showed that PSMD4 and UBXN1 occupy central positions with high connectivity (**Figure 4D-F**). PSMD4 is a ubiquitin receptor subunit of the 26S proteasome regulatory particle that recognizes ubiquitinated substrates and has been reported to promote malignant progression and drug resistance in several cancers. In CRC, PSMD4 may serve as a potential target for inhibiting tumor invasiveness and improving prognosis in patients with altered cNrf2 expression [37]. UBXN1 is a classical negative regulator of NF-κB. Recent studies show that in CRC, SUB1 promotes the ubiquitination and degradation of UBXN1 by increasing expression of the E3 ubiquitin ligase UBR5, thereby sustaining NF-κB signaling and driving tumor metastasis and malignant progression [38]. In this cohort, high UBXN1 expression was significantly associated with poorer survival in CRC patients. This pattern may reflect a complex regulatory state in which UBXN1 is transcriptionally upregulated but suppressed at the protein level or functionally by the ubiquitin-proteasome system, consistent with the reported UBXN1-NF-κB pathway.

UBXN1 inhibits the NF-κB signaling pathway to promote malignant transformation in CRC

To further investigate the functional role of UBXN1 in CRC progression, the study performed overexpression and knockdown experiments in two CRC cell lines, HCT116 and SW480. Post-infection with siUBXN1 markedly reduced *UBXN1* expression compared with siNC, whereas transfection with a *UBXN1* OE plasmid significantly increased *UBXN1* expression relative to the empty vector (**Figure 5A**). Analysis of the downstream NF-κB pathway showed that UBXN1 knockdown increased p-NF-κB protein levels, whereas *UBXN1* overexpression reduced p-NF-κB expression (**Figure 5B**), indicating that UBXN1 negatively regulates NF-κB signaling, consistent with previous reports. Correspondingly, inflammatory factors were positively correlated with the activation status of the NF-κB pathway (**Figure 5C**).

This study also examined the relationship between *UBXN1* expression and the malignant phenotype of CRC cells. *UBXN1* knockdown significantly reduced colony formation, migration, invasion, and anti-apoptotic capacity in HCT-116 and SW480 cells, whereas *UBXN1* overexpression markedly enhanced these properties (**Figure 5D-F**), supporting the oncogenic role of UBXN1 in CRC progression. To determine whether these effects depend on NF-κB signaling, an NF-κB pathway inhibitor was applied to UBXN1 KD cell lines. The inhibitor suppressed NF-κB activation in HCT116 cells (Supplementary Figure 1A). The results revealed that inhibition of the NF-κB signaling pathway counteracted the suppression of malignant phenotypes induced by UBXN-1 knockdown (Supplementary Figure 1B, 1C). This finding confirmed that the regulatory effect of UBXN-1 on CRC phenotypes is dependent on the NF-κB signaling pathway.

Discussion

CRC exhibits pronounced intratumoral heterogeneity, with persistent genomic instability and marked metabolic reprogramming being among its hallmark features [39]. To systematically elucidate the intrinsic relationship between these two processes, the present study integrated large-scale single-cell transcriptomic data and performed functional experimental validation. The results demonstrated that the transcriptional heterogeneity of CRC epithelial cells can be categorized into six stable gene expression programs, among which metabolism-related programs show the strongest association with malignant states defined by copy number variations (CNVs) [40]. By inte-

Table 4. Correlation between copy number variation (CNV) and area-under-the-curve (AUC) at the gene level for the first 13 genes

