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.

Spatial and single-cell transcriptomics decipher butyrate metabolic reprogramming and its impact on colorectal cancer progression and tumor microenvironment remodeling

Jiuyuan Fang
Peking Union Medical College Hospital
Tianqi Liu
Peking Union Medical College Hospital
Lulu Wen
17100560810qq.com

Xuan Wu Hospital of the Capital Medical University
Miao Qu
Xuan Wu Hospital of the Capital Medical University

Research Article

Keywords: Colorectal cancer, butyrate metabolism, spatial transcriptomics, single-cell transcriptomics, tumor microenvironment.

Posted Date: September 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10773427/v1

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

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Cancer on October 21st, 2026. See the published version at https://doi.org/10.1186/s12885-026-17211-8.

Abstract

Background

Colorectal cancer (CRC) is closely linked to dietary and microbial factors. Butyrate, a major gut microbial metabolite with histone deacetylase-inhibitory and immunomodulatory effects, contributes to epithelial homeostasis, but the spatial heterogeneity and clinical relevance of butyrate metabolism in CRC remain unclear.

Methods

Multi-algorithm single-cell metabolic scoring was used to quantify butyrate metabolism and define tumor epithelial subgroups, which were mapped spatially using cell2location. CellChat, Monocle, pySCENIC, inferCNV, PROGENy, and stLearn were applied to characterize lineage states, transcriptional regulation, signaling activity, and intercellular crosstalk. Ensemble machine learning was used for prognostic modeling, followed by molecular and functional validation.

Results

We identified malignant epithelial cells with low butyrate metabolism (LowBTAepis), which were associated with poor prognosis, early pseudotime states, increased stemness and proliferation, and distinct copy-number variation (CNV) patterns. Spatial analysis revealed preferential interactions between LowBTAepis and fibroblasts. HOXB8 emerged as a key transcriptional regulator of LowBTAepis. Sodium butyrate (NaB) suppressed CRC cell proliferation, migration, and invasion in a dose-dependent manner, accompanied by reduced HOXB8 expression, inhibition of AKT/mTOR signaling, and changes in apoptosis- and metastasis-related proteins. A butyrate metabolism-related model showed strong prognostic performance and was associated with an immunosuppressive tumor microenvironment.

Conclusions

LowBTAepis represents an aggressive CRC epithelial state characterized by stromal crosstalk and malignant transcriptional features. Butyrate may suppress CRC progression partly through the HOXB8/AKT/mTOR axis, supporting its potential relevance to metabolic intervention and patient stratification.

1 Background

Colorectal cancer (CRC) is one of the most commonly diagnosed malignancies worldwide and remains the second leading cause of cancer-related mortality. With 1.9 million new cases and nearly 935,000 deaths reported annually [1], its incidence continues to rise, particularly among younger populations.

Alarmingly, CRC cases in individuals younger than 50 are projected to increase by approximately 140% by 2030 [2], highlighting the urgent need for enhanced prevention and treatment strategies. Despite substantial advances in surgery, chemotherapy, targeted therapy, and immunotherapy, the prognosis for metastatic CRC remains poor, with a 5-year survival rate below 15%, largely because of therapeutic resistance, disease recurrence, and metastatic progression [3]. CRC is characterized by marked molecular heterogeneity, including recurrent alterations in APC, KRAS, and TP53, which affect the Wnt/β-catenin, RAS/MAPK, and p53 signaling pathways, respectively [4–6]. In addition, molecular features such as microsatellite instability-high (MSI-H), the CpG island methylator phenotype (CIMP), and chromosomal instability (CIN) are associated with distinct biological behaviors, clinical outcomes, and treatment responses [7–9]. Given the substantial heterogeneity and limited therapeutic efficacy in advanced CRC, a deeper understanding of its molecular mechanisms is imperative to identify novel oncogenic drivers and therapeutic vulnerabilities.

Short-chain fatty acids (SCFAs), pivotal microbial metabolites, are instrumental in preserving intestinal homeostasis. Among them, butyrate is predominantly synthesized in the colon via bacterial fermentation of dietary fiber, with Faecalibacterium prausnitzii representing one of the major butyrate-producing species [10]. Butyrate uptake by colonic epithelial cells occurs through monocarboxylate transporters (MCT1, SMCT1). It contributes to epithelial barrier integrity and regulates local inflammatory and immune responses [11]. In addition to its physiological functions, butyrate exerts tumor-suppressive effects in CRC by inhibiting tumor-cell proliferation and inducing apoptosis. These effects are mediated through histone deacetylase (HDAC) inhibition, ERK modulation, and metabolic reprogramming, collectively orchestrating a shift toward tumor suppression [12, 13]. The context-dependent effects of butyrate on normal and malignant colonic epithelial cells are commonly referred to as the butyrate paradox. Differentiated colonic epithelial cells efficiently oxidize butyrate as an energy substrate, whereas CRC cells preferentially depend on aerobic glycolysis. Consequently, butyrate is less efficiently metabolized in CRC cells and accumulates in the nucleus, where it inhibits HDAC activity and promotes cell-cycle arrest and apoptosis [12]. Previous studies have also suggested that butyrate may enhance the antitumor effects of irinotecan, 5-fluorouracil, and radiotherapy and may alleviate chemotherapy-associated intestinal mucosal injury [14, 15].

Our study unveils a pronounced suppression of butanoate metabolism in colorectal tumor epithelial cells, which correlates with unfavorable clinical outcomes. Spatial transcriptomic profiling further revealed that tumor epithelial cells with reduced butyrate metabolism exhibited enhanced activation of oncogenic pathways associated with cell-cycle progression, metastatic potential, and tumor hypoxia. These findings suggest that impaired butyrate metabolism is associated with a more aggressive tumor-cell state in CRC.

Cancer-associated fibroblasts (CAFs) are integral to the CRC microenvironment, fostering tumor initiation and metastasis through the production of cytokines, growth factors, and extracellular matrix (ECM) components [16]. Single-cell analyses have uncovered CAF heterogeneity, with distinct subsets exerting either pro-tumorigenic or tumor-restraining effects, though their precise origins and roles in CRC

progression remain elusive [17]. Although the tumor-suppressive effects of butanoate metabolism in epithelial cells have been extensively investigated, its potential role in regulating tumor epithelial cell-fibroblast interactions remains unclear. Our single-cell and spatial analyses identified enhanced spatial association and putative intercellular communication between COL1A2-positive fibroblasts and CD44-positive tumor epithelial cells with low butanoate metabolic activity. These observations suggest that alterations in epithelial butyrate metabolism may be linked to fibroblast activation and ECM remodeling, thereby contributing to the formation of a tumor-supportive stromal microenvironment.

In addition, HOXB8 transcriptional activity is significantly elevated in tumor epithelial cells with low butanoate metabolism, suggesting that HOXB8 may link altered butyrate metabolism to malignant tumor-cell behavior. HOXB8, a homeobox transcription factor, has been implicated in several malignancies, including colorectal, breast, bladder, and thyroid cancers [18–22]. In CRC, HOXB8 is frequently overexpressed and has been reported to promote tumor-cell proliferation, migration, and invasion through activation of Wnt/β-catenin signaling and upregulation of oncogenic mediators, including MMP2, c-Myc, cyclin D1, and vimentin [23]. Previous microarray analyses have shown that butyrate treatment alters the expression of multiple Wnt-associated genes, including HOXB8, indicating a potential regulatory relationship between butyrate exposure and HOXB8 expression [24]. However, whether butyrate directly regulates HOXB8 and whether HOXB8 mediates the tumor-suppressive effects of butyrate in CRC remain unknown.

Collectively, our findings highlight butanoate metabolism as a crucial determinant of CRC progression by orchestrating tumor-intrinsic oncogenic pathways, modulating transcription factor activity, and shaping the dynamic crosstalk between tumor epithelial cells and cancer-associated fibroblasts. The identified butyrate/HOXB8/AKT/mTOR axis represents a novel regulatory mechanism, reinforcing the potential of metabolic interventions to suppress tumor growth. Moreover, the interplay between low butanoate metabolism, fibroblast activation, and HOXB8 upregulation underscores the complexity of CRC pathogenesis, suggesting that targeting these metabolic and signaling networks could offer a promising therapeutic strategy to improve patient outcomes.

2 Methods

2.1 Data acquisition

In this study, we integrated three publicly available datasets to comprehensively analyze single-cell RNA sequencing (scRNA-seq) data in colorectal cancer. The single-cell RNA sequencing datasets included GSE144735, GSE166555, and GSE132465 [17, 25], which were obtained from the Gene Expression Omnibus (GEO) database. Additionally, bulk RNA sequencing data and corresponding clinical and survival information were retrieved from The Cancer Genome Atlas (TCGA) COAD and READ cohort (https://xena.ucsc.edu) [26]. These datasets provide a robust foundation for investigating the molecular landscape of colorectal cancer at both single-cell and population levels.

2.2 Gene Set Score Calculation

Genes associated with butanoate metabolism were retrieved from the 113 metabolism associated signatures, as described in Chen Yang's [27] study. Additionally, two key transporters involved in butyrate uptake, SMCT-1 (SLC5A8) and MCT-1 (SLC16A1), were included in the analysis due to their known roles in mediating cellular butyrate transport [28] (Additional file 1, Supplementary Table S2).

To quantify butanoate metabolism gene set activity, we employed the irGSEA (https://github.com/chuiqin/irGSEA), leveraging eight distinct scoring methods specifically designed for single-cell transcriptomic data. The specific parameters used were: min.cells = 3, min.feature = 0, custom = F, msigdb = T, species = "Homo sapiens", and kcdf = 'Gaussian'.

2.3 ScRNA-seq Analysis

To ensure data integrity and consistency, the single-cell RNA sequencing data underwent preprocessing using the Seurat R package (v4.0.5) [29]. Quality control procedures involved filtering out genes detected in fewer than three cells, discarding cells with fewer than 50 expressed genes, and eliminating cells in which mitochondrial gene content exceeded 5%. Given that prior studies had already performed initial quality assessments, no further filtering was applied.

To standardize the expression data, SCTransform was utilized [30], followed by batch effect correction across the three datasets using the Harmony algorithm (v0.1.0) [31]. Afterward, Seurat objects were merged into a comprehensive dataset for downstream analysis. Dimensionality reduction was performed using Principal Component Analysis (PCA), and cell clustering was executed through the FindNeighbors and FindClusters functions. To account for cell cycle heterogeneity, Cell-cycle scores were determined using Seurat's CellCycle Scoring function, allowing for phase-specific adjustments. Stemness scores were calculated using the AddModuleScore function, with gene sets referenced from Hang Zheng's study [32]. Finally, data visualization and interpretation were carried out using Uniform Manifold Approximation and Projection (UMAP) to reveal cellular heterogeneity and structure.

2.4 Cell Type Recognition

A differential expression analysis was executed across all genes within identified cell clusters, employing Seurat's FindAllMarkers function to delineate cluster-specific marker genes. The selection criteria encompassed an adjusted P-value threshold of < 0.05, an expression fraction surpassing 25%, and an absolute log2 fold change (FC) exceeding 0.25 (Additional file 1, Supplementary Table S3). Following this, distinct cellular clusters were annotated utilizing the GPTCelltype R package [33], which capitalizes on marker gene signatures for precise classification. To ensure the robustness and fidelity of these classifications, manual curation was performed, with refinements corroborated through comparative assessment against the CellMarker database.

2.5 Intercellular Communication Analysis

To elucidate intercellular communication dynamics within the tumor microenvironment (TME), we employed CellChat, a computational framework designed to infer and analyze cell-cell communication networks based on scRNA-seq data. CellChat [34] systematically identifies and quantifies ligand-

receptor interactions, enabling a comprehensive understanding of signaling pathways mediating intercellular crosstalk. The analysis was conducted by assessing receptor expression in one cell type and the corresponding ligand expression in another, facilitating the identification of significant cell type-specific interactions. Only ligand-receptor pairs with a p-value < 0.05 were considered statistically significant. Furthermore, CellChat's network-based approach allowed us to characterize both the incoming communication patterns of target cells and the outgoing signaling patterns of secreting cells, thereby offering a refined perspective on intercellular signaling within the TME.

2.6 Single-Cell Trajectory Inference

To reconstruct cellular state transitions at the single-cell level, we employed the Monocle2 R package (v2.16.0) for trajectory analysis [35]. Pseudotime trajectories were established by filtering cells based on mean expression > 0.1 and selecting genes with dispersion_empirical ≥ 1 * dispersion_fit, as discerned by Monocle2, ensuring robust modeling of dynamic processes. Dimensionality reduction was performed using the reduceDimension function with DDRTree as the reduction method and max_components set to two, allowing visualization via plot_cell_trajectory. Genes exhibiting significant variation along the pseudotime trajectory were identified using the differentialGeneTest function and subsequently visualized through plot_pseudotime_heatmap. These genes were further categorized into distinct expression-based subgroups, facilitating the characterization of transcriptional dynamics across the inferred trajectory.

2.7 InferCNV Analysis

We isolated epithelial cells and constructed a new gene-cell matrix. Single-cell copy number variation (CNV) analysis was conducted using the inferCNV R package (v1.6.0; https://github.com/broadinstitute/inferCNV) to evaluate chromosomal alterations in epithelial cells [36]. A new gene-cell matrix was constructed, incorporating epithelial cells while designating B cells as reference populations. The analysis was executed with default parameters, including "denoise," the hidden Markov model (HMM) configurations, and a threshold value of 0.1 for "cutoff" to enhance CNV inference. To mitigate false-positive CNV calls, the Bayesian latent mixture model was applied with a posterior probability threshold of 0.5.

2.8 Stemness Potential Assessment via CytoTRACE

To delineate cellular differentiation hierarchies and infer stemness potential, we employed the CytoTRACE algorithm, a computational framework developed by Gulati et al [37]. CytoTRACE systematically quantifies gene expression dynamics by leveraging correlations between transcript abundance and total gene counts in scRNA-seq data, offering a refined measure of cellular plasticity.

The analysis was conducted using the CytoTRACE R package (v0.3.3), which assigns a numerical score ranging from 0 to 1 to each cell, where higher scores denote greater stemness and a more undifferentiated state, while lower scores indicate progressive differentiation. This approach has been extensively validated across diverse large-scale datasets, consistently outperforming conventional

computational methods for stemness evaluation. By applying CytoTRACE to malignant cells, we inferred differentiation trajectories, unveiling heterogeneity in tumor cell plasticity and its potential implications within the tumor microenvironment.

2.9 Metabolic Landscape Reconstruction via scFEA

Metabolic fluxes and metabolite abundances in malignant epithelial cells were computed using scFEA (v1.1.2) in Python [38]. This framework integrates flux balance analysis with probabilistic modeling, leveraging a curated metabolic network and graph neural network–based optimization. The analysis incorporated 70 metabolic modules and their associated metabolites, which were visualized via heatmaps to facilitate metabolic characterization.

2.10 Spatial Transcriptomics Data Processing and Integration

Spatial transcriptomics (ST) data from GSE225857 were processed and visualized using the Seurat R package [39] Data normalization was conducted via the SCTransform (SCT) method, and integration was performed using SelectIntegrationFeatures, PrepSCTIntegration, FindIntegrationAnchors, and IntegrateData functions to ensure a harmonized dataset. Unsupervised clustering was employed to categorize spatially distinct tissue regions based on transcriptomic similarity. Cell population annotations were refined by integrating hematoxylin and eosin (HE) staining images with cluster-specific marker genes, further corroborated by the GPTCelltype R package. To delineate spatial expression patterns, the SpatialDimPlot and SpatialFeaturePlot functions were utilized, providing an in-depth visualization of transcriptomic distributions within the tissue architecture.

2.11 Spatial Cell Type Decomposition via Cell2location

To deconvolute spatial transcriptomics data into absolute cell-type abundances, we employed Cell2location, a Bayesian framework designed to map cell types from reference scRNA-seq data onto spatial transcriptomic profiles with high resolution [40]. The expected number of cells per spot was estimated via histological inspection, and the regularization parameter was adjusted to mitigate batch-specific variability. Key parameters were set following the official documentation. Gene filtering excluded those detected in fewer than 5 cells, present in less than 3% of cells, or with an average expression below 1.12. The model was trained for up to 30,000 epochs to ensure convergence. A negative binomial distribution was applied to infer cell-type proportions, incorporating spot- and gene-specific rate adjustments to correct for contamination and technical artifacts. This method provided a high-resolution spatial reconstruction of cellular composition, offering deeper insights into tissue organization.

2.12 Functional Pathway Estimation from Spatial Data

Signaling pathway activity across spatial transcriptomic regions was inferred using PROGENy (v1.12.0) [41, 42]. A model matrix derived from the top 1,000 genes of each transcriptional footprint was applied to sctransform-normalized data. Pathway scores were computed for individual spatial locations, enabling a spatially resolved assessment of functional heterogeneity.