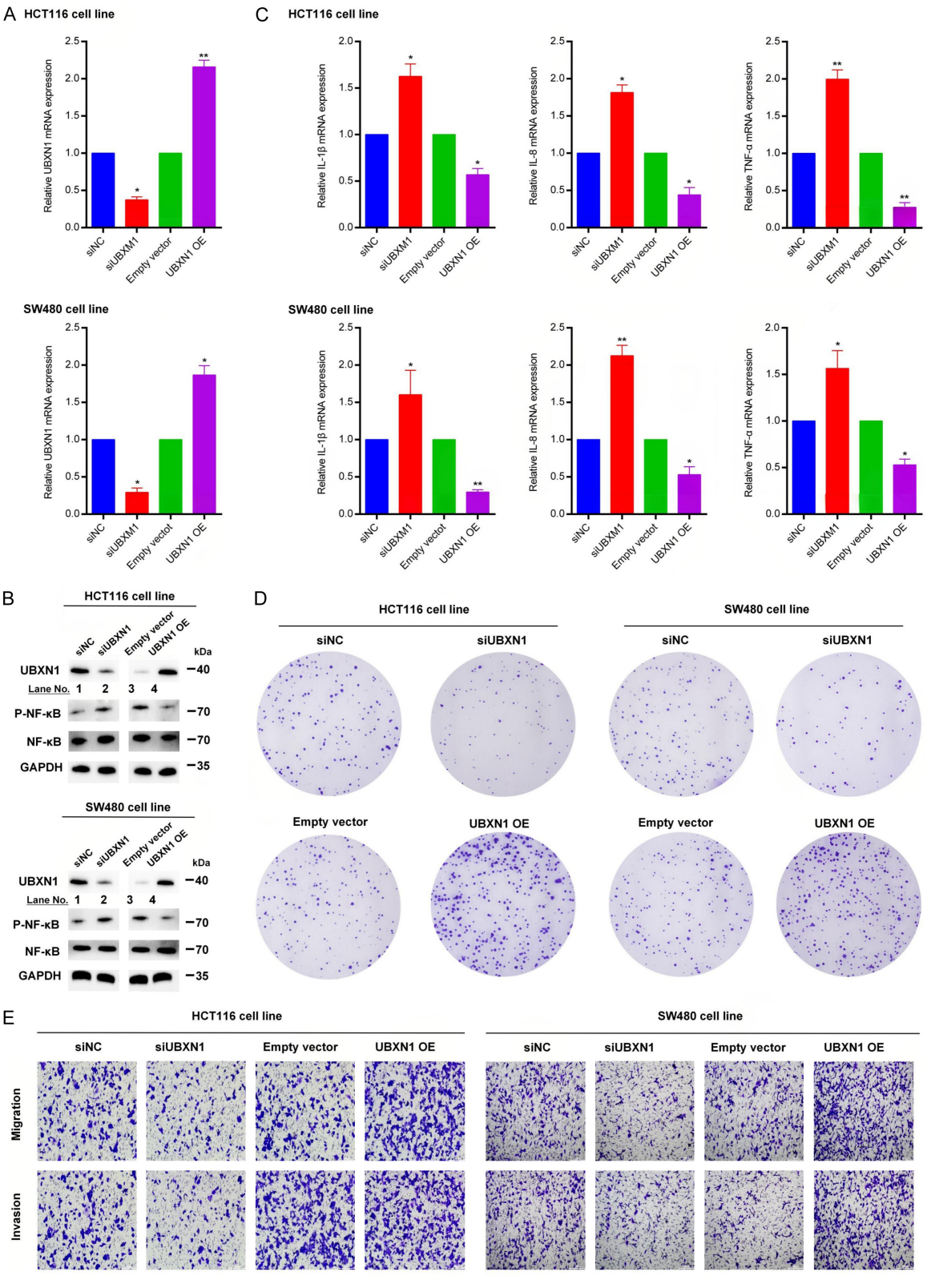
Gene	The AUC value of gene-predicted CNV	Correlation between genes and CNV_score under the Pearson algorithm	Correlation between genes and cnv_score under the Spearman algorithm
TRAM1	0.784561675	0.439449235	0.41384412
CANX	0.767593825	0.460094767	0.372633186
RAB10	0.755012291	0.393513141	0.34351339
DDOST	0.783653623	0.429421661	0.396612254
MRPL37	0.774735376	0.40814547	0.395168445
PCMT1	0.770322275	0.403514401	0.393435421
CCT2	0.790926129	0.474750003	0.431740035
PSMA5	0.759778787	0.404042525	0.350892311
PSMB2	0.754337604	0.407664636	0.346160305
PSMD4	0.773775353	0.433867627	0.396818915
GSPT1	0.777922964	0.435238051	0.392222214
TCP1	0.767500247	0.430387518	0.39339189
UBXN1	0.752452855	0.416339941	0.355408553

grating multiple datasets, we constructed a comprehensive single-cell transcriptomic atlas of CRC encompassing nearly 500,000 cells. This atlas not only reveals substantial epithelial heterogeneity and diverse functional states but also delineates the potential evolutionary trajectories of tumor cells.

InferCNV-based analysis indicated that malignant epithelial cells universally harbor large-scale chromosomal structural variations, suggesting that genomic instability is a key driver of transcriptional diversity. Further single-cell metabolic pathway activity analysis revealed a close correlation between the degree of metabolic reprogramming and the level of genomic instability, implying that metabolic states may serve as functional indicators reflecting tumor malignancy. Notably, a machine learning model constructed solely based on metabolic pathway activity features effectively distinguished patients with high CNV from those with low CNV, further highlighting the critical role of metabolic characteristics in representing genomic stress and tumor aggressiveness.

Among the metabolic pathways analyzed, protein metabolism and TP53-related pathways contributed most significantly to the prediction of malignant states. Protein metabolism supports rapid cancer cell proliferation and adaptation to stress by maintaining proteostasis [41]. Beyond its classical role in cell cycle regulation, TP53 also functions as a key regulator of metabolic reprogramming, coupling genomic stress signals with metabolic processes such as glycolysis and oxidative phosphorylation [42]. Abnormal regulation of the TP53 pathway may therefore facilitate tumor cell survival and proliferation under conditions of high genomic instability.

Through integrative multi-omics analysis and validation using The Cancer Genome Atlas (TCGA) cohort, we ultimately identified 13 genes significantly associated with patient prognosis that were concurrently linked to metabolic dysregulation and CNV status. Within the protein-protein interaction network constructed from these genes, PSMD4 and UBXN1 were identified as central hub nodes. PSMD4, a key component of the 26S proteasome, promotes tumor progression by regulating the degradation of tumor suppressor proteins and modulating NF-κB signaling [43]. UBXN1, as a cofactor of the p97/VCP complex, is involved in protein quality control mediated by the ubiquitin-proteasome system [44]. Although UBXN1 is generally considered a negative regulator of NF-κB signaling [45], our functional experiments revealed that, in CRC, *UBXN1* overexpression suppresses NF-κB phosphorylation while paradoxically promoting cell proliferation, migration, invasion, and resistance to apoptosis. This seemingly contradictory finding suggests that UBXN1 may exert context-dependent and com-


A
HCT116 cell line
Relative UBXN1 mRNA expression
SW480 cell line
Relative UBXN1 mRNA expression
siNC
siUBXM1
Empty vector
UBXN1 OE
C
HCT116 cell line
Relative IL-1β mRNA expression
Relative IL-8 mRNA expression
Relative TNF-α mRNA expression
SW480 cell line
Relative IL-1β mRNA expression
Relative IL-8 mRNA expression
Relative TNF-α mRNA expression
B
HCT116 cell line
siNC
siUBXN1
Empty vector
UBXN1 OE
kDa
UBXN1
Lane No. 1 2 3 4
P-NF-κB
NF-κB
GAPDH
40
70
70
35
SW480 cell line
D
HCT116 cell line
SW480 cell line
siNC
siUBXN1
Empty vector
UBXN1 OE
E
HCT116 cell line
SW480 cell line
siNC
siUBXN1
Empty vector
UBXN1 OE
Migration
Invasion