2.13 Spatial Coordination of Cell Types and Functional Pathways

To elucidate spatial dependencies among cell types, we employed MISTy in mistyR (v1.2.1), leveraging a multi-view modeling approach [43]. Cell-type proportions inferred by Cell2location were integrated into the spatial analysis framework. This analysis incorporated three spatial perspectives: (1) an intrinsic perspective examining within-spot cellular associations, (2) a juxta perspective capturing proximal interactions within a 10-spot radius, and (3) a para perspective integrating long-range influences extending up to 15 spots. The standardized importances derived from these spatial contexts were analyzed to infer cell-type interrelations, identifying patterns of spatial colocalization and segregation. Predictors with an R^2 value below 10% were excluded to ensure robust estimations.

To establish links between tissue organization and functional activity, we applied MISTy to model the spatial distribution of PROGENy pathway activity scores. This framework incorporated multiple perspectives, including localized interactions, neighboring influences (max distance = 10), and extended tissue-wide correlations (radius = 15). Additionally, two perspectives were included to capture interactions between cell-type compositions and pathway activities. Pathway predictors with an R^2 value below 10% were discarded before aggregation. This approach provided a refined spatial characterization of cellular communication and functional dynamics within the colorectal cancer microenvironment.

2.14 Spatial trajectory analysis and Cellular Communication

To reconstruct spatially informed cell-state transitions and infer cell-cell interactions, we utilized stLearn (v1.0) [44], an integrated framework for spatial transcriptomic analysis. Cell-cell communication was estimated by leveraging spatial gene expression dependencies, followed by spatially constrained pseudotime trajectory inference using PAGA, built upon SME-normalized gene expression data.

2.15 Transcription Factor Regulatory Network Analysis

Gene regulatory network inference was performed using pySCENIC on both single-cell and spatial transcriptomic data [45]. AUCell was employed to evaluate regulon activity across individual cells. For spatial transcriptomics, transcription factor activity was compared between tumor epithelial and normal epithelial cells, as well as between tumor epithelial cells with high and low butanoate metabolism scores. Specific transcription factors and their target genes were further predicted, enhancing the resolution of regulatory interactions.

Additionally, Connection Specificity Index (CSI) analysis [46, 47], a context-dependent approach for identifying specific associating partners, was conducted to identify regulon modules. Pearson correlation coefficients of regulon activity scores were computed, followed by CSI-based hierarchical clustering with Euclidean distance to delineate regulon modules. A CSI threshold of 0.7 was applied to construct a regulon association network, visualized using Cytoscape. For each regulon module, its

activity score was defined as the average activity of its regulon members across all cells within a given cell type. Pearson correlation coefficients were then calculated to assess the association between each module and cell types, identifying the top-correlated cell types as the primary regulatory context for each module. This framework enabled the identification of key regulatory patterns and their associations with distinct cell types across spatial and single-cell contexts.

2.16 Machine Learning-Derived Signature Construction

To establish a robust and high-performance prognostic signature, we integrated 10 machine learning algorithms and 101 algorithmic combinations. The ensemble models encompassed random survival forest (RSF), elastic network (Enet), Lasso, Ridge, stepwise Cox, CoxBoost, partial least squares regression for Cox (plsRcox), supervised principal components (SuperPC), generalized boosted regression modeling (GBM), and survival support vector machine (survival-SVM).

The signature development process proceeded as follows: (a) Marker genes associated with tumor epithelial cells exhibiting low butanoate scores were first identified using Seurat's FindAllMarkers function; (b) These marker genes were then subjected to Univariate Cox regression to determine prognostic relevance; (c) Feature selection was further refined using the Boruta algorithm to eliminate non-informative genes; (d) Finally, 101 algorithmic combinations were evaluated to construct prognostic models using 10-fold cross-validation in the TCGA cohort. The performance of these models was further assessed in the GSE17536 and GSE39582 cohorts using Harrell's concordance index (C-index). The model with the highest average C-index across the two external evaluation cohorts was selected as the final model.

2.17 Immunotherapy Response and Precision Drug Prediction

To comprehensively assess immunotherapeutic efficacy and explore potential precision-targeted therapies, we leveraged datasets GSE91061 and PRJEB23709 from the TIGER database. These datasets facilitated the identification of immune response signatures and their association with therapeutic outcomes.

Additionally, drug sensitivity profiles for cancer cell lines (CCLs) were extracted from the Cancer Therapeutics Response Portal (CTRP v.2.0, released October 2015, https://portals.broadinstitute.org/ctrp) and the PRISM Repurposing dataset (19Q4, released December 2019, https://depmap.org/portal/prism/). Drug response was estimated by calculating the area under the dose–response curve (AUC) for each compound using ridge regression via the calcPhenotype function in the R package "pRRophetic." Lower AUC values indicated heightened sensitivity to treatment. The differences in AUC values between risk-stratified groups and their correlation with risk scores were further assessed.

2.18 Survival analysis

To investigate the prognostic significance of butanoate metabolism and the constructed risk model, we performed Kaplan–Meier survival analysis. The median was chosen as the threshold to stratify patients into high- and low-risk categories. The “Survival” and “survminer” packages in R were employed for survival differences evaluation, while the “ggplot2” package was used for visualization. Additionally, receiver operating characteristic (ROC) and time-dependent ROC analyses were conducted to evaluate the predictive accuracy of the risk model. The “pROC” and “timeROC” packages in R were utilized to quantify sensitivity and specificity, providing a robust assessment of prognostic performance.

2.19 Immune Landscape Analysis

We first examined the expression of immune checkpoint molecules across tumor patients stratified by different risk scores. To further characterize immune activity, we employed CIBERSORT (R package) [48] to estimate the infiltration levels of 22 immune cell types in patients with varying risk scores based on gene expression profiles.

Additionally, the single-sample gene set enrichment analysis (ssGSEA) algorithm, an extension of the GSEA approach, was utilized to quantify the abundance of 29 immune signatures [49] across different risk groups, providing insights into immune heterogeneity in tumor patients with varying risk scores.

2.20 Reagents

Anti-HOXB8 antibody (1:1000 dilution; Cat. No. 31387-1-AP) was purchased from Proteintech (Wuhan, China), anti-Bax (1:1000 dilution; Cat. No. GB15690-100), anti-Bcl-2 (1:1000 dilution; Cat. No. GB154380-100), and anti-GAPDH (1:1000 dilution; Cat. No. GB15002-100) antibodies were obtained from Servicebio (Wuhan, China). *HOXB8* plasmids were acquired from Beijing Maijing Biotechnology Co., Ltd. (Beijing, China). Lipofectamine 3000 (Cat. No. L3000015) and Dulbecco’s modified Eagle’s medium (DMEM) were purchased from Invitrogen (Thermo Fisher Scientific, Inc., Waltham, MA, USA). Fetal bovine serum (FBS; Cat. No. C04001-500) was obtained from Vivacell (Shanghai, China). Cell Counting Kit-8 (CCK-8; Cat. No. C0038) and crystal violet staining solution (Cat. No. C0121) were procured from Beyotime Biotechnology (Shanghai, China). Opti-MEM (Cat. No. 11058021) was sourced from Gibco (Thermo Fisher Scientific, Inc., Waltham, MA, USA). Transwell chambers (24-well plate, 6.5 mm insert, 8 μm pore size, PC membrane) were obtained from LABSELECT (Labgic Bio, Beijing, China). Sodium butyrate (Cat. No. S817488) was purchased from MACKLIN (Shanghai, China). Matrigel (Cat. No. BL1833B) was acquired from Biosharp (Anhui, China). Culture-insert 2 well was procured from ibidi (Grafelfing, Germany).

2.21 Cell Culture

HCT-116 human colorectal cancer cells were obtained from Shanghai Fuheng Biotechnology Co., Ltd. (Shanghai, China; Cat. No. FH0027) in 2025. The cell line was authenticated by short tandem repeat (STR) profiling by the supplier. Cells were cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and maintained at 37°C in a humidified atmosphere containing 5% CO_2.

2.22 Plasmid Transfection

Twenty-four hours prior to transfection, cells were digested with trypsin, resuspended, and seeded into 6-well plates at a density of approximately 1 × 10^6 cells per well, aiming for 60%-70% confluence at the time of transfection. For transfection mixture preparation: in Tube A, *HOXB8* plasmid DNA and P3000™ reagent were diluted in Opti-MEM™ Reduced Serum Medium according to the manufacturer's instructions. In a separate Tube B, Lipofectamine™ 3000 reagent was diluted in Opti-MEM™. The contents of Tube A and Tube B were then combined, mixed gently, and incubated at room temperature for 10–15 minutes to allow DNA-lipid complex formation. The resulting complexes were added dropwise to the cell culture wells along the sidewall, and the plate was gently swirled to ensure even distribution. Cells were returned to the 37°C, 5% CO_2 incubator for 4 hours, after which the medium was replaced with fresh complete culture medium. Transfected cells were harvested for subsequent experiments 72 hours post-transfection.

2.23 Sodium Butyrate Treatment

A stock solution of sodium butyrate was prepared at 160 mM in DMEM and filter-sterilized. For formal experiments, this stock solution was diluted in complete medium to obtain working concentrations of 2 mM, 4 mM, and 8 mM.

2.24 Cell Viability Assay

Cell viability was assessed using the Cell Counting Kit-8 (CCK-8) according to the manufacturer's protocol. Briefly, HCT116 cells were digested, resuspended, and seeded into 96-well plates at a density of 3,000–5,000 cells per well. Each experimental group consisted of four replicate wells, with blank control wells (medium only) included. After the designated treatment periods, 10 µL of CCK-8 solution was added to each well, followed by incubation for 2 hours at 37°C. The absorbance of each well was measured at 450 nm using a microplate reader (BioTek Instruments, Winooski, VT, USA). Cell viability curves were generated from three independent experiments. Cell viability was calculated as follows: Viability (%) = [(OD_Experimental - OD_Blank) / (OD_Control - OD_Blank)] × 100%. Data are presented with different time points on the X-axis and cell viability percentage on the Y-axis.

2.25 Wound Healing Assay

HCT116 cells were digested and resuspended in serum-free medium to a density of 3 × 10^5 cells/mL. A Culture-insert 2 well (ibidi) was placed securely in a 24-well plate. Each well of the insert was filled with 100 µL of the cell suspension, while the area outside the insert received 400 µL of the same suspension. After 12 hours of incubation at 37°C, the culture insert was carefully removed to create a defined cell-free gap (wound). The initial wound area was photographed using a digital camera (400× magnification). Following relevant drug treatments, cells were further incubated for 48 hours, after which the same areas were re-photographed. Wound closure was analyzed by comparing the changes in the wound area over time.

2.26 Cell Invasion Assay (Transwell)

Matrigel was thawed overnight at 4°C and diluted 1:8 with cold DMEM to prepare Solution A. Each Transwell chamber (8 μm pore size, 6.5 mm diameter, LABSELECT) was coated with 100 μL of Solution A and incubated at 37°C overnight. The following day, excess liquid was aspirated, and the chambers were hydrated with 200 μL DMEM for 3 hours. HCT116 cells were digested, resuspended in DMEM containing 2% FBS at a density of 5×10^5 cells/mL, and 100 μL of this cell suspension was added to the upper chamber. The lower chamber was filled with 500 μL of DMEM supplemented with 10% FBS as a chemoattractant. After 48 hours of incubation at 37°C, non-migrated cells on the upper surface of the membrane were carefully removed using a cotton swab. Cells that had migrated to the lower surface were fixed with 4% paraformaldehyde for 15 minutes and stained with 0.1% crystal violet for 5 minutes. After gentle rinsing with PBS, the membranes were examined under a microscope. The number of migrated cells was counted in five randomly selected fields per membrane.

2.27 Western Blot Analysis

The culture medium in the 6-well plates was discarded, and the cells were washed twice with ice-cold PBS. An appropriate volume of RIPA lysis buffer containing protease and phosphatase inhibitors was added to each well, and cells were lysed on ice for 30 minutes. The lysates were scraped, collected into microcentrifuge tubes, and centrifuged at 12,000 rpm for 15 minutes at 4°C. The supernatant (total protein extract) was collected. Protein concentration was determined using a BCA protein assay kit, and all samples were adjusted to the same concentration. Protein samples were mixed with 5× loading buffer, denatured by boiling at 100°C for 10 minutes, and equal amounts of protein were loaded onto SDS-PAGE gels for electrophoresis under constant voltage. Proteins were subsequently transferred onto PVDF membranes using a wet transfer system, typically at a constant current of 200–300 mA for 1.5 hours. The membranes were blocked with 5% non-fat milk in TBST for 1 hour at room temperature with gentle shaking. After blocking, membranes were incubated with primary antibodies diluted in TBST overnight at 4°C with shaking. Following three washes with TBST (10 minutes each), membranes were incubated with HRP-conjugated secondary antibodies for 1 hour at room temperature. After another three washes with TBST, protein bands were visualized using an ECL chemiluminescent substrate and detected with a chemiluminescence imaging system. The expression levels of target proteins were quantified using ImageJ software (Version 1.43u, National Institutes of Health, Bethesda, MD, USA) and normalized to GAPDH.

3 Results

3.1 Single-cell Transcriptomic Landscape of Colorectal Cancer

The overall workflow of our study is depicted in Fig. 1. We incorporated three publicly available single-cell RNA sequencing (scRNA-seq) datasets (GSE144735, GSE166555, and GSE132465) to bolster the robustness and comprehensiveness of our findings. These datasets comprised tumor and adjacent normal tissue samples collected from a total of 41 colorectal cancer patients following surgical

resection. Supplementary Table S1 provides a detailed summary of the clinical and pathological characteristics of all patients included in this study. To integrate the scRNA-seq data from different sources, we applied the Harmony algorithm to effectively eliminate batch effects while simultaneously accounting for multiple experimental and biological covariates. After dimensionality reduction and clustering, cells were annotated according to canonical marker expression, yielding nine distinct cell types: T cells (CD3E, CD3G), B cells (MS4A1), plasma cells (SDC1, MZB1), myeloid cells (CD14, FCGR3A), fibroblasts (COL1A1), mast cells (KIT, MS4A2), endothelial cells (PECAM1), glial cells (S100B), and epithelial cells (EPCAM) (Fig. 2A-B). The distinctive cellular markers within each major cell population were visualized using a violin plot (Additional file 2, Supplementary Fig. S1A, Additional file 1, Supplementary Table S3). The functional enrichment profiles of the top 50 marker genes in each cell type shed light on the transcriptional features underpinning their unique physiological roles (Additional file 2, Supplementary Fig. S1B). As representative examples, fibroblast-specific genes were primarily enriched in extracellular matrix-related pathways, including collagen fibril organization, whereas B cell markers were associated with immunological functions such as activation, proliferation, and B cell receptor signaling (Additional file 2, Supplementary Fig. S1B).

3.2 Construction and Calculation of a Butyrate Metabolism Signature

As a major microbial metabolite in the colon, butyrate is increasingly recognized as a key regulator of the colorectal cancer microenvironment, offering anti-inflammatory, immunomodulatory, and potential therapeutic benefits [50]. To quantify butyrate metabolism at the single-cell level, we employed irGSEA, an R package that integrates distinct scoring algorithms based on gene expression ranks, emphasizing relative expression levels rather than absolute values. All eight scoring algorithms consistently revealed that epithelial cells exhibited the highest butyrate metabolism scores among all identified cell types, followed by fibroblast cells (Fig. 3A). Previous studies have shown that in normal colonic epithelial cells, butyrate facilitates cell growth by serving as an oxidative energy source. However, in cancerous colonocytes, butyrate is inefficiently metabolized, as glucose replaces butyrate as the primary energy source (the Warburg effect) [51–53]. In line with these findings, butyrate metabolism scores were markedly reduced in tumor tissues compared to adjacent normal tissues, with this difference being particularly pronounced between malignant and non-malignant epithelial cells (Fig. 3C). We next calculated butyrate metabolism scores for each patient in TCGA cohort. Based on the median score, patients were stratified into highBTA and lowBTA groups. Survival analysis revealed that patients in the lowBTA group exhibited significantly poorer overall survival (OS) and disease-free survival (DFS) compared to those in the highBTA group (Fig. 3B), underscoring the association between high butyrate metabolism and favorable prognosis.

3.3 Differentiation States and Phenotypic Profiling of Tumor Epithelial Cells with High and Low Butyrate

Metabolism

UMAP analysis of epithelial cells from core tumor regions uncovered five distinct subpopulations (Additional file 2, Supplementary Fig. S2A). To further identify malignant cells, we extracted these five epithelial subgroups along with B cells, and applied the inferCNV algorithm to assess large-scale chromosomal alterations. The results divulged that all five epithelial subgroups exhibited remarkably higher CNV levels than compared to normal B cells, suggesting their high degree of malignancy (Additional file 2, Supplementary Fig. S2B-C). Across lesions, inferCNV analysis revealed mostly noticed gains in 20p, 20q, 7p, 7q, 8p,8q, and 19q, consistent with well-defined CNVs reported in previous WGS-based colorectal cancer studies (Additional file 2, Supplementary Fig. S3A) [54].