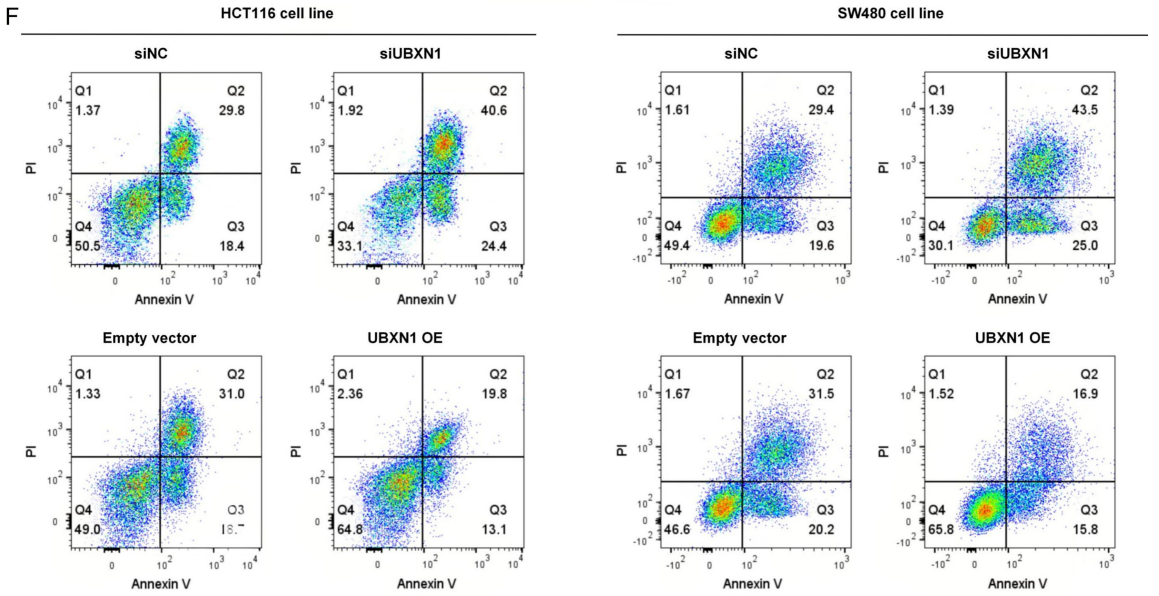


Figure 5. UBXN1 suppresses NF-κB signaling pathway to promote colorectal cancer progression. A. In HCT116 and SW480 cells, transfection with siUBXN1 inhibits *UBXN1* gene expression, while transfection with *UBXN1* OE (overexpression) enhances UBXN1 expression, two groups were analyzed using the independent samples t-test statistical method, in addition, Bonferroni correction was applied to all multiple comparisons ($^{*}P < 0.05$, $^{**}P < 0.01$). B. Expression levels of NF-κB signaling pathway following *UBXN1* knockdown and OE. C. Expression levels of inflammatory factors IL-8, IL-1β, and TNF-α following *UBXN1* knockdown and OE, the statistical method was the same as above ($^{*}P < 0.05$, $^{**}P < 0.01$). D. Differences in colony formation ability of HCT116 and SW480 cells after *UBXN1* knockdown and OE. E. Differences in migration and invasion capabilities of HCT116 and SW480 cells detected by Transwell assay following *UBXN1* knockdown and OE. F. Differences in anti-apoptotic capacity of HCT116 and SW480 cells in response to 5-fluorouracil after *UBXN1* knockdown and OE.

plex functions through feedback regulation within the proteostasis network, consistent with recent studies implicating UBXN1 in metabolic stress responses and tumor progression [46, 47]. Nevertheless, several limitations of this study should be acknowledged. First, although multiple single-cell datasets were integrated, the lack of spatial transcriptomic data limits the ability to resolve direct cell-cell interactions. Second, the precise regulatory mechanisms among UBXN1, NF-κB, and TP53-mediated metabolic adaptation require further in vivo validation.

In summary, this study systematically elucidates the close relationship between metabolic reprogramming and genomic instability in CRC and provides preliminary evidence that UBXN1 may serve as a key mediator of this association and a driver of malignant progression. These findings underscore the potential value of targeting protein metabolism, the TP53 pathway, and the ubiquitin-proteasome network in the precise treatment of colorectal cancer.

Acknowledgements

The authors would like to express their gratitude to EditSprings (https://www.editsprings.com) for the expert linguistic services provided. This work was supported by the 12th Five-Year Plan Key Clinical Specialty Program (Finance and Society [2013] No. 239), the Anhui Province 14th Five-Year Plan Clinical Key Specialty Construction Recommendation Program (Wanwei Yimi [2022] No. 105), and the 2026 Ministry of Education Anhui Advanced Research Institute University-Enterprise Joint Talent Training and Scientific Research Project (a collaborative project between Anhui University of Chinese Medicine and the Space Research Institute).

Disclosure of conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Address correspondence to: Qingsheng Yu, Hui Peng, Lixiang Li and Yi Shen, Institute of Chinese Medicine Surgery, Anhui Academy of Chinese Medicine, Hefei 230031, Anhui, P. R. China. Tel: +86-13866141906; E-mail: qsy6312@163.com (QSY); Tel: +86-15155916472; E-mail: ph136382333@126.com (HP); Tel: +86-13865200766; E-mail: llxwxm2022@163.com (LXL); Tel: +86-15395100404; E-mail: Shenyi612@126.com (YS)

References

[1] Li J, Ma X, Chakravarti D, Shalapour S and DePinho RA. Genetic and biological hallmarks of colorectal cancer. Genes Dev 2021; 35: 787-820.

[2] Dekker E, Tanis PJ, Vleugels JLA, Kasi PM and Wallace MB. Colorectal cancer. Lancet 2019; 394: 1467-1480.