The dysregulation of butyrate metabolism in malignant epithelial cells warrants further attention due to its potential relevance to tumor progression and cellular differentiation. To explore this, we stratified epithelial cells into two subgroups based on their butyrate metabolism scores: HighBTAepis and LowBTAepis. To investigate the evolutionary dynamics of tumor epithelial cells within the BTA group, we conducted unsupervised cell trajectory analysis using Monocle2 (Fig. 4A). The trajectory analysis elucidated that malignant epithelial cells could be deconvoluted into nine distinct differentiation states, with State 1 corresponding to the initiation of the trajectory and State 7 representing the terminal stage (Fig. 4B). Strikingly, LowBTAepis were predominantly aggregated in the early and middle stages of pseudotime trajectory, whereas HighBTAepis were primarily distributed at the terminal end of the trajectory (Fig. 4C). Expanding on this, we visualized the differentiation trajectories of BTA genes (butyrate metabolism-related genes) in nine states and their dynamic expression changes during differentiation. Heatmap showcased that key regulators of butyrate uptake and oxidation, including SLC5A8 (SMCT1), SLC16A1 (MCT1), HMGCS1/2, and BDH1/2, were upregulated in the intermediate-to-terminal stages. In contrast, ACAT2, OXCT2, and L2HGDH were mainly prominent during the incipient phase (Fig. 4D). Trajectory analysis dissected a notable expansion of HighBTAepis within terminal pseudotime 7–9 states, while LowBTAepis were predominantly confined to the early phases, specifically 1–2 states (Fig. 4E). Moreover, cell-cycle interrogation demonstrated that LowBTAepis exhibited a significantly higher fraction of S and G2/M phase cells, pointing to increased proliferative potential (Fig. 4F). Stemness inference via CytoTRACE highlighted a pronounced increase of stemness scores in LowBTAepis (Fig. 4G-H), concurrently with increased CNV levels (Fig. 4I), collectively reflecting a more aggressive and undifferentiated phenotype. Such results are concordant with trajectory inference, which positioned LowBTAepis as residing at the differentiation origin — a position commonly attributed to cells with high stemness. Through inferCNV analysis, LowBTAepis exhibited focal upregulation in CNV levels across 19q gain, 19p gain, and 7p gain, while showing reduced CNV levels including 19q loss, 16q gain, 9q gain, 3p gain and 3q loss (Fig. 4J).

3.4 Intercellular Communication Landscape of BTA-Stratified Epithelial Cells and the Tumor Microenvironment

Recent studies have highlighted the pivotal role of butyrate in modulating the tumor immune microenvironment [55]. Specifically, butyrate has been shown to promote T cell differentiation into both effector and regulatory (Treg) subsets, thereby shaping immune responses within the tumor [56]. In order to uncover how stratification by butyrate metabolism correlates with distinct intercellular signaling programs in the tumor microenvironment, we leveraged CellChat to delineate the communication networks between HighBTAepis/LowBTAepis cells and other cellular populations, Remarkably, the GDF signaling pathway conveyed pronounced specificity within HighBTAepis/LowBTAepis. In addition, with respect to LowBTAepis, the predominant incoming signals included *PTN* and MK pathways, whereas the main output signals comprised MIF, MK, VEGF, GDF, and PARs (Fig. 5A-B). Upon quantifying interaction strength, the majority of intercellular communications occurred among T cells, B cells, plasma cells, myeloid cells, fibroblasts, mast cells, endothelial cells, glial cells, and HighBTAepis/LowBTAepis (Fig. 5C). Besides, the count of ligand–receptor interactions between HighBTAepis and other cell types surpassed that observed in LowBTAepis (Fig. 5D). The likelihood of ligand–receptor interactions between HighBTAepis/LowBTAepis and other cell type were further evaluated. Results indicated that myeloid cells preferentially interacted with LowBTAepis through *SPP1–CD44* and *LGALS9–CD44*. In particular, the *CD44–SPP1* axis has been previously implicated in suppressing $CD8^+$ T cell activation, thereby facilitating tumor immune tolerance and promoting immune evasion [57]. Beyond that, several key ligand–receptor pairs originating from fibroblasts were pinpointed as specifically targeting LowBTAepis, which encompass MDK–*NCL*, MDK–(*ITGA6* + *ITGB1*), MDK–*SDC4*, PTN–*NCL*, and PTN–*SDC4*. Midkine (MDK) and pleiotrophin (PTN) are closely related heparin-binding growth factors implicated in both neural development and tumorigenesis, reflecting their dual roles in physiological and pathological contexts [58, 59]. Such functionally significant interactions are capable of activating pivotal oncogenic signaling cascades—including PI3K–AKT, MAPK or JAK–STAT, thereby fostering tumor cell proliferation, migration, angiogenesis, and facilitating an immunosuppressive microenvironment [60, 61].

3.5 Transcription Factors Profiling and Metabolic flux analysis in Butyrate Metabolism–Stratified Tumor Epithelial Cells

To dissect the regulatory mechanisms underlying HighBTAepis/LowBTAepis subgroups, we implemented the pySCENIC pipeline, which involves three key steps: inference of TF–target co-expression modules, motif-based refinement to delineate high-confidence regulons, and AUCell-based scoring to evaluate regulon activity across single cells [62]. Transcription factors uniquely indicative of each subtype were ranked based on their respective regulon specificity scores (RSS) (Fig. 6A). Strikingly, the LowBTAepis were characterized by the selective upregulation of *STAT1* in JAK-STAT signaling [63], *TCF4* in Wnt signaling, *TP53* in p53 signaling, *TFAP2A* in PI3K-AKT signaling [64], *KLF6/13* in TGF-β signaling [65, 66] and *NFE2L2* (NRF2), which governs antioxidant and detoxification gene programs [67]. For HighBTAepis, critical transcription factors with elevated regulon activity included *CDX1/2*, E2F2, *CEBPA*, *HNF4G*, *NR2F6*. Emerging evidence suggests that butyrate, a major microbial derived short chain fatty acid, reprograms colorectal cancer cell metabolism by suppressing glycolysis [14], reducing SCAD-

mediated beta-oxidation [68], and enhancing TCA cycle activity [69]—thereby coupling metabolic and epigenetic remodeling to arrest proliferation and induce apoptosis. To further investigate the potential role of butyrate in the metabolic regulation of tumor epithelial cells, a deep learning-based scRNA-seq analysis tool, scFEA was employed to computationally infer cellular metabolic fluxes in HighBTAepis/LowBTAepis subgroups. The observed enhancement of metabolic fluxes through key TCA cycle reactions (e.g., Citrate→2OG, Pyruvate→Acetyl-CoA, Succinate→Fumarate) suggests a metabolically oxidative state in LowBTAepis, potentially reflecting a compensatory metabolic adaptation to altered butyrate utilization. Additionally, stimulated degradation of lysine, valine, and isoleucine into TCA cycle intermediates (lysine→Acetyl − CoA, Valine→Succinyl − CoA, Isoleucine→Acetyl − CoA, Isoleucine→Succinyl − CoA) in LowBTAepis likely represents a compensatory mechanism to replenish mitochondrial carbon pools and sustain oxidative metabolism. Moreover, increased flux through the N-linked glycan synthesis pathway was observed in LowBTAepis, manifested by elevated dolichol-linked oligosaccharide assembly and complex-type glycan maturation, indicating augmented glycoprotein processing capacity (Fig. 6C). Other metabolic pathways enriched in LowBTAepis—such as Pyruvate → Acetyl-CoA (M_5), Acetyl-CoA → (E,E)-Farnesyl-PP (M_113), Acetyl-CoA → Fatty Acid (M_34), and Valine → Succinyl-CoA (M_54)—further support a compensatory shift toward mitochondrial oxidative metabolism and reduced reliance on aerobic glycolysis (Fig. 6B). These alterations are indicative of a suppressed Warburg effect and substantiate that LowBTAepis flexibly reprogram carbon metabolism to maintain energy homeostasis through amplified TCA cycle activity.

3.6 Decoding Butyrate Metabolism through the Convergence of Spatial Transcriptomics and Single-Cell RNA Sequencing

Spatial transcriptomics offers a unique advantage by enabling the precise localization of gene expression, which allows for better visualization of the heterogeneity of butyrate metabolism. In this work, we re-analyzed previously published spatial transcriptome datasets from three colorectal cancer (CRC) patients (ST-C1, ST-C2 and ST-C3). Leveraging hematoxylin and eosin (HE) stained tissue sections, coupled with the intelligent annotation of marker genes for each cluster using GPTCelltype, we annotated and cataloged the spatial locations into six predominant clusters: tumor cells, fibroblasts, smooth muscle cells, epithelial cells, endothelial cells, and immune cells (Fig. 7A-B). A pronounced decrease in butyrate metabolism scores was discerned within the neoplastic region (Fig. 7C). We scrutinized and interrogated the ST data with the scRNA-seq datasets through the application of Cell2location, and eleven distinct cell types identified in scRNA-seq were spatially mapped to the CRC tissue slice (Fig. 7D). The mapping results obtained from Cell2location exhibit a high degree of consistency with the prior annotation outcomes. To explore the spatial dynamics of cellular interactions and abundance in CRC, we employed MISTy [43], a cutting-edge fusion multiview framework designed for establishing the associations between the detected expression profiles of the markers, for instance, intracellular modulation or paracrine signaling. Throughout tissue sections, spatially coincident interactions within specific tissue spots (intra) manifested as the most informative indicator, unveiling

co-occurrences between LowBTAepis and fibroblasts in ST-C1 and ST-C3, which were also substantiated by the cellchat results derived from scRNA data. A further observation demonstrated that LowBTAepis and B cells co-interactions occurred in intra-position or at a distal position in neighboring (10 pots) and extended neighboring regions (15 pots) in ST-C2 and ST-C3. Thereupon, we quantified pathway activities by exploiting PROGENy for each individual tissue spot based on spatial gene expression profiling. Our findings authenticate a significant upregulation of critical pathways in the LowBTAepis, including Androgen, EGFR, Hypoxia, JAK–STAT, MAPK, NFkB, p53, TNFa and TGFβ (Additional file 2, Supplementary Fig. S4A-B).

To optimally harness spatial information, MISTy was once again applied to the colocalization analysis of pathways, through which we uncovered intimate interactions between NFkB and TNFa signaling, TNFa and TGFβ signaling, EGFR and MAPK signaling, which have been extensively validated in colorectal cancer (Additional file 2, Supplementary Fig. S5A) [70]. These pathways displayed a hierarchical cascade of interactions, where NFkB is activated by TNFa, subsequently initiating TGFβ signaling to regulate cell differentiation and immune responses [71]. Furthermore, EGFR induces the MAPK pathway, driving cell proliferation and migration [72]. The spatial interrelations of these HighBTAepis or LowBTAepis-related pathways highlights the cellular heterogeneity within colon cancer cells. As it depicted, Androgen, WNT, TNFa, TGFβ, *TP53*, NFkB, JAK–STAT, Hypoxia and EGFR were detected in concomitant presence with LowBTAepis in ST-C1, ST-C2 and ST-C3 (Additional file 2, Supplementary Fig. S5B).

3.7 Combinatorial Regulation Landscapes elucidated through connection specificity index (CSI) in Spatial Transcriptomic Data

Transcription factors (TFs) orchestrate the transcriptional regulation of target genes by specifically binding to distinct DNA sequences. A comparative analysis of the atlas-wide similarity in RAS scores across regulon pairs was conducted, employing the Connection Specificity Index (CSI) [46, 47] to clarify the combinatorial patterns. Remarkably, these 389 regulons were hierarchically partitioned into 7 distinct modules (Fig. 8A, Additional file 1, Supplementary Table S4). Several indissociable regulators and cell types, based on their mean activity scores, were discerned within each module. The activity of M2, encompassing *ATF3*, *ATF6*, *BCL6*, *FOXC1* and *ZEB1*, is particularly elevated in fibroblasts. Previous research has confirmed that *FOXC1* and *ZEB1* govern fibroblast differentiation, primarily through involvement in the signal transduction cascades of key pathways such as TGF-β and Wnt, which modulate cellular migration and ECM remodeling [73–75]. Module M3 harbors *E2F5* and *KLF5*, which are pivotal in regulating cell cycle progression [76, 77], alongside *POU5F1B* and *NEUROD1* [78, 79], which are indispensable for sustaining stem cell pluripotency and differentiation. Additionally, members of the homeobox superfamily, including *HOXA4*, *HOXB6*, *HOXB8*, and *HOXB9*, influence tumor initiation and growth, invasion and metastasis, angiogenesis and anti-cancer drug resistance [80]. When projecting the regulon activity of module 3 onto histological slides, we noted a spatially coherent distribution pattern between LowBTAepis and those with high M3 scores, confirming their substantial associations (Fig. 8B). Capitalizing on the precisely computed CSI, we further visualized the transcription factor regulatory

networks pertinent to each module (Fig. 8C). To fully incorporate spatially resolved transcriptomic information, regulon-specific score (RSS) analysis was realized through the pySCENIC algorithm. Intriguingly, we noticed that *HOXB8*, *NFE2L2*, and *NELFE* were predominantly expressed in neoplastic epithelial cells, particularly in the LowBTAepis (Additional file 2, Supplementary Fig. S6A-C). Ultimately, we also depicted the regulatory networks of these three core TFs and their potential target genes (Additional file 2, Supplementary Fig. S6D-F). The target genes of *NFE2L2*, including *SDC4* and *SDCBP*, are involved in regulating extracellular matrix remodeling and cell adhesion [81, 82]. *HOXB8* modulates the expression of genes such as *ITGA9* and *MSLN*, which are integral to cellular migration and invasion [83, 84]. Additionally, *NELFE* regulates genes like *TMEM200A* and *TMEM205*, which are crucial for membrane dynamics and material exchange [85, 86].

3.8 Spatial Trajectory Inference and Cell–Cell Interaction Analysis in Tumor Core and Periphery Regions

Focusing on the dynamic, branching nature of cancerous cells or clones within the tumor microenvironment is critical, as it mirrors the complex processes observed in biological phenomena such as cell division, morphogenesis, clonal drug resistance, and evolutionary differentiation. Pseudotime, an advanced concept in scRNA-seq analysis, is a computational tool widely adopted to address the challenge of quantifying transcriptional changes as cells undergo biological processes and transition between different states [35]. In the course of our investigation, we zoomed into a tracking algorithm called stLearn to infer changes in cell states or progression within the tissue context of the profiled cells. The asynchrony process captured by stlearn in spatial transcriptomics exhibited an evolutionary route from LowBTAepis to HighBTAepis (Fig. 9A-C, left panel). The complex spatial trajectories of cell clusters were portrayed through an overarching hierarchical tree diagram (Fig. 9A-C, right panel). Of interest, the cluster located proximal to the tumor's core invasive and differentiate radially outward the deeper regions of periphery tissue. As an illustration, cluster 12 displayed a differentiation trajectory towards cluster 0 and cluster 46 in ST-C1, while ST-C2 showed a transition from cluster 3 to clusters 1, as well as from cluster 2 to clusters 18. Moreover, ST-C3 emphasized a progression from cluster 10 to clusters 9/4, and from cluster 5 to clusters 2/11/12/38/40. In brief, with the passage of time, branches characterized by intensified invasiveness often transform into subpopulations with disseminative invasive propensity.

Cellular state progression kinetics were typically reflected by transcriptomic data, in which stage-specific gene expression signatures demarcate a seamless continuum from nascent to fully advanced phases. We here undertook transition marker analysis to ascertain those genes that either increase or decrease their expression along the trajectory/trajectories of each clade (Fig. 10D).

Moreover, stLearn also enables the high-resolution delineation of cell–cell interactions between neighbouring spots or within spots, thereby encompassing juxtacrine, paracrine, and endocrine ligand–receptor mechanisms. Based on this, we detected a suite of candidate ligand–receptor interactions throughout the tissue section (Additional file 1, Supplementary Table S5), with *COL1A2–CD44* emerging

as the most salient pairing. We can also visualise where the ligand or receptor are expressed/coexpressed simultaneously with the dominant spots (Fig. 10A). Results indicated that *COL1A2–CD44* interactions were markedly prevalent within the tumor-associated mesenchyme. Upon annotating spots with specific cellular identities and executing cell–cell communication modeling, we corroborated an exceptionally potential crosstalk axis linking LowBTAepis to fibroblasts in all patients (Fig. 10B–C). To elucidate the possible biological process mediated by the postulated CCI events for the LR pairs, we ultimately implemented GO and KEGG enrichment analysis for the top 100 LR with the greatest number of significant spots. These LR pairs displayed significant enrichment of known biological processes regulated by ECM–cell adhesion axis (collagen-containing extracellular matrix, cell-matrix adhesion, regulation of cell migration and cell adhesion molecule binding), where collagen-rich matrix remodeling and integrin-mediated focal adhesions, and pathways including ECM-receptor interaction and PI3K-Akt signaling, thereby driving the choreographed progression of migration and invasive infiltration (Additional file 2, Supplementary Fig. S7A-B, Additional file 1, Supplementary Table S6).

3.9 Development and Validation of a Consensus BTA-risk score via Integrative Bulk RNA-Seq Analysis and Machine Learning Algorithms

We initiated our analytic pipeline with FindMarkers-based identification of LowBTAepis markers in scRNA-seq data. Univariate Cox analysis subsequently refined 47 genes as statistically robust predictors of clinical outcomes, warranting further integrative modeling (Additional file 1, Supplementary Table S7). Integrating feature selection into our pipeline, we harnessed the Boruta algorithm to extract 7 highly reliable genes (Fig. 11A, Additional file 1, Supplementary Table S8). Then we implemented a comprehensive model construction strategy leveraging 101 pairwise combinations of classical and ensemble machine learning algorithms. These models were constructed using 10-fold cross-validation in the TCGA cohort and subsequently evaluated in the GSE17536 and GSE39582 cohorts. Model performance was compared using the concordance index (C-index). Among the 101 algorithmic combinations, the CoxBoost + Random Survival Forest (RSF) model achieved the highest average C-index across the two external evaluation cohorts (0.715) and was therefore selected as the final BTA-risk model (Fig. 11B). A patient-specific prognostic score was formulated by applying the regression weights from the selected model to the corresponding gene expression profiles. This enabled robust bifurcation of the cohort into prognostically divergent subgroups. Kaplan–Meier survival analysis demonstrated a pronounced survival disadvantage in patients classified into the high-risk group compared to their low-risk counterparts, in both the discovery and validation datasets (Fig. 11C). Furthermore, Receiver operating characteristic (ROC) curve analysis demonstrated that the BTA-risk score possessed robust predictive performance across multiple time horizons, achieving AUCs of 0.993, 0.99, 0.985 at 1-, 3- and 5-year intervals in the TCGA cohort (Fig. 11D). The selected model also showed consistent prognostic stratification in GSE17536 and GSE39582, supporting its prognostic relevance across multiple cohorts.

3.10 Clinical Applications of the BTA-risk score in Shaping the Tumor Immune Microenvironment and Predicting Immunotherapy Outcomes

Given the well-established influence of the tumor immune microenvironment (TME) on cancer progression and therapeutic outcomes, we next systematically analyzed the immune landscape in relation to BTA-risk score. A notable upregulation of multiple immune checkpoint genes, including *CD274*, *PDCD1*, *CTLA4*, and *CD247* were observed in patients with elevated risk scores (Additional file 2, Supplementary Fig. S8A), suggesting a heightened state of immune suppression. Concomitantly, elevated TIDE scores (Additional file 2, Supplementary Fig. S8B) and decreased immunotherapy response rates (Fig. 12A-B) in patients with high BTA-risk scores underscore a more evasive immunophenotype, potentially contributing to their unfavorable clinical outcomes (Fig. 12C). Subsequent ssGSEA analysis revealed that patients in the high BTA-risk score group were characterized by pronounced enrichment of effector immune cell populations, particularly CD8 T cells, Tfh, Th1 cells and TILs. Additionally, immune-related functional pathways, such as check point, T-cell co-stimulation and co-inhibition, APC co-stimulation and co-inhibition were significantly more active in the high BTA-risk group (Additional file 2, Supplementary Fig. S8C). In pursuit of personalized treatment strategies, we systematically leveraged CTRP and PRISM pharmacogenomic datasets to identify compounds with differential activity in BTAscore-defined subgroups. For patients in high BTA-risk group, owing to their immunoevasive phenotype and resistance to checkpoint inhibition, prioritizing the identification of effective small-molecule interventions is clinically imperative. Notably, these individuals predicted pronounced sensitivity to agents such as BRD-K64610608, brivanib, hyperforin, adapalene, clonazepam, eltrombopag and RGFP966 (Fig. 12D-G). In parallel, the low BTA-risk group exhibited increased responsiveness to agents like cucurbitacin I, oligomycin A and ML21 (Fig. 12H).

3.11 Sodium butyrate suppresses the malignancy of HCT116 cells in a dose- and time-dependent manner

Our comprehensive single-cell transcriptomic analysis of colorectal cancer (CRC) tissues revealed a significant reduction in butyrate metabolism scores within malignant epithelial cells compared to their normal counterparts (Fig. 3C). This metabolic reprogramming was not merely an incidental observation; rather, it was critically associated with tumor progression. Patients exhibiting lower butyrate metabolism activity (lowBTA) demonstrated markedly poorer overall and disease-free survival rates (Fig. 3B). Notably, the lowBTA epithelial subpopulation (LowBTAepis) was characterized by a more undifferentiated, stem-like state, enhanced proliferative potential, and exhibited a distinct copy number variation (CNV) signature that included gains on chromosomes 19q and 7p (Fig. 4F-J).

Most significantly, our regulon activity analysis identified the transcription factor *HOXB8* as being specifically and highly active in these aggressive LowBTAepis populations (Additional file 2, Supplementary Fig. S6B). *HOXB8*, a member of the homeobox gene family, has been implicated in tumor initiation, invasion, and metastasis. Its pronounced enrichment within the poor-prognosis population characterized by low butyrate metabolism presents a compelling hypothesis: that *HOXB8* may serve as a key functional effector mediating the aggressive phenotype associated with impaired butyrate metabolism in CRC.

To experimentally validate the hypothesis derived from our bioinformatic analyses, we conducted functional studies utilizing the HCT116 colorectal cancer cell. Initially, we aimed to confirm the direct anti-tumor effects of its namesake metabolite—sodium butyrate (NaB). Treatment of HCT116 cells with varying concentrations of NaB (2, 4, and 8 mM) and timepoints (24, 48 and 72 h) demonstrated a significant inhibition of cell proliferation in a clear dose- and time-dependent manner (Fig. 13A). Beyond proliferation, functional assays revealed that a 48-hour treatment with NaB effectively suppressed both the migratory (Fig. 13B, C) and invasive (Fig. 13D, E) capabilities of HCT116 cells in a dose-dependent manner. These consistent findings across multiple phenotypic endpoints firmly establish that NaB exerts broad and potent suppressive effects on the core malignant behaviors of colorectal cancer cells.

3.12 The inhibitory effects of sodium butyrate are mediated through the downregulation of *HOXB8*.

We subsequently utilized this experimental system to mechanistically elucidate the relationship between sodium butyrate (NaB) and our computational candidate, *HOXB8*. Western blot analysis confirmed that NaB treatment resulted in a significant, dose-dependent reduction in HOXB8 protein expression (Fig. 13F), thereby establishing *HOXB8* as a potential molecular target of NaB signaling. To definitively ascertain whether the downregulation of *HOXB8* is functionally necessary for the effects of NaB, we conducted a genetic rescue experiment by overexpressing *HOXB8* (OE-*HOXB8*) in HCT116 cells. As expected, overexpression of *HOXB8* alone acted as a potent oncogenic driver, significantly enhancing cell proliferation, migration, and invasion (Fig. 14A-E). However, the critical finding was that while NaB treatment effectively suppressed these malignant behaviors in HCT116 cells, its inhibitory effects were markedly diminished in cells overexpressing *HOXB8* (Fig. 14A-E). This partial yet substantial rescue strongly indicates that the anti-tumor efficacy of NaB is primarily mediated through its capacity to suppress *HOXB8* expression.

3.13 Sodium butyrate inhibits the AKT/mTOR signaling pathway and regulates downstream functional proteins by targeting *HOXB8*.

To further elucidate the downstream signaling cascade through which the NaB-*HOXB8* axis exerts its effects, we investigated the AKT/mTOR pathway, a critical regulator of cell growth, survival, and metabolism that is frequently dysregulated in cancer. Our data demonstrated that NaB treatment significantly inhibited the phosphorylation of AKT and phosphorylation of mTOR, indicating a suppression of this key oncogenic pathway (Fig. 15A, B). Notably, overexpression of *HOXB8* itself activated the AKT/mTOR pathway and partially reversed the suppressive effect of NaB, positioning *HOXB8* upstream in the AKT/mTOR signaling cascade (Fig. 15A, B).

We then examined key downstream functional proteins regulated by this axis. On one hand, NaB treatment resulted in a decrease in the Bcl-2/Bax ratio—a crucial indicator that shifts cellular balance

towards apoptosis (Fig. 15C, D). On the other hand, NaB modulated central mediator of metastasis by MMP9 (Fig. 15E, F). This expression profile aligns with a suppression of epithelial-mesenchymal transition (EMT) and invasive capacity. Importantly, all these molecular changes induced by NaB—the pro-apoptotic shift along with anti-EMT modulation—were significantly counteracted by *HOXB8* overexpression.

Collectively, our findings from both computational biology and lab experimentation delineate a novel and coherent signaling axis in colorectal cancer. We propose that sodium butyrate suppresses aggressive cellular behaviors by directly or indirectly inhibiting the expression of the oncogenic transcription factor *HOXB8*, which is hyperactive in cell populations characterized by poor prognosis and low butyrate metabolism. This inhibition of *HOXB8* subsequently leads to the inactivation of the AKT/mTOR pathway, ultimately modulating its downstream regulators involved in apoptosis and metastasis to impede tumor progression. This mechanistic insight bridges the gap between microbial metabolite sensing, transcriptional reprogramming, and canonical oncogenic signaling, thereby providing a more comprehensive understanding of butyrate's role in colorectal cancer.

4 Discussion

This research integrates multi-algorithmic single-cell metabolic scoring strategies and functional experiments to comprehensively and systematically expound on the novel role of butyrate metabolism in the progression of CRC. Our findings demonstrate that the impairment of butyrate metabolism in tumor epithelial cells is not merely an incidental phenomenon but an active driving force behind malignancy. It is closely associated with poor prognosis, stem-cell-like characteristics, enhanced proliferation, and the remodeling of the tumor microenvironment (TME). This metabolic reprogramming operates via a well-defined NaB/*HOXB8*/AKT/mTOR signaling axis, thereby uncovering potential therapeutic targets on CRC.

Butyrate has shown remarkable anticancer efficacy in diseases such as CRC, endometrial cancer, and pancreatic cancer [87–89]. However, the underlying mechanisms vary significantly. An increasing body of evidence suggests that cancer represents a disease of cell identity crisis, where differentiated cells lose their original characteristics and acquire features typical of precursor cells [90, 91]. A central discovery of our study is the identification of a tumor epithelial cell subpopulation characterized by low butyrate-metabolism activity (LowBTAepis), which is strongly associated with poor clinical outcomes. Further characterization reveals that the LowBTAepis subpopulation exhibits a more invasive phenotype. Evidence includes their enrichment in the early stages of the pseudotime trajectory, higher stemness scores, enhanced proliferative capacity, and distinct copy number variation (CNV) characteristics. This indicates that LowBTAepis represents a highly malignant cell state with progenitor-like features, whose malignant behavior is driven by metabolic reprogramming that diverges from normal butyrate utilization.

Our research further uncovers the key molecular mediator linking abnormal butyrate metabolism to the malignant phenotype of tumors: the transcription factor *HOXB8*. *HOXB8* belongs to the highly conserved HOX gene family. As a transcription factor, it plays a pivotal role in cell proliferation, differentiation, and

migration by binding to the regulatory regions of specific genes [92]. *HOXB8* is significantly upregulated in CRC tissues, and its expression level is closely correlated with advanced clinical stages, lymph node metastasis, and distant metastasis, suggesting its role as an important oncogene [93]. To elucidate the role of *HOXB8* in butyrate metabolism, we conducted a regulatory activity analysis and discovered that *HOXB8* exhibits high activity in LowBTAepis. Subsequent functional experiments demonstrated that treatment with sodium butyrate (NaB) significantly inhibits the expression of *HOXB8* in HCT116 cells in a dose-dependent manner. Importantly, functional rescue experiments underscored the pivotal role of *HOXB8*: when *HOXB8* is forcibly overexpressed in the presence of butyrate, the anti-proliferative, anti-migratory, and anti-invasive effects induced by NaB are markedly diminished. This finding provides direct evidence that downregulation of *HOXB8* is a crucial and integral step for butyrate to exert its anti-cancer effects.

Having established the functional significance of *HOXB8*, we further explored its downstream signaling pathways. The AKT/mTOR signaling axis is one of the most frequently activated pathways in human cancers. It plays a central role in tumor growth, metastasis, and treatment resistance by promoting cell survival, proliferation, metabolic reprogramming, and resistance to apoptosis [94]. Our research results indicate that treatment with sodium butyrate (NaB) effectively inhibits the phosphorylation and activation of the AKT/mTOR pathway, and this effect is also contingent upon the downregulation of *HOXB8*. Further investigations revealed that this butyrate-*HOXB8*-dominated signaling axis, by regulating the AKT/mTOR pathway, influences a series of key apoptosis- and invasion-related proteins downstream. Specifically, it significantly reduces the ratio of the anti-apoptotic protein Bcl-2 to the pro-apoptotic protein Bax and simultaneously inhibits the expression of matrix metalloproteinase MMP-9 [95]. These changes functionally promote apoptosis in cancer cells and weaken their ability to degrade the extracellular matrix. Collectively, our research outlines a complete NaB/*HOXB8*/AKT/mTOR signaling axis.

Beyond the intracellular effects, our study also elucidates how butyrate metabolism shapes the TME through intercellular communication. CellChat analysis indicates that LowBTAepis engages in specific communication with stromal cells such as myeloid cells and fibroblasts via specific ligand-receptor pairs, including *SPP1-CD44* and the *MDK*-related signaling axis derived from fibroblasts. Previous research has shown that *SPP1* (osteopontin), a multifunctional, secreted, and phosphorylated glycoprotein, not only serves as a key factor in cell adhesion but also initiates a potent signaling event upon binding to *CD44*. This signaling axis has been shown to promote tumor cell survival, proliferation, and chemoresistance by activating the PI3K-AKT and MAPK pathways, while directly regulating the polarization of myeloid cells towards an immunosuppressive M2-like phenotype [96]. Midkine (MDK), a heparin-binding growth factor enriched in cancer-associated fibroblasts (CAFs), activates multiple pro-survival pathways, including PI3K-AKT, by binding to its receptors (such as ALK, PTPRZ1, etc.), directly stimulating tumor growth and participating in angiogenesis and immune regulation [97]. Spatial transcriptomics analysis further validates the close spatial colocalization between LowBTAepis and fibroblasts in tissue sections and, through the MISTy framework, reveals the coordinated activation and spatial hierarchical interactions of multiple oncogenic pathways (such as TNFα, TGFβ, EGFR, MAPK, NF-

κB, and hypoxia signaling) within LowBTAepis-enriched regions. These pathways engage in crosstalk, jointly promoting tumor progression [98, 99]. These findings underscore the central role of butyrate metabolic status in coordinating intercellular communication and the signaling pathway network within the TME.

Machine learning approaches have demonstrated substantial potential in predicting survival, recurrence, and metastasis in CRC, holding significant clinical translational implications [100, 101]. To translate basic research findings into potential clinical applications, we utilized machine learning algorithms to develop and validate a robust BTA risk score model based on the signature genes of LowBTAepis. This model has demonstrated excellent prognostic prediction performance across multiple independent cohorts. Notably, high-risk scores are associated with upregulated expression of immune checkpoint molecules and increased infiltration of effector immune cells. However, this is accompanied by higher TIDE scores and lower immunotherapy response rates, suggesting a complex immune state of "immune enrichment with functional inhibition" within this subgroup. This may be attributed to the fact that although effector immune cells are recruited to the tumor site, their functions are rendered inactive or exhausted due to the combined effects of an inhibitory metabolic environment and immunosuppressive cells, ultimately leading to resistance to immunotherapy [102]. Leveraging pharmacogenomic databases, we employed computational biology methods to predict potential small-molecule therapeutic strategies for patients with high-risk BTA scores. The analysis results suggest that drugs such as Brivanib (a dual inhibitor of FGFR/VEGFR) and RGFP966 (a selective HDAC3 inhibitor) may exhibit higher sensitivity in patient models with this molecular signature [103, 104].

This study has certain limitations. Despite integrating multiple datasets, the sample size remains a factor that requires further expansion. The upstream mechanism by which butyrate precisely regulates *HOXB8* remains an area that demands further exploration. Additionally, the generalizability of the BTA risk model across different ethnic groups and cancer stages represents an important avenue for future research.