[3] Abedizadeh R, Majidi F, Khorasani HR, Abedi H and Sabour D. Colorectal cancer: a comprehensive review of carcinogenesis, diagnosis, and novel strategies for classified treatments. Cancer Metastasis Rev 2024; 43: 729-753.

[4] Mo S, Tang P, Luo W, Zhang L, Li Y, Hu X, Ma X, Chen Y, Bao Y, He X, Fu G, Xu X, Rao X, Li X, Guan R, Chen S, Deng Y, Lv T, Mu P, Zheng Q, Wang S, Liu F, Li Y, Sheng W, Huang D, Hu C, Gao J, Zhang Z, Cai S, Clevers H, Peng J and Hua G. Patient-derived organoids from colorectal cancer with paired liver metastasis reveal tumor heterogeneity and predict response to chemotherapy. Adv Sci (Weinh) 2022; 9: e2204097.

[5] Gaiani F, Marchesi F, Negri F, Greco L, Malesci A, de'Angelis GL and Laghi L. Heterogeneity of colorectal cancer progression: molecular gas and brakes. Int J Mol Sci 2021; 22: 5246.

[6] Pino MS and Chung DC. The chromosomal instability pathway in colon cancer. Gastroenterology 2010; 138: 2059-2072.

[7] Molparia B, Oliveira G, Wagner JL, Spencer EG and Torkamani A. A feasibility study of colorectal cancer diagnosis via circulating tumor DNA derived CNV detection. PLoS One 2018; 13: e0196826.

[8] Sato K, Masuda T, Hu Q, Tobo T, Gillaspie S, Niida A, Thornton M, Kuroda Y, Eguchi H, Nakagawa T, Asano K and Mimori K. Novel oncogene 5MP1 reprograms c-Myc translation initiation to drive malignant phenotypes in colorectal cancer. EBioMedicine 2019; 44: 387-402.

[9] Chan TL, Curtis LC, Leung SY, Farrington SM, Ho JW, Chan AS, Lam PW, Tse CW, Dunlop MG, Wyllie AH and Yuen ST. Early-onset colorectal cancer with stable microsatellite DNA and near-diploid chromosomes. Oncogene 2001; 20: 4871-4876.

[10] Madunić K, Zhang T, Mayboroda OA, Holst S, Stavenhagen K, Jin C, Karlsson NG, Lageveen-Kammeijer GSM and Wuhrer M. Colorectal cancer cell lines show striking diversity of their O-glycome reflecting the cellular differentiation phenotype. Cell Mol Life Sci 2021; 78: 337-350.

[11] Li TT and Zhu HB. LKB1 and cancer: the dual role of metabolic regulation. Biomed Pharmacother 2020; 132: 110872.

[12] Robert A, Crottès D, Bourgeais J, Gueguen N, Chevrollier A, Dumas JF, Servais S, Domingo I, Chadet S, Sobilo J, Hérault O, Lecomte T, Vandier C, Raoul W and Guéguinou M. MICU2 upregulation enhances tumor aggressiveness and metabolic reprogramming during colorectal cancer development. PLoS Biol 2024; 22: e3002854.

[13] Aden D, Sureka N, Zaheer S, Chaurasia JK and Zaheer S. Metabolic reprogramming in cancer: implications for immunosuppressive microenvironment. Immunology 2025; 174: 30-72.

[14] Zhang F, Ma Y, Li D, Wei J, Chen K, Zhang E, Liu G, Chu X, Liu X, Liu W, Tian X and Yang Y. Cancer associated fibroblasts and metabolic reprogramming: unraveling the intricate crosstalk in tumor evolution. J Hematol Oncol 2024; 17: 80.

[15] Raggi C, Taddei ML, Rae C, Braconi C and Marra F. Metabolic reprogramming in cholangiocarcinoma. J Hepatol 2022; 77: 849-864.

[16] Yan L, Wu M, Wang T, Yuan H, Zhang X, Zhang H, Li T, Pandey V, Han X, Lobie PE and Zhu T. Breast cancer stem cells secrete MIF to mediate tumor metabolic reprogramming that drives immune evasion. Cancer Res 2024; 84: 1270-1285.

[17] Liu S, Zhang X, Wang W, Li X, Sun X, Zhao Y, Wang Q, Li Y, Hu F and Ren H. Metabolic reprogramming and therapeutic resistance in primary and metastatic breast cancer. Mol Cancer 2024; 23: 261.