5 Conclusions

In conclusion, this study reveals that butyrate metabolism drives the malignant evolution of an invasive epithelial cell subpopulation with high stemness and proliferation potential, mediated by *HOXB8*. It also shapes the microenvironment conducive to tumor growth by remodeling intercellular communication and the signaling pathway network. The NaB/*HOXB8*/AKT/mTOR axis elucidated herein and the developed BTA risk model offer invaluable insights and resources for understanding the pathogenesis of CRC and for the development of novel metabolic intervention and stratified treatment strategies.

Abbreviations

CRC
Colorectal cancer

SCFA
Short-chain fatty acid
scRNA-seq
Single-cell RNA sequencing
ST
Spatial transcriptomics
TME
Tumor microenvironment
CAF
Cancer-associated fibroblast
ECM
Extracellular matrix
CNV
Copy-number variation
UMAP
Uniform Manifold Approximation and Projection
BTA
Butyrate metabolism
HighBTAepis
Tumor epithelial cells with high butyrate metabolism scores
LowBTAepis
Tumor epithelial cells with low butyrate metabolism scores
CSI
Connection Specificity Index
RSS
Regulon specificity score
RSF
Random survival forest
CTRP
Cancer Therapeutics Response Portal
TIDE
Tumor Immune Dysfunction and Exclusion
NaB
Sodium butyrate
HDAC
Histone deacetylase

Declarations

Ethics approval and consent to participate

Ethics approval and informed consent were not required for the present study because all human transcriptomic and clinical data were obtained from publicly available, de-identified datasets. No new human participants or human specimens were recruited or collected for this study.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China [grant number 82505473], Beijing Municipal Science and Technology Commission [Support for high-level innovative and entrepreneurial talents], Xuanwu Hospital Talent Convergence Program [grant number HZ2025ZCBJ004] and the Fundamental Research Funds for the Central Universities [grant number 3332024008].

Author Contribution

LLW and MQ jointly conceived and designed the study. JYF and TQL performed data interpretation, statistical analyses, figure preparation, and drafted the manuscript. LLW contributed to statistical analysis and critically revised the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgements

We sincerely thank all colleagues and collaborators who provided valuable support and assistance during this study.

Availability of data and materials

The gene expression datasets GSE144735, GSE166555, GSE132465, GSE225857, GSE17536, GSE39582, and GSE91061 were retrieved from public databases, including GEO (https://www.ncbi.nlm.nih.gov/geo/) and the TIGER database (http://tiger.canceromics.org/; PRJEB23709). The TCGA (COAD and READ) datasets were obtained via the UCSC Xena data portal (https://xenabrowser.net/). Data generated during this study are included in the article and its Supplementary Information. Additional raw experimental data are available from the corresponding author upon reasonable request.

References

1. Cosso F, Lavacchi D, Fancelli S, Caliman E, Brugia M, Rossi G, et al. Long-term response of more than 9 years to regorafenib in a heavily pretreated patient with metastatic colorectal cancer. Anticancer Drugs. 2023;34(3):451–4.
2. Cavestro GM, Mannucci A, Zuppardo RA, Di Leo M, Stoffel E, Tonon G. Early onset sporadic colorectal cancer: Worrisome trends and oncogenic features. Dig Liver Dis. 2018;50(6):521–32.
3. Fu MS, Pan SX, Cai XQ, Pan QC. Urban vs. rural: colorectal cancer survival and prognostic disparities from 2000 to 2019. Front Public Health. 2024;12:1319977.
4. Boutin AT, Liao WT, Wang M, Hwang SS, Karpinets TV, Cheung H, et al. Oncogenic Kras drives invasion and maintains metastases in colorectal cancer. Genes Dev. 2017;31(4):370–82.
5. Nguyen LH, Goel A, Chung DC. Pathways of Colorectal Carcinogenesis. Gastroenterology. 2020;158(2):291–302.
6. Rudolph KL, Millard M, Bosenberg MW, DePinho RA. Telomere dysfunction and evolution of intestinal carcinoma in mice and humans. Nat Genet. 2001;28(2):155–9.
7. Mouradov D, Greenfield P, Li S, In EJ, Storey C, Sakthianandeswaren A, et al. Oncomicrobial Community Profiling Identifies Clinicomolecular and Prognostic Subtypes of Colorectal Cancer. Gastroenterology. 2023;165(1):104–20.
8. Ogino S, Kawasaki T, Kirkner GJ, Ohnishi M, Fuchs CS. 18q loss of heterozygosity in microsatellite stable colorectal cancer is correlated with CpG island methylator phenotype-negative (CIMP-0) and inversely with CIMP-low and CIMP-high. BMC Cancer. 2007;7:72.
9. Wielandt AM, Hurtado C, Moreno CM, Villarroel C, Castro M, Estay M, et al. Characterization of Chilean patients with sporadic colorectal cancer according to the three main carcinogenic pathways: Microsatellite instability, CpG island methylator phenotype and Chromosomal instability. Tumour Biol. 2020;42(7):1010428320938492.
10. Louis P, Flint HJ. Formation of propionate and butyrate by the human colonic microbiota. Environ Microbiol. 2017;19(1):29–41.
11. Liu H, Wang J, He T, Becker S, Zhang G, Li D, et al. Butyrate: A Double-Edged Sword for Health? Adv Nutr. 2018;9(1):21–9.
12. Salvi PS, Cowles RA. Butyrate and the Intestinal Epithelium: Modulation of Proliferation and Inflammation in Homeostasis and Disease. Cells. 2021;10(7).
13. Xiao M, Liu YG, Zou MC, Zou F. Sodium butyrate induces apoptosis of human colon cancer cells by modulating ERK and sphingosine kinase 2. Biomed Environ Sci. 2014;27(3):197–203.
14. Geng HW, Yin FY, Zhang ZF, Gong X, Yang Y. Butyrate Suppresses Glucose Metabolism of Colorectal Cancer Cells via GPR109a-AKT Signaling Pathway and Enhances Chemotherapy. Front Mol Biosci. 2021;8:634874.
15. Ferreira TM, Leonel AJ, Melo MA, Santos RR, Cara DC, Cardoso VN, et al. Oral supplementation of butyrate reduces mucositis and intestinal permeability associated with 5-Fluorouracil

administration. Lipids. 2012;47(7):669–78.
16. Kobayashi H, Enomoto A, Woods SL, Burt AD, Takahashi M, Worthley DL. Cancer-associated fibroblasts in gastrointestinal cancer. Nat Rev Gastroenterol Hepatol. 2019;16(5):282–95.
17. Lee HO, Hong Y, Etlioglu HE, Cho YB, Pomella V, Van den Bosch B, et al. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. Nat Genet. 2020;52(6):594–603.
18. Cantile M, Scognamiglio G, La Sala L, La Mantia E, Scaramuzza V, Valentino E, et al. Aberrant expression of posterior HOX genes in well differentiated histotypes of thyroid cancers. Int J Mol Sci. 2013;14(11):21727–40.
19. Haria D, Naora H. Homeobox Gene Deregulation: Impact on the Hallmarks of Cancer. Cancer Hallm. 2013;1(2–3):67–76.
20. Hur H, Lee JY, Yun HJ, Park BW, Kim MH. Analysis of HOX gene expression patterns in human breast cancer. Mol Biotechnol. 2014;56(1):64–71.
21. Kanai M, Hamada J, Takada M, Asano T, Murakawa K, Takahashi Y, et al. Aberrant expressions of HOX genes in colorectal and hepatocellular carcinomas. Oncol Rep. 2010;23(3):843–51.
22. Li X, Lin H, Jiang F, Lou Y, Ji L, Li S. Knock-Down of HOXB8 Prohibits Proliferation and Migration of Colorectal Cancer Cells via Wnt/beta-Catenin Signaling Pathway. Med Sci Monit. 2019;25:711–20.
23. Marra L, Cantile M, Scognamiglio G, Perdona S, La Mantia E, Cerrone M, et al. Deregulation of HOX B13 expression in urinary bladder cancer progression. Curr Med Chem. 2013;20(6):833–9.
24. Lazarova DL, Chiaro C, Bordonaro M. Butyrate induced changes in Wnt-signaling specific gene expression in colorectal cancer cells. BMC Res Notes. 2014;7:226.
25. Uhlitz F, Bischoff P, Peidli S, Sieber A, Trinks A, Luthen M, et al. Mitogen-activated protein kinase activity drives cell trajectories in colorectal cancer. EMBO Mol Med. 2021;13(10):e14123.
26. Goldman MJ, Craft B, Hastie M, Repecka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol. 2020;38(6):675–8.
27. Yang C, Huang X, Liu Z, Qin W, Wang C. Metabolism-associated molecular classification of hepatocellular carcinoma. Mol Oncol. 2020;14(4):896–913.
28. Kurata N, Tokashiki N, Fukushima K, Misao T, Hasuoka N, Kitagawa K, et al. Short chain fatty acid butyrate uptake reduces expressions of prostanoid EP(4) receptors and their mediation of cyclooxygenase-2 induction in HCA-7 human colon cancer cells. Eur J Pharmacol. 2019;853:308–15.
29. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573–e873529.
30. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 2019;20(1):296.
31. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16(12):1289–96.

32. Zheng H, Liu H, Li H, Dou W, Wang J, Zhang J, et al. Characterization of stem cell landscape and identification of stemness-relevant prognostic gene signature to aid immunotherapy in colorectal cancer. Stem Cell Res Ther. 2022;13(1):244.
33. Hou W, Ji Z. Assessing GPT-4 for cell type annotation in single-cell RNA-seq analysis. Nat Methods. 2024;21(8):1462–5.
34. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021;12(1):1088.
35. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014;32(4):381–6.
36. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science. 2014;344(6190):1396–401.
37. Gulati GS, Sikandar SS, Wesche DJ, Manjunath A, Bharadwaj A, Berger MJ, et al. Single-cell transcriptional diversity is a hallmark of developmental potential. Science. 2020;367(6476):405–11.
38. Alghamdi N, Chang W, Dang P, Lu X, Wan C, Gampala S, et al. A graph neural network model to estimate cell-wise metabolic flux using single-cell RNA-seq data. Genome Res. 2021;31(10):1867–84.
39. Wang F, Long J, Li L, Wu ZX, Da TT, Wang XQ, et al. Single-cell and spatial transcriptome analysis reveals the cellular heterogeneity of liver metastatic colorectal cancer. Sci Adv. 2023;9(24):eadf5464.
40. Kleshchevnikov V, Shmatko A, Dann E, Aivazidis A, King HW, Li T, et al. Cell2location maps fine-grained cell types in spatial transcriptomics. Nat Biotechnol. 2022;40(5):661–71.
41. Holland CH, Szalai B, Saez-Rodriguez J. Transfer of regulatory knowledge from human to mouse for functional genomics analysis. Biochim Biophys Acta Gene Regul Mech. 2020;1863(6):194431.
42. Schubert M, Klinger B, Klunemann M, Sieber A, Uhlitz F, Sauer S, et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. Nat Commun. 2018;9(1):20.
43. Tanevski J, Flores ROR, Gabor A, Schapiro D, Saez-Rodriguez J. Explainable multiview framework for dissecting spatial relationships from highly multiplexed data. Genome Biol. 2022;23(1):97.
44. Pham D, Tan X, Balderson B, Xu J, Grice LF, Yoon S, et al. Robust mapping of spatiotemporal trajectories and cell-cell interactions in healthy and diseased tissues. Nat Commun. 2023;14(1):7739.
45. Kumar N, Mishra B, Athar M, Mukhtar S. Inference of Gene Regulatory Network from Single-Cell Transcriptomic Data Using pySCENIC. Methods Mol Biol. 2021;2328:171–82.
46. Fuxman Bass JI, Diallo A, Nelson J, Soto JM, Myers CL, Walhout AJ. Using networks to measure similarity between genes: association index selection. Nat Methods. 2013;10(12):1169–76.
47. Suo S, Zhu Q, Saadatpour A, Fei L, Guo G, Yuan GC. Revealing the Critical Regulators of Cell Identity in the Mouse Cell Atlas. Cell Rep. 2018;25(6):1436–e451433.

48. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol. 2019;37(7):773–82.
49. Lin Y, Pan X, Chen Z, Lin S, Shen Z, Chen S. Prognostic value and immune infiltration of novel signatures in colon cancer microenvironment. Cancer Cell Int. 2021;21(1):679.
50. Kazmierczak-Siedlecka K, Marano L, Merola E, Roviello F, Polom K. Sodium butyrate in both prevention and supportive treatment of colorectal cancer. Front Cell Infect Microbiol. 2022;12:1023806.
51. Donohoe DR, Collins LB, Wali A, Bigler R, Sun W, Bultman SJ. The Warburg effect dictates the mechanism of butyrate-mediated histone acetylation and cell proliferation. Mol Cell. 2012;48(4):612–26.
52. Donohoe DR, Garge N, Zhang X, Sun W, O'Connell TM, Bunger MK, et al. The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. Cell Metab. 2011;13(5):517–26.
53. Donohoe DR, Wali A, Brylawski BP, Bultman SJ. Microbial regulation of glucose metabolism and cell-cycle progression in mammalian colonocytes. PLoS ONE. 2012;7(9):e46589.
54. Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, et al. The genomic landscapes of human breast and colorectal cancers. Science. 2007;318(5853):1108–13.
55. Gomes S, Rodrigues AC, Pazienza V, Preto A. Modulation of the Tumor Microenvironment by Microbiota-Derived Short-Chain Fatty Acids: Impact in Colorectal Cancer Therapy. Int J Mol Sci. 2023;24(6).
56. Park J, Kim M, Kang SG, Jannasch AH, Cooper B, Patterson J, et al. Short-chain fatty acids induce both effector and regulatory T cells by suppression of histone deacetylases and regulation of the mTOR-S6K pathway. Mucosal Immunol. 2015;8(1):80–93.
57. Klement JD, Paschall AV, Redd PS, Ibrahim ML, Lu C, Yang D, et al. An osteopontin/CD44 immune checkpoint controls CD8 + T cell activation and tumor immune evasion. J Clin Invest. 2018;128(12):5549–60.
58. Xu C, Zhu S, Wu M, Han W, Yu Y. Functional receptors and intracellular signal pathways of midkine (MK) and pleiotrophin (PTN). Biol Pharm Bull. 2014;37(4):511–20.
59. Kadomatsu K, Muramatsu T. Midkine and pleiotrophin in neural development and cancer. Cancer Lett. 2004;204(2):127–43.
60. Yildirim B, Kulak K, Bilir A. Midkine (MDK) in cancer and drug resistance: from inflammation to therapy. Discov Oncol. 2025;16(1):1062.
61. Hu B, Qin C, Li L, Wei L, Mo X, Fan H, et al. Midkine promotes glioblastoma progression via PI3K-Akt signaling. Cancer Cell Int. 2021;21(1):509.
62. Van de Sande B, Flerin C, Davie K, De Waegeneer M, Hulselmans G, Aibar S, et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat Protoc. 2020;15(7):2247–76.

63. Slattery ML, Lundgreen A, Kadlubar SA, Bondurant KL, Wolff RK. JAK/STAT/SOCS-signaling pathway and colon and rectal cancer. Mol Carcinog. 2013;52(2):155–66.
64. Sun Q, Qu W, Wang Y, Yang K, Weng Y. TFAP2A upregulates FAM83A to suppress ferroptosis and diminish cisplatin sensitivity in non-small cell lung cancer. Cell Div. 2025;20(1):19.
65. Yang S, Xiang J, Ma C, Yang G, Wang X, Liu H, et al. Sp1-like protein KLF13 acts as a negative feedback regulator of TGF-beta signaling and fibrosis. Cell Rep. 2023;42(4):112367.
66. Botella LM, Sanz-Rodriguez F, Komi Y, Fernandez LA, Varela E, Garrido-Martin EM, et al. TGF-beta regulates the expression of transcription factor KLF6 and its splice variants and promotes co-operative transactivation of common target genes through a Smad3-Sp1-KLF6 interaction. Biochem J. 2009;419(2):485–95.
67. Chang LC, Fan CW, Tseng WK, Chen JR, Chein HP, Hwang CC, et al. Immunohistochemical study of the Nrf2 pathway in colorectal cancer: Nrf2 expression is closely correlated to Keap1 in the tumor and Bach1 in the normal tissue. Appl Immunohistochem Mol Morphol. 2013;21(6):511–7.
68. Han A, Bennett N, Ahmed B, Whelan J, Donohoe DR. Butyrate decreases its own oxidation in colorectal cancer cells through inhibition of histone deacetylases. Oncotarget. 2018;9(43):27280–92.
69. Sun J, Chen S, Zang D, Sun H, Sun Y, Chen J. Butyrate as a promising therapeutic target in cancer: From pathogenesis to clinic (Review). Int J Oncol. 2024;64(4).
70. Koveitypour Z, Panahi F, Vakilian M, Peymani M, Seyed Forootan F, Nasr Esfahani MH, et al. Signaling pathways involved in colorectal cancer progression. Cell Biosci. 2019;9:97.
71. Zhang YE, Stuelten CH. Alternative splicing in EMT and TGF-beta signaling during cancer progression. Semin Cancer Biol. 2024;101:1–11.
72. Krasinskas AM. EGFR Signaling in Colorectal Carcinoma. Patholog Res Int. 2011;2011:932932.
73. Zhang H, Nie J, Bao Z, Shi Y, Gong J, Li H. FOXC1 promotes EMT and colorectal cancer progression by attracting M2 macrophages via the TGF-beta/Smad2/3/snail pathway. Cell Signal. 2025;130:111680.
74. Wang J, Gan L, Liu F, Yang Q, Deng Q, Jiang D, et al. USP10 promotes pancreatic ductal adenocarcinoma progression by attenuating FOXC1 protein degradation to activate the WNT signaling pathway. Int J Biol Sci. 2024;20(13):5343–62.
75. Saitoh M. Transcriptional regulation of EMT transcription factors in cancer. Semin Cancer Biol. 2023;97:21–9.
76. Rogerson C, Ogden S, Britton E, Consortium O, Ang Y, Sharrocks AD. Repurposing of KLF5 activates a cell cycle signature during the progression from a precursor state to oesophageal adenocarcinoma. Elife. 2020;9.
77. Park SY, Seo YR, Ko MJ, Lee JH, Chun KS, Kim MJ, et al. Targeting CALM2 Inhibits Hepatocellular Carcinoma Growth and Metastasis by Suppressing E2F5-mediated Cell Cycle Progression. Anticancer Res. 2021;41(3):1315–25.

78. Romer AI, Singer RA, Sui L, Egli D, Sussel L. Murine Perinatal beta-Cell Proliferation and the Differentiation of Human Stem Cell-Derived Insulin-Expressing Cells Require NEUROD1. Diabetes. 2019;68(12):2259–71.
79. Breyer JP, Dorset DC, Clark TA, Bradley KM, Wahlfors TA, McReynolds KM, et al. An expressed retrogene of the master embryonic stem cell gene POU5F1 is associated with prostate cancer susceptibility. Am J Hum Genet. 2014;94(3):395–404.
80. Yu M, Zhan J, Zhang H. HOX family transcription factors: Related signaling pathways and post-translational modifications in cancer. Cell Signal. 2020;66:109469.
81. Onyeisi JOS, Nader HB, Lopes CC. Effects of syndecan-4 silencing on the extracellular matrix remodeling in anoikis-resistant endothelial cells. Cell Biol Int. 2024;48(6):883–97.
82. Du R, Huang C, Chen H, Liu K, Xiang P, Yao N, et al. SDCBP/MDA-9/syntenin phosphorylation by AURKA promotes esophageal squamous cell carcinoma progression through the EGFR-PI3K-Akt signaling pathway. Oncogene. 2020;39(31):5405–19.
83. Hou H, Li J, Wang J, Zhou L, Li J, Liang J, et al. ITGA9 Inhibits Proliferation and Migration of Dermal Microvascular Endothelial Cells in Psoriasis. Clin Cosmet Investig Dermatol. 2022;15:2795–806.
84. He X, Wang L, Riedel H, Wang K, Yang Y, Dinu CZ, et al. Mesothelin promotes epithelial-to-mesenchymal transition and tumorigenicity of human lung cancer and mesothelioma cells. Mol Cancer. 2017;16(1):63.
85. Saini U, Smith BQ, Dorayappan KDP, Yoo JY, Maxwell GL, Kaur B, et al. Targeting TMEM205 mediated drug resistance in ovarian clear cell carcinoma using oncolytic virus. J Ovarian Res. 2022;15(1):130.
86. Fang F, Zhang T, Lei H, Shen X. TMEM200A is a potential prognostic biomarker and correlated with immune infiltrates in gastric cancer. PeerJ. 2023;11:e15613.
87. Zhu L, Chen G, Huang C, Gao H, Wang Y, Shen Y. SUMO3 inhibition by butyric acid suppresses cell viability and glycolysis and promotes gemcitabine antitumor activity in pancreatic cancer. Biol Direct. 2024;19(1):74.
88. Zhang Q, Qin Y, Sun X, Bian Z, Liu L, Liu H, et al. Sodium butyrate blocks the growth of colorectal cancer by inhibiting the aerobic glycolysis mediated by SIRT4/HIF-1α. Chem Biol Interact. 2024;403:111227.
89. Tao M, Wu T, Zhou X, Li F, Chen Y, Ling K, et al. Butyrate enhances gut dysbiosis by activating the cAMP/PKA/CREB signaling pathway to inhibit the progression of endometrial carcinoma. BMC Microbiol. 2025;25(1):516.
90. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov. 2022;12(1):31–46.
91. Benhassoun R, Morel AP, Jacquot V, Puisieux A, Ouzounova M. The epipliancy journey: Tumor initiation at the mercy of identity crisis and epigenetic drift. Biochim Biophys Acta Rev Cancer. 2025;1880(3):189307.
92. de Bessa Garcia SA, Araújo M, Pereira T, Mouta J, Freitas R. HOX genes function in Breast Cancer development. Biochim Biophys Acta Rev Cancer. 2020;1873(2):188358.

93. Abuderman AA, Harb OA, Gertallah LM. Prognostic and clinic-pathological significances of HOXB8, ILK and FAT4 expression in colorectal cancer. Contemp Oncol (Pozn). 2020;24(3):183–92.
94. Glaviano A, Foo ASC, Lam HY, Yap KCH, Jacot W, Jones RH, et al. PI3K/AKT/mTOR signaling transduction pathway and targeted therapies in cancer. Mol Cancer. 2023;22(1):138.
95. Li J, Chen Y, Hao Z, Zhang Z, Fan J, Liu X, et al. RAD23B Promotes Colorectal Cancer Metastasis via the Talin1/Integrin/PI3K/AKT/MMP9 Axis. Oncol Res. 2025;33(11):3523–41.
96. Ding D, Li W, Ren J, Li Y, Jiang G, Liang R et al. SPP1(high) macrophage-induced T-cell stress promotes colon cancer liver metastasis through SPP1/CD44/PI3K/AKT signaling. J Immunother Cancer. 2025;13(10).
97. Ito S, Sakakibara-Konishi J, Sato M, Shoji T, Furuta M, Takahashi H, et al. Midkine Promotes Tumor Growth and Attenuates the Effect of Cisplatin in Small Cell Lung Cancer. Cancer Med. 2025;14(13):e71034.
98. Ma M, Shi F, Zhai R, Wang H, Li K, Xu C, et al. TGF-β promote epithelial-mesenchymal transition via NF-κB/NOX4/ROS signal pathway in lung cancer cells. Mol Biol Rep. 2021;48(3):2365–75.
99. Taniguchi K, Karin M. NF-κB, inflammation, immunity and cancer: coming of age. Nat Rev Immunol. 2018;18(5):309–24.
100. Buk Cardoso L, Cunha Parro V, Verzinhasse Peres S, Curado MP, Fernandes GA, Wünsch Filho V, et al. Machine learning for predicting survival of colorectal cancer patients. Sci Rep. 2023;13(1):8874.
101. Zhao W, Li T, Zong J, Zhang F, Zhou M, Hou Y et al. Crosstalk between the wnt signaling pathway and axon guidance molecules reveals a novel signature for accurately predicting diagnosis, prognosis, and immune landscape in colorectal cancer. Int J Surg. 2025.
102. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med. 2018;24(10):1550–8.
103. Hong CS, Sun EG, Choi JN, Kim DH, Kim JH, Ryu KH, et al. Fibroblast growth factor receptor 4 increases epidermal growth factor receptor (EGFR) signaling by inducing amphiregulin expression and attenuates response to EGFR inhibitors in colon cancer. Cancer Sci. 2020;111(9):3268–78.
104. Lasko LM, Jakob CG, Edalji RP, Qiu W, Montgomery D, Digiammarino EL, et al. Discovery of a selective catalytic p300/CBP inhibitor that targets lineage-specific tumours. Nature. 2017;550(7674):128–32.

Figures

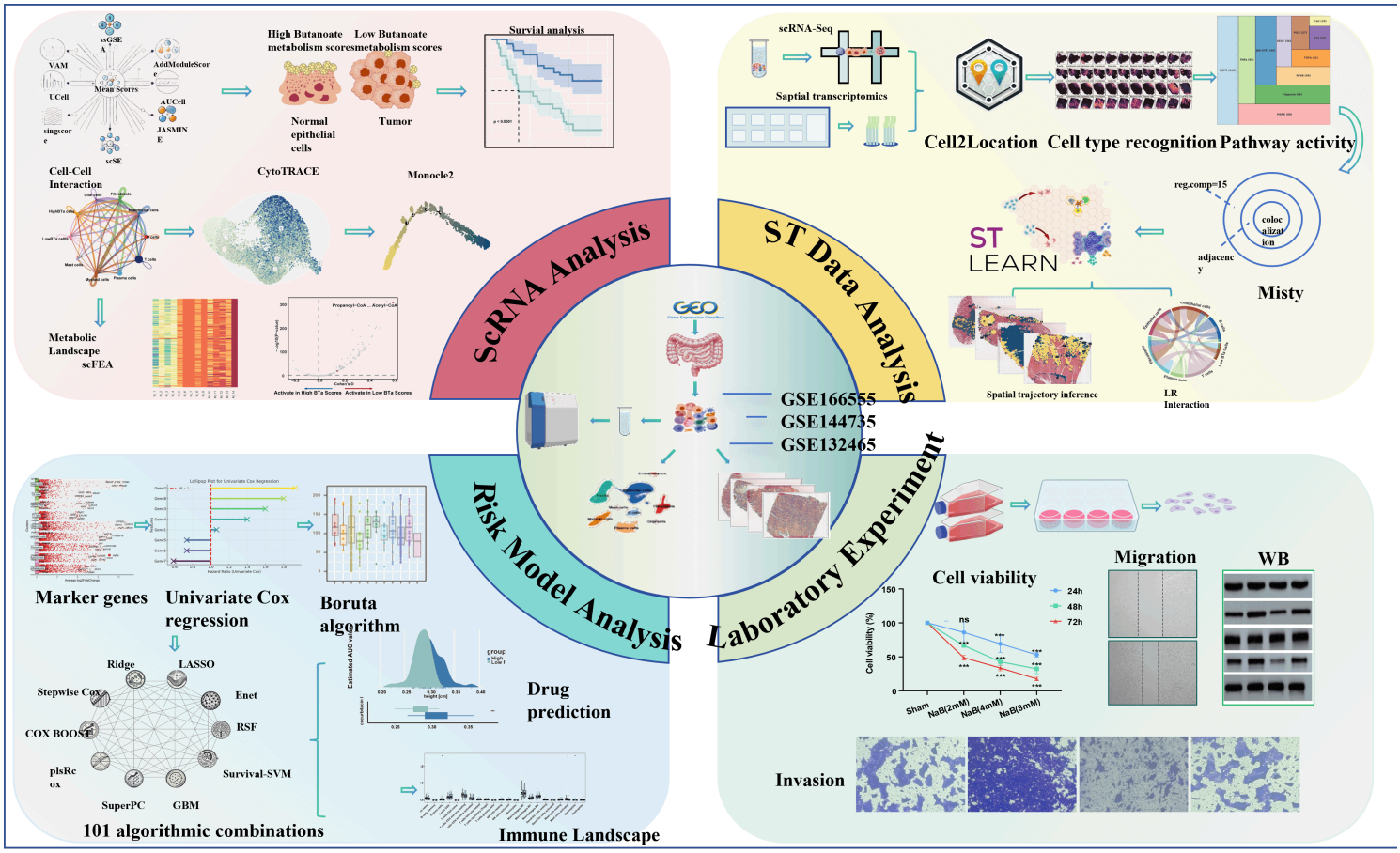


Figure 1

Overview of the Study Workflow

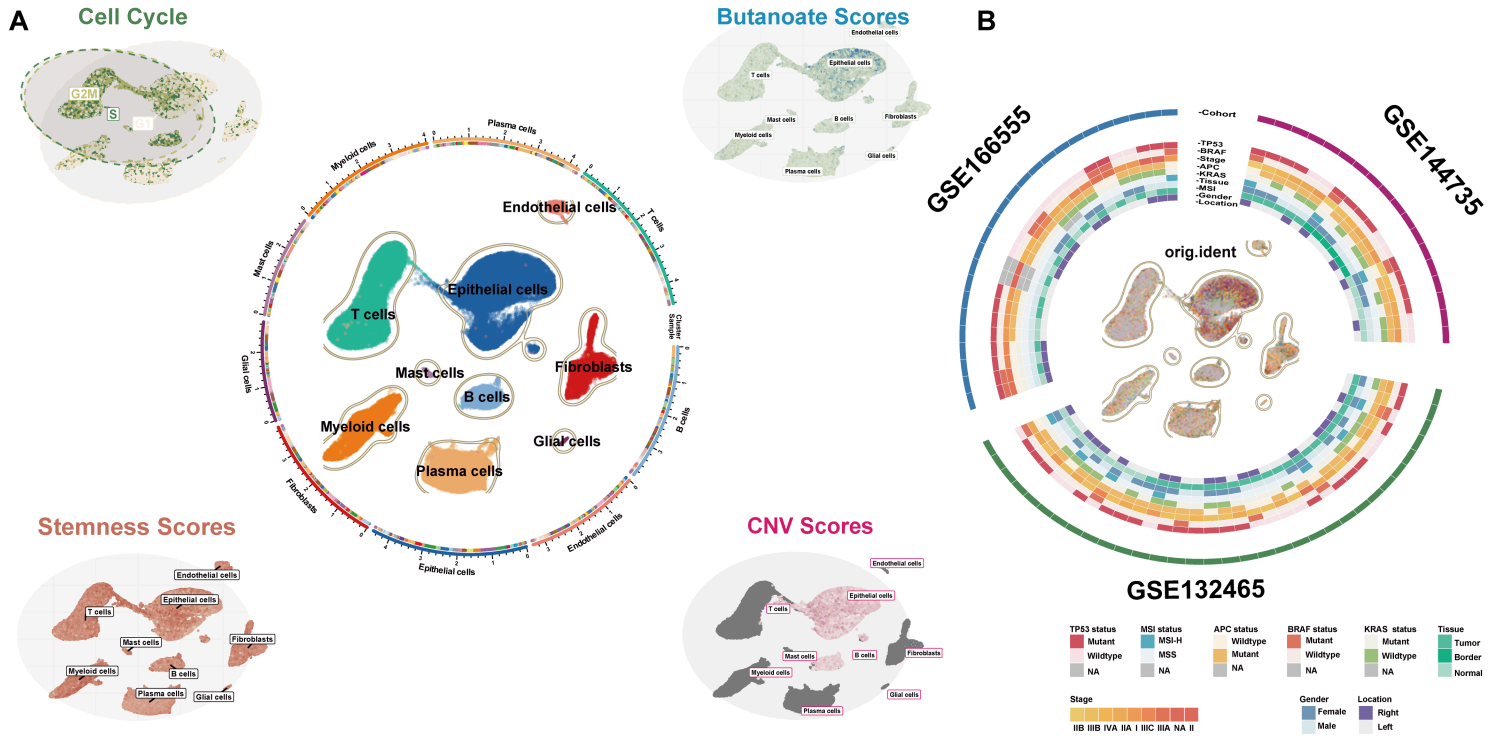


Figure 2

Single-Cell Transcriptional Landscape of Colorectal Cancer

(A) Uniform Manifold Approximation and Projection (UMAP) visualization of colorectal cancer patients, displaying distinct cell populations. Clockwise from the top left, inset panels depict variations in cell cycle status, butanoate metabolism score, stemness score, and CNV score, providing insights into functional and genomic diversity among tumor cells.

(B) UMAP representation of single-cell transcriptomic data, colored by patient ID. This plot illustrates patient-specific clustering, highlighting the impact of batch effect correction using Harmony and capturing inter-patient variability. The colored tracks (from outside to inside) represent *TP53* status, *BRAF*mutation status, tumor stage, *APC*mutation status, *KRAS* mutation status, tissue origin, MSI status, gender, and tumor location, offering a comprehensive view of the clinical and genetic landscape.