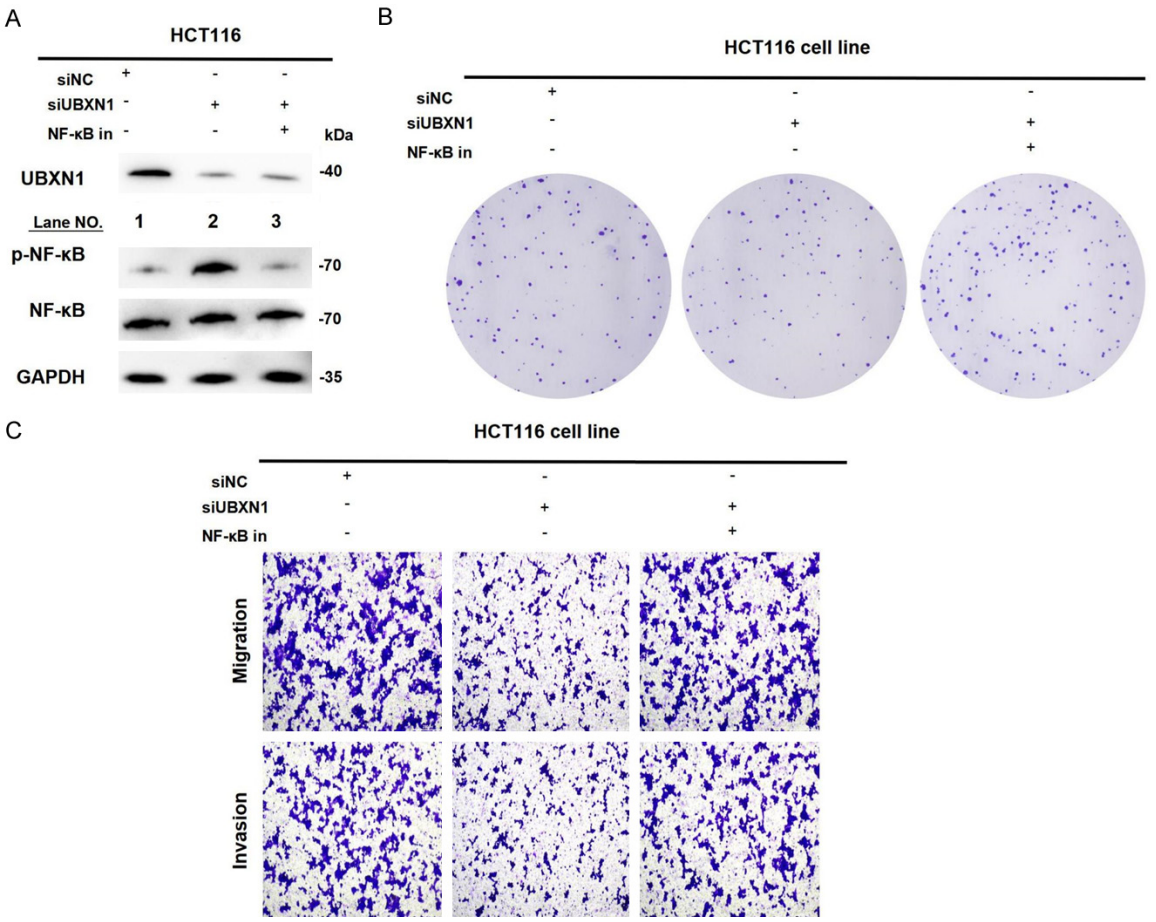
[18] Ju HQ, Lin JF, Tian T, Xie D and Xu RH. NADPH homeostasis in cancer: functions, mechanisms and therapeutic implications. Signal Transduct Target Ther 2020; 5: 231.

[19] Ye M, Chen J, Lu F, Zhao M, Wu S, Hu C, Yu P, Kan J, Bai J, Tian Y and Tang Q. Down-regulated FTO and ALKBH5 co-operatively activates FOXO signaling through m6A methylation modification in HK2 mRNA mediated by IGF2BP2 to enhance glycolysis in colorectal cancer. Cell Biosci 2023; 13: 148.

[20] Chen Z, Hong W, Li B, He D, Ren Z, Cai M, Cheng Y, Liu J, Xu E, Du Y, Dong Y, Cai S, Shi Q, Qi Z and Zhong Y. HDAC2 promotes colorectal tumorigenesis by triggering dysregulation of lipid metabolism through YAP1. Cell Signal 2025; 128: 111627.

[21] Lin C, Li Z, Zhu X, Zhou W, Lu X, Zheng J and Lin J. Abnormal β-hydroxybutyrylation modification of ARG1 drives reprogramming of arginine metabolism to promote the progression of colorectal cancer. Adv Sci (Weinh) 2025; 12: e02402.
[22] Pakiet A, Kobiela J, Stepnowski P, Sledzinski T and Mika A. Changes in lipids composition and metabolism in colorectal cancer: a review. Lipids Health Dis 2019; 18: 29.
[23] Bian S, Hou Y, Zhou X, Li X, Yong J, Wang Y, Wang W, Yan J, Hu B, Guo H, Wang J, Gao S, Mao Y, Dong J, Zhu P, Xiu D, Yan L, Wen L, Qiao J, Tang F and Fu W. Single-cell multiomics sequencing and analyses of human colorectal cancer. Science 2018; 362: 1060-1063.
[24] Qian J, Olbrecht S, Boeckx B, Vos H, Laoui D, Etlioglu E, Wauters E, Pomella V, Verbandt S, Busschaert P, Bassez A, Franken A, Bempt MV, Xiong J, Weynand B, van Herck Y, Antoranz A, Bosisio FM, Thienpont B, Floris G, Vergote I, Smeets A, Tejpar S and Lambrechts D. A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. Cell Res 2020; 30: 745-762.
[25] Li Y, Hu X, Lin R, Zhou G, Zhao L, Zhao D, Zhang Y, Li W, Zhang Y, Ma P, Ren H, Liao X, Niu P, Wang T, Zhang X, Wang W, Gao R, Li Q, Church G, He J and Chen Y. Single-cell landscape reveals active cell subtypes and their interaction in the tumor microenvironment of gastric cancer. Theranostics 2022; 12: 3818-3833.
[26] Xu C, Pan Y, Zhang S, Lin M, Liu H and Yu Z. Dynamic remodelling of epithelial plasticity in colorectal cancer from single-cell and spatially resolved perspectives. J Transl Med 2025; 23: 1341.
[27] Hu H, Wu J and Lu Y. Unveiling metabolic pathway dysregulation in the malignant transformation of polyps to colorectal cancer: a single-cell analysis of epithelial cell trajectories. Int J Surg 2026; 112: 3071-3089.
[28] You Y, Tian L, Su S, Dong X, Jabbari JS, Hickey PF and Ritchie ME. Benchmarking UMI-based single-cell RNA-seq preprocessing workflows. Genome Biol 2021; 22: 339.
[29] Xi NM and Li JJ. Benchmarking computational doublet-detection methods for single-cell RNA sequencing data. Cell Syst 2021; 12: 176-194, e176.
[30] Nguyen HCT, Baik B, Yoon S, Park T and Nam D. Benchmarking integration of single-cell differential expression. Nat Commun 2023; 14: 1570.
[31] Zhang Y, Liu W and Duan J. On the core segmentation algorithms of copy number variation detection tools. Brief Bioinform 2024; 25: bbae022.
[32] Qi J, Sun H, Zhang Y, Wang Z, Xun Z, Li Z, Ding X, Bao R, Hong L, Jia W, Fang F, Liu H, Chen L, Zhong J, Zou D, Liu L, Han L, Ginhoux F, Liu Y, Ye Y and Su B. Single-cell and spatial analysis reveal interaction of FAP(+) fibroblasts and SPP1(+) macrophages in colorectal cancer. Nat Commun 2022; 13: 1742.
[33] Fehm T, Sagalowsky A, Clifford E, Beitsch P, Saboorian H, Euhus D, Meng S, Morrison L, Tucker T, Lane N, Ghadimi BM, Heselmeyer-Haddad K, Ried T, Rao C and Uhr J. Cytogenetic evidence that circulating epithelial cells in patients with carcinoma are malignant. Clin Cancer Res 2002; 8: 2073-2084.
[34] Puleo F, Nicolle R, Blum Y, Cros J, Marisa L, Demetter P, Quertinmont E, Svrcek M, Elarouci N, Iovanna J, Franchimont D, Verset L, Galdon MG, Devière J, de Reyniès A, Laurent-Puig P, Van Laethem JL, Bachet JB and Maréchal R. Stratification of pancreatic ductal adenocarcinomas based on tumor and microenvironment features. Gastroenterology 2018; 155: 1999-2013, e1993.
[35] Gryfe R, Swallow C, Bapat B, Redston M, Gallinger S and Couture J. Molecular biology of colorectal cancer. Curr Probl Cancer 1997; 21: 233-300.
[36] An Y and Duan H. The role of m6A RNA methylation in cancer metabolism. Mol Cancer 2022; 21: 14.
[37] Lin PL, Chang JT, Wu DW, Huang CC and Lee H. Cytoplasmic localization of Nrf2 promotes colorectal cancer with more aggressive tumors via upregulation of PSMD4. Free Radic Biol Med 2016; 95: 121-132.
[38] Wang H, Chen W, Wang Y, Gao Y, Zhang Z, Mi S, Wang L and Xue M. SUB1 promotes colorectal cancer metastasis by activating NF-κB signaling via UBR5-mediated ubiquitination of UBXN1. Sci China Life Sci 2024; 67: 1199-1211.
[39] Zhou Z, Dong D, Yuan Y, Luo J, Liu XD, Chen LY, Wang G and Yin Y. Single cell atlas reveals multilayered metabolic heterogeneity across tumour types. EBioMedicine 2024; 109: 105389.
[40] Xiao J, Yu X, Meng F, Zhang Y, Zhou W, Ren Y, Li J, Sun Y, Sun H, Chen G, He K and Lu L. Integrating spatial and single-cell transcriptomics reveals tumor heterogeneity and intercellular networks in colorectal cancer. Cell Death Dis 2024; 15: 326.
[41] Chen Z, Xu J, Fang K, Jiang H, Leng Z, Wu H, Zhang Z, Wang Z, Li Z, Sun M, Zhao Z, Feng A, Zhang S, Chu Y, Ye L, Xu M, He L and Chen T. FOXC1-mediated serine metabolism reprogramming enhances colorectal cancer growth and 5-FU resistance under serine restriction. Cell Commun Signal 2025; 23: 13.
[42] Malki A, ElRuz RA, Gupta I, Allouch A, Vranic S and Al Moustafa AE. Molecular mechanisms of