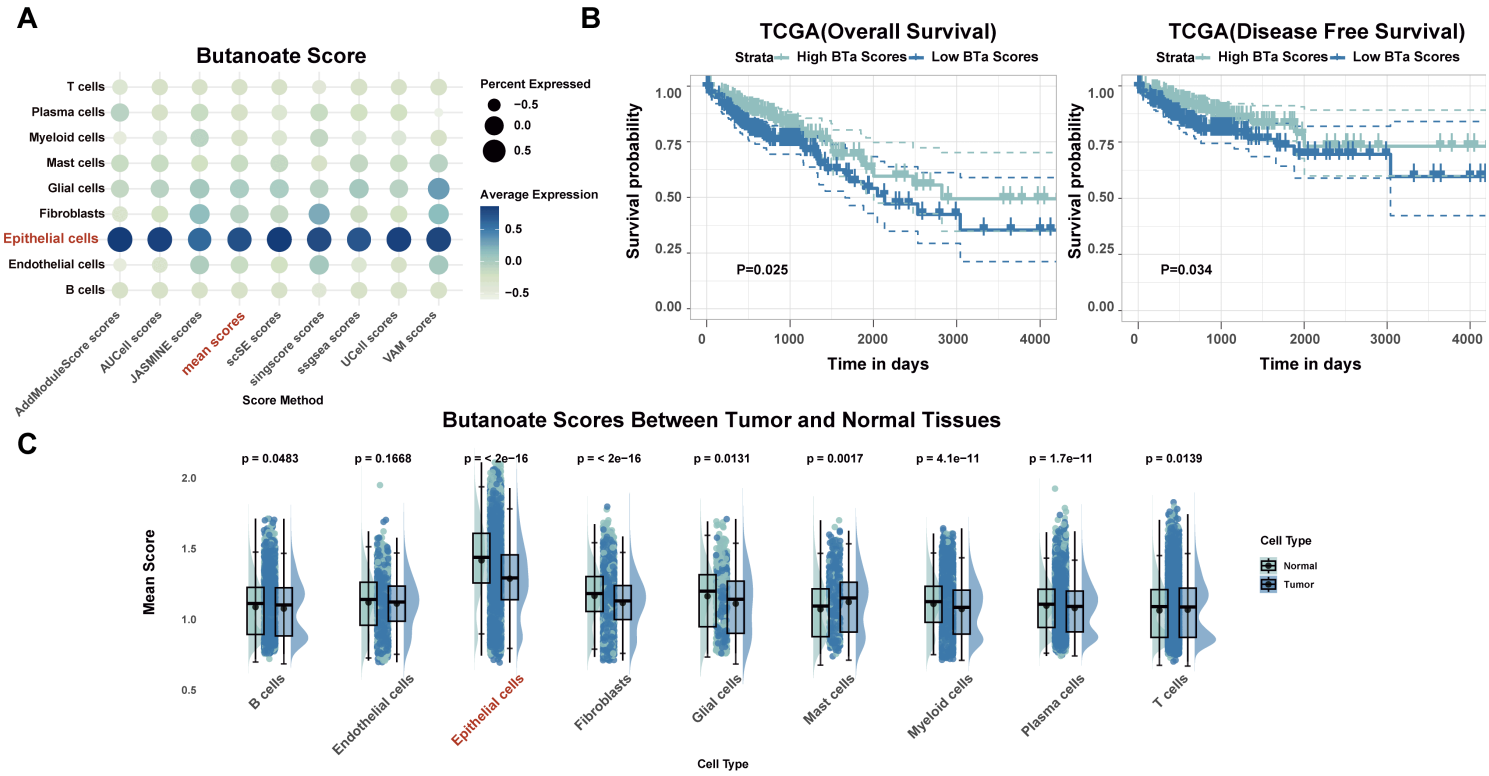


Figure 3

Decreased Butanoate Metabolism Score in Tumor Epithelial Cells and Its Prognostic Significance

(A) Bubble plot displaying butanoate metabolism scores across different cell types using various computational scoring approaches.

(B) Kaplan–Meier survival curves illustrating the prognostic impact of butanoate metabolism scores on overall survival (OS) and disease-free survival (DFS) in the TCGA cohort.

(C) Violin plots depicting the distribution of butanoate metabolism scores across various cell types, highlighting metabolic heterogeneity in colorectal cancer.

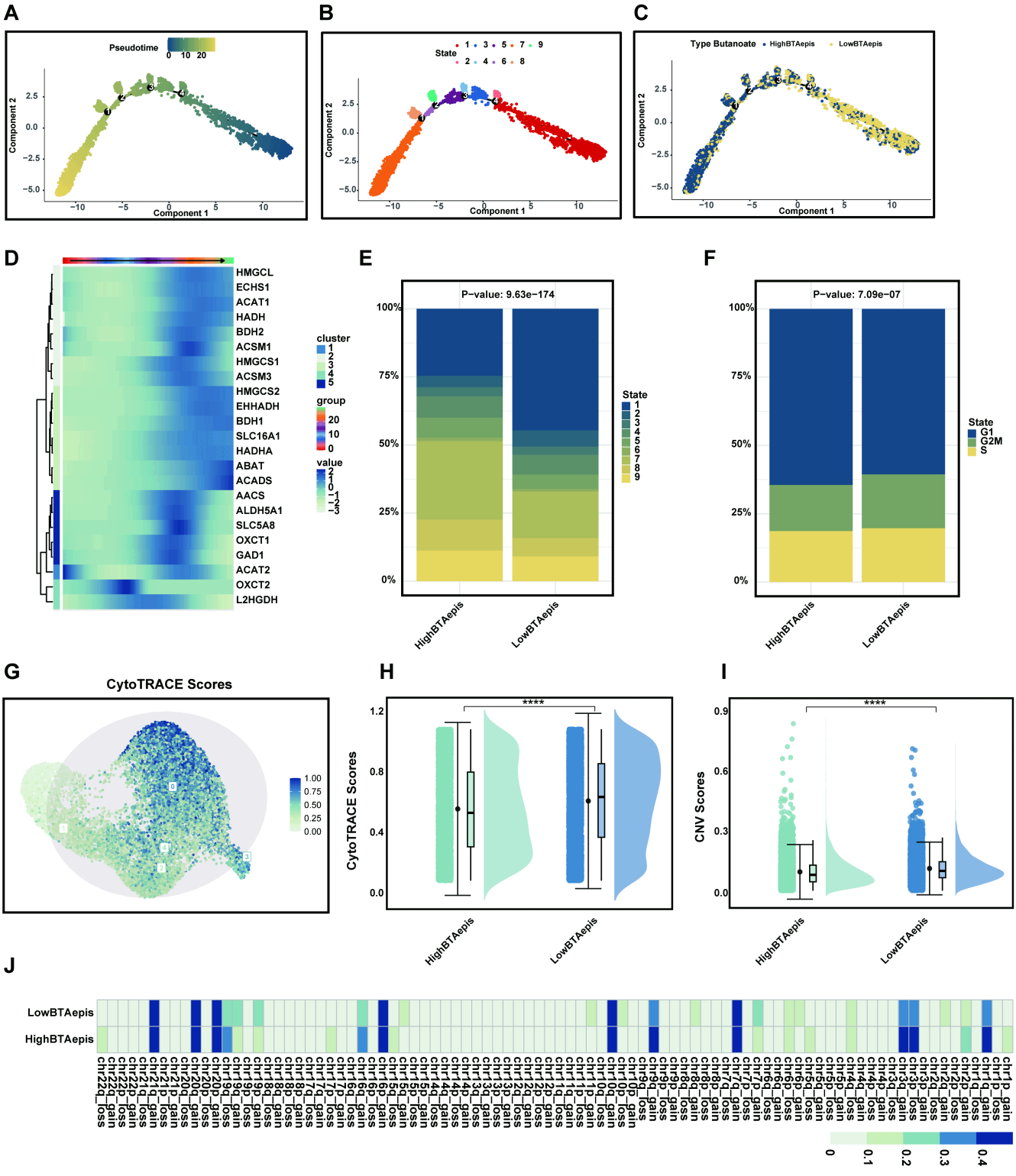


Figure 4

Evolutionary Trajectories of Tumor Epithelial Cells with Distinct Butanoate Metabolism Profiles

(A-C) Pseudotime trajectory analysis of tumor epithelial cells using Monocle 2. Each dot represents a single cell, color-coded by its inferred differentiation timeline (A), pseudotime-defined cellular state (B), and high or low butanoate metabolism (C).

(D) Heatmap illustrating the temporal expression dynamics of butanoate metabolism-associated genes along the pseudotime trajectory.

(E-F) Bar plots comparing pseudotime state distributions (E) and cell cycle phase proportions (F) between tumor epithelial cells with high and low butanoate metabolism scores.

(G) CytoTRACE analysis quantifying differentiation potential in tumor epithelial cells.

(H-I) Raincloud plots illustrating differences in CytoTRACE scores (H) and CNV burden (I) between tumor epithelial cells with high and low butanoate metabolism scores, revealing their association with differentiation potential and genomic alterations.

(J) Heatmap displaying inferCNV profiles of tumor epithelial cells with high and low butanoate metabolism scores, highlighting genomic instability patterns.

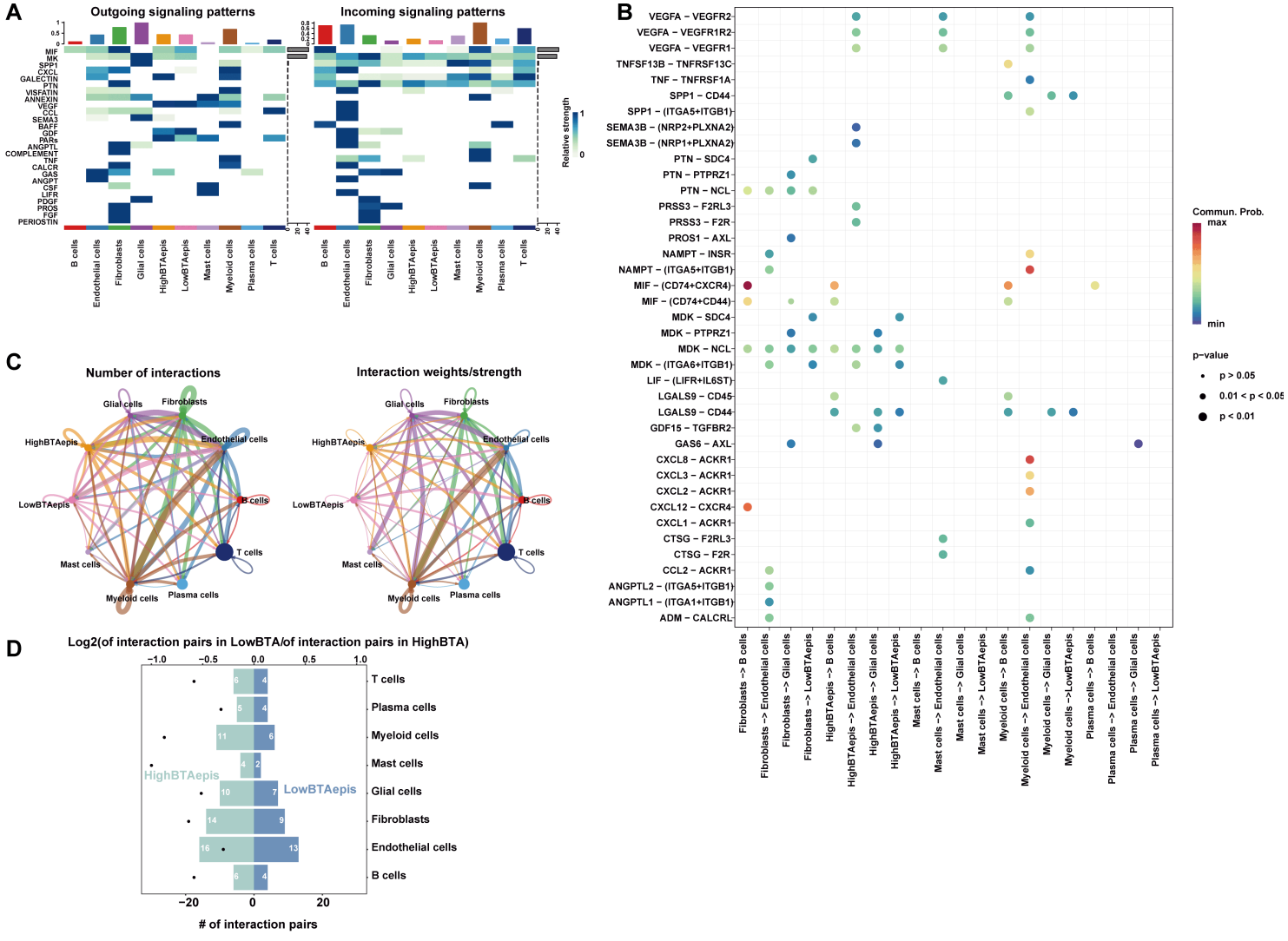


Figure 5

Cell-Cell Communication in the Tumor Microenvironment

(A) Inferred incoming and outgoing signaling patterns among distinct cell types in the colorectal cancer tumor microenvironment (TME), highlighting communication differences between tumor epithelial cells with high and low butanoate metabolism scores. The visualization highlights the predominant intercellular communication networks shaping the tumor microenvironment.

(B) Network representation of intercellular communication, where nodes denote distinct cell subpopulations and edge thickness corresponds to the strength of inferred signaling interactions.

(C) Dot plot illustrating ligand-receptor interactions across all cell types in colorectal cancer, providing an overview of intercellular communication within the TME.

(D) Bar plot quantifying the number of cell types engaging in communication with the BTa groups, demonstrating its interaction landscape within the tumor microenvironment.

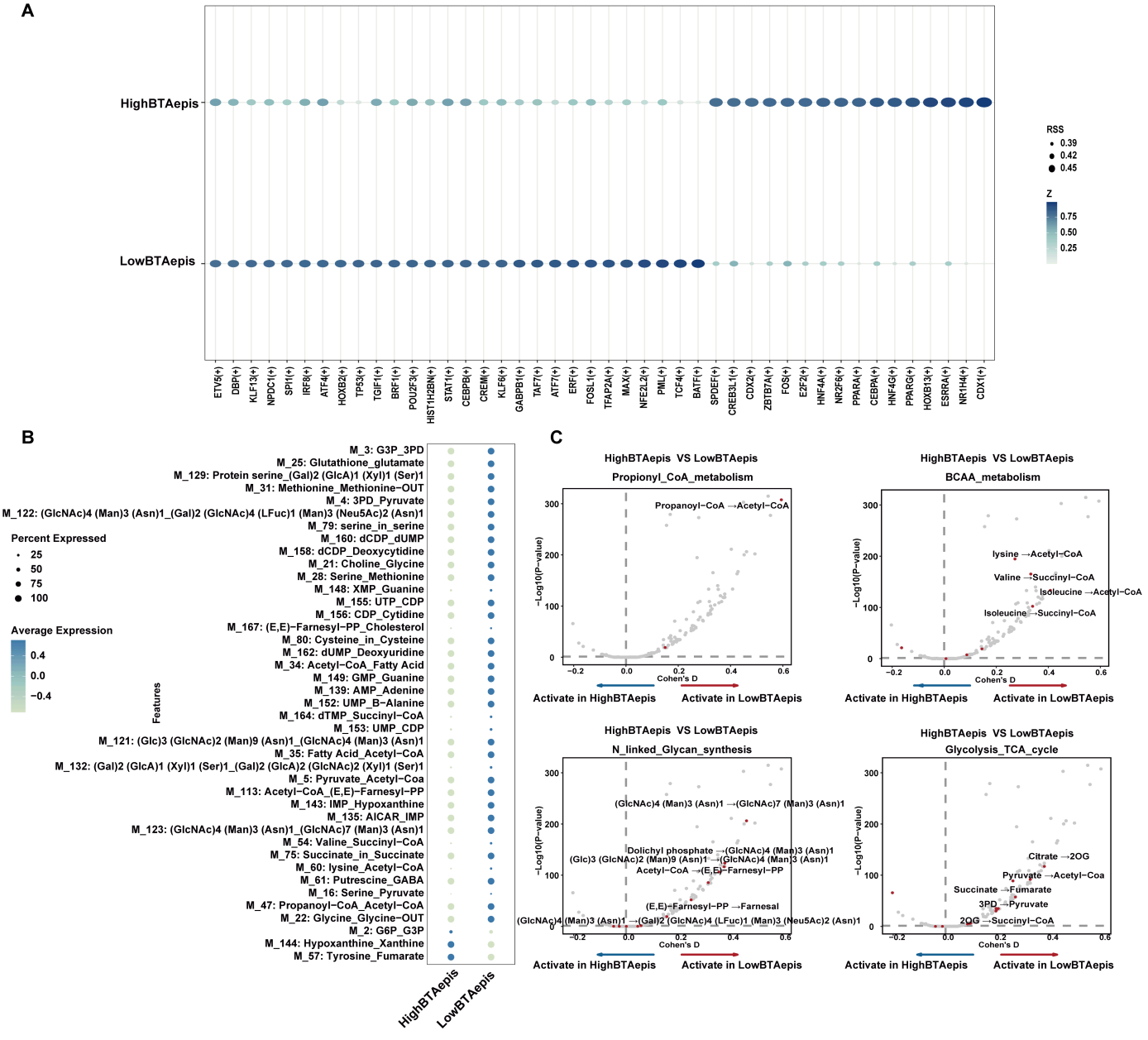


Figure 6

Transcriptional and Metabolic Dynamics in Tumor Epithelial Cells With Distinct Butanoate Scores

(A) Dot plot illustrating transcription factor activity across tumor epithelial cells with high and low butanoate metabolism scores, inferred through SCENIC analysis. The intensity of color represents the average regulon activity, providing insights into key regulatory networks shaping cellular states.

(B) Dot plot depicting metabolite abundance across tumor epithelial cells with distinct butanoate metabolism scores. The analysis incorporated 70 significantly altered metabolic modules and their associated metabolites, highlighting metabolic variations.

(C) Volcano plots comparing metabolic activity between tumor epithelial cells with high and low butanoate metabolism scores. Key metabolites within Propionyl-CoA metabolism, BCAA metabolism, N-linked glycan synthesis, and Glycolysis-TCA cycle are highlighted to reveal metabolic distinctions.

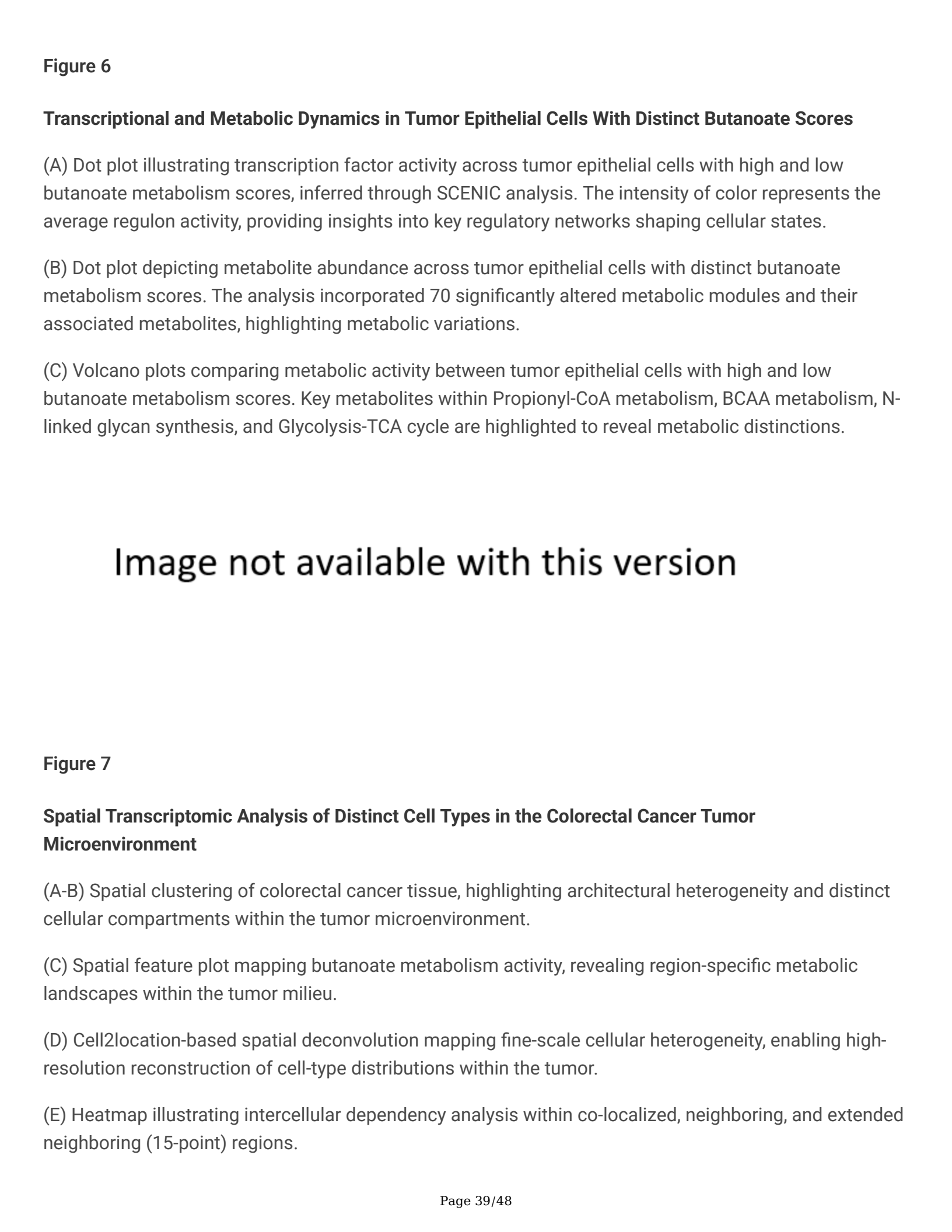


Figure 7

Spatial Transcriptomic Analysis of Distinct Cell Types in the Colorectal Cancer Tumor Microenvironment

(A-B) Spatial clustering of colorectal cancer tissue, highlighting architectural heterogeneity and distinct cellular compartments within the tumor microenvironment.

(C) Spatial feature plot mapping butanoate metabolism activity, revealing region-specific metabolic landscapes within the tumor milieu.

(D) Cell2location-based spatial deconvolution mapping fine-scale cellular heterogeneity, enabling high-resolution reconstruction of cell-type distributions within the tumor.

(E) Heatmap illustrating intercellular dependency analysis within co-localized, neighboring, and extended neighboring (15-point) regions.

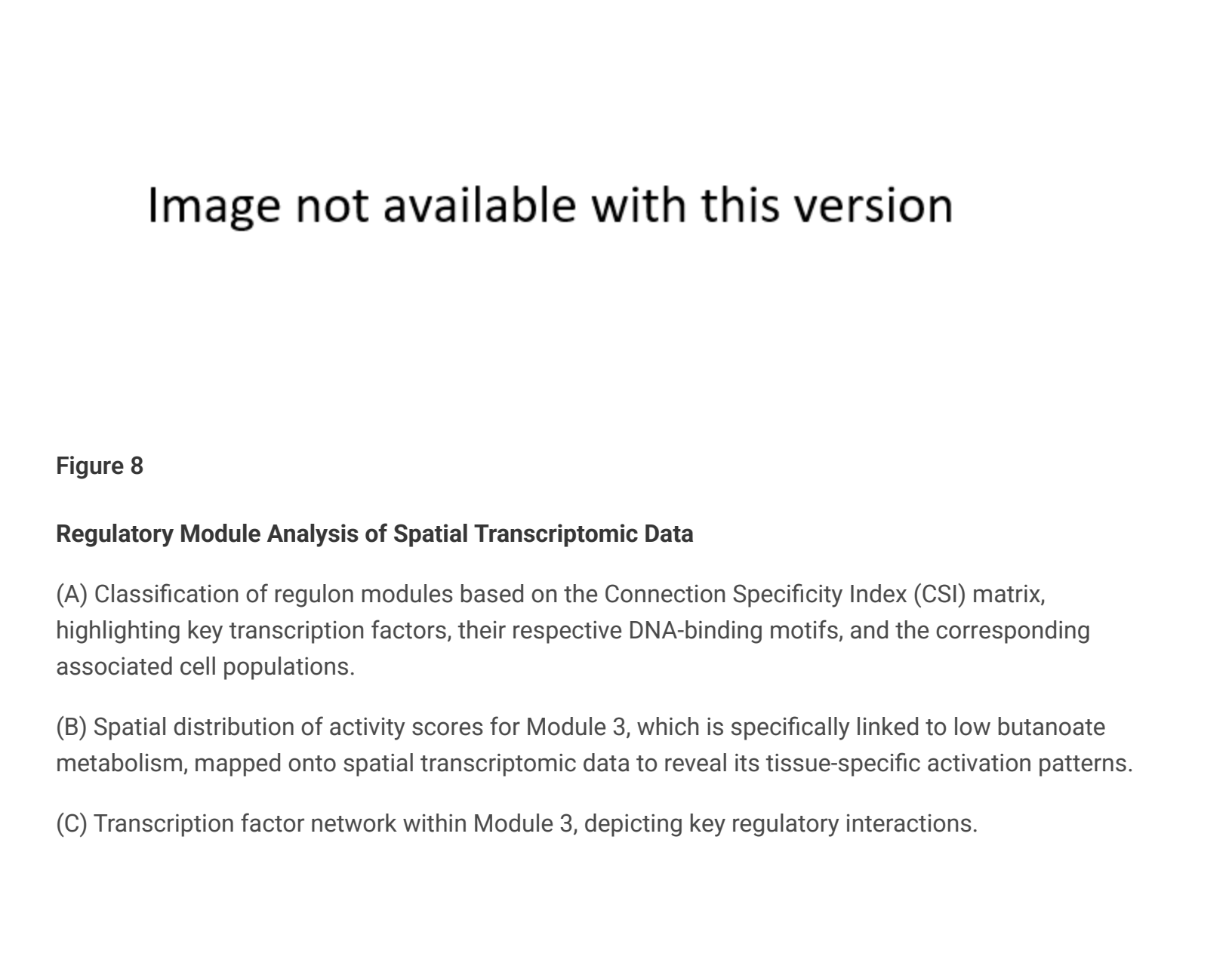


Figure 8

Regulatory Module Analysis of Spatial Transcriptomic Data

(A) Classification of regulon modules based on the Connection Specificity Index (CSI) matrix, highlighting key transcription factors, their respective DNA-binding motifs, and the corresponding associated cell populations.

(B) Spatial distribution of activity scores for Module 3, which is specifically linked to low butanoate metabolism, mapped onto spatial transcriptomic data to reveal its tissue-specific activation patterns.

(C) Transcription factor network within Module 3, depicting key regulatory interactions.

Figure 9

Spatial Trajectory Inference of Tumor Epithelial Cells With Distinct Butanoate Metabolism

(A-C) The pseudo-time spatial trajectory algorithm was applied to delineate the spatial evolution trajectories of tumor epithelial cells within the ST GSE225857 dataset, revealing dynamic cellular transitions and lineage progression.

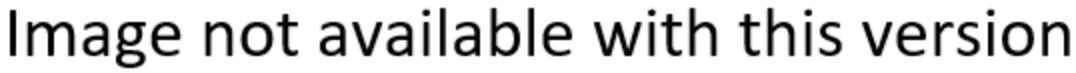


Figure 10

Selective Ligand-Receptor Coordinated Cell Interactions in Spatial Transcriptomic Data

(A) Scoring and spatial co-localization analysis of specific ligand-receptor pairs, highlighting their distribution and potential functional relevance.

(B) Analysis of interaction strength of specific ligand-receptor pairs across different cell types, revealing variations in intercellular communication patterns.

(C) Chord diagram illustrating spatial communication mediated by specific ligand-receptor pairs among different cell types, depicting key intercellular signaling relationships.

(D) Spatial trajectory-associated genes identified at ST-C1, ST-C2 and ST-C3, with negatively correlated genes highlighted in yellow and positively correlated genes in blue, elucidating spatially regulated gene expression dynamics.

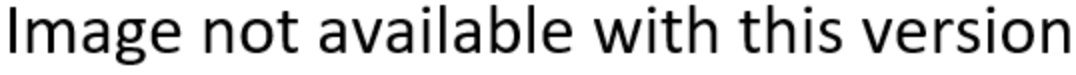


Figure 11

Development and Validation of a Consensus Signatures Using Machine Learning-Based Integration

(A) The Boruta algorithm pinpointed seven key marker genes associated with butanoate metabolism. Features validated with high confidence are depicted in deep blue, while shadow attributes are represented in other colors. The accompanying boxplots illustrate variations in the concordance index (C-index) values, offering insights into feature robustness and predictive efficacy.

(B) Comparison of the C-index of 101 machine-learning models across the TCGA cohort and the two external evaluation cohorts, GSE17536 and GSE39582.

(C) Kaplan–Meier survival analysis according to the selected CoxBoost + RSF model in the TCGA, GSE17536, and GSE39582 cohorts.

(D) Time-dependent ROC curves illustrating the predictive accuracy of the risk model for OS at 1, 3, and 5 years.

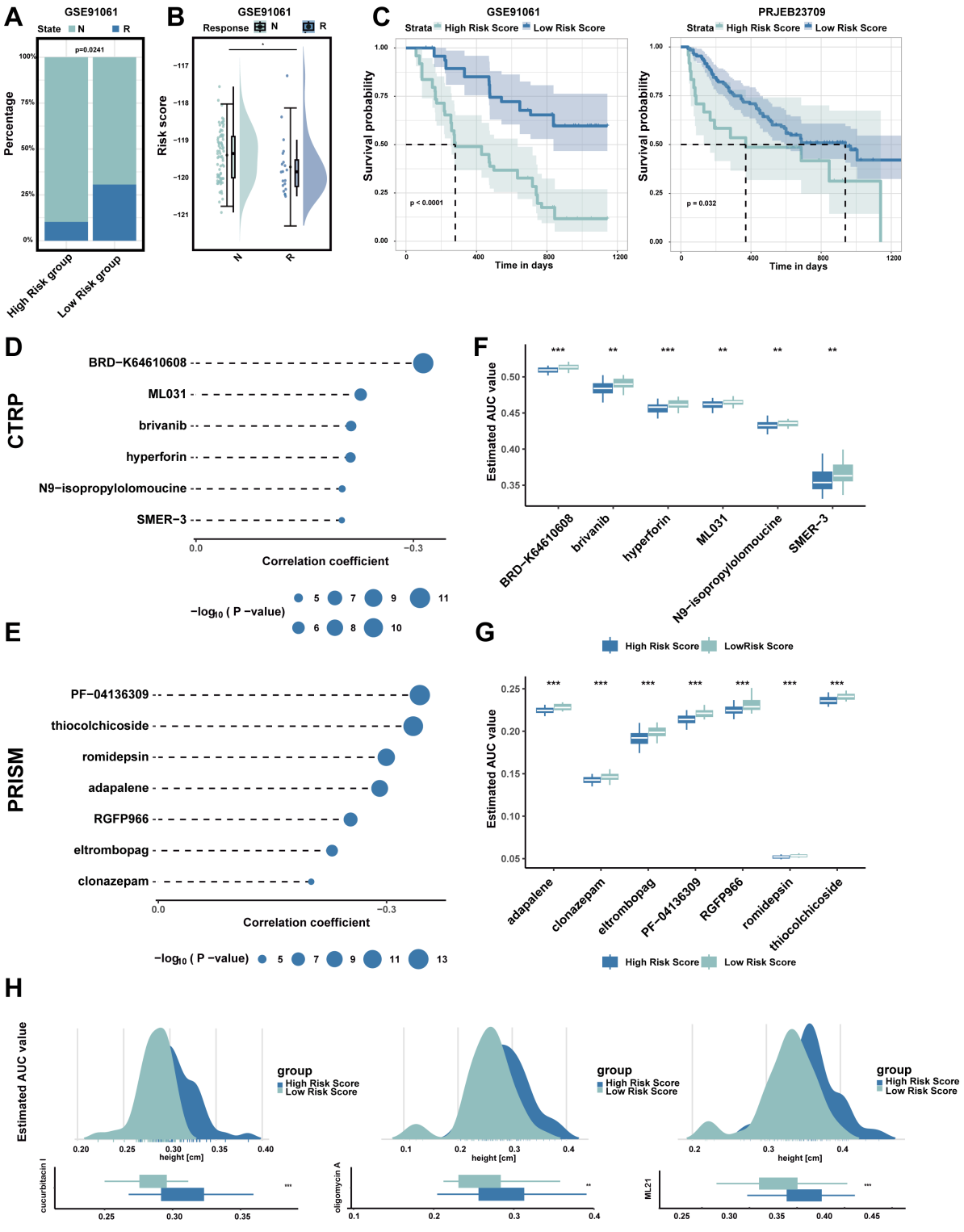


Figure 12

Predicting Immunotherapy Response and Drug Sensitivity Based on Risk Score Model

(A) Bar plot comparing immunotherapy response rates between high-risk and low-risk groups in the GSE91061 cohort.

(B) Raincloud plot illustrating differences in risk scores between responders and non-responders in the GSE91061 cohort.

(C) Kaplan-Meier survival curves evaluating the prognostic significance of the risk score model in patients from the GSE91061 and PRJEB23709 cohorts.

(D-E) Spearman's correlation analysis and differential drug response analysis for compounds derived from the CTRP dataset (D) and the PRISM dataset (E), assessing