colon cancer progression and metastasis: recent insights and advancements. Int J Mol Sci 2020; 22: 130.
[43] Li Y, Zhou Q, Shen J and Zhu L. Down-regulation of PSMD4 can attenuate autophagy, enhance the accumulation of intracellular ROS, and increase the sensitivity of epithelial ovarian cancer to carboplatin by inhibiting the NF-κB pathway. Transl Cancer Res 2021; 10: 4756-4772.
[44] Hu Y, O'Boyle K, Auer J, Raju S, You F, Wang P, Fikrig E and Sutton RE. Multiple UBXN family members inhibit retrovirus and lentivirus production and canonical NFκB signaling by stabilizing IκBα. PLoS Pathog 2017; 13: e1006187.
[45] Li Y, Liu X, Cui X, Tan Y, Wang Q, Wang Y, Xu C, Fang C and Kang C. LncRNA PRADX-mediated recruitment of PRC2/DDX5 complex suppresses UBXN1 expression and activates NF-κB activity, promoting tumorigenesis. Theranostics 2021; 11: 4516-4530.
[46] Chai RC, Chang YZ, Chang X, Pang B, An SY, Zhang KN, Chang YH, Jiang T and Wang YZ. YTHDF2 facilitates UBXN1 mRNA decay by recognizing METTL3-mediated m(6)A modification to activate NF-κB and promote the malignant progression of glioma. J Hematol Oncol 2021; 14: 109.
[47] Shi K, Chen Y, Gao T, Guo H, Fu X, Wu Y and Yu H. The oncogenic role of UBXN1 in gastric cancer is attributed to the METTL16-mediated m6A methylation and histone modifications. Cancer Med 2025; 14: e70772.



Supplementary Figure 1. UBXN1's regulation of crc progression is dependent on the NF-κB signaling pathway. A. In the HCT116 cell line, the activation of the NF-κB signaling pathway induced by transfection with siUBXN1 can be reversed by an NF-κB pathway inhibitor. B. Following transfection with siUBXN1, inhibition of the NF-κB pathway with an inhibitor enhances clonogenic colony formation ability compared to when the NF-κB pathway is activated. C. Following transfection with siUBXN1, inhibition of the NF-κB pathway with an inhibitor enhances migration and invasion capabilities compared to when the NF-κB pathway is activated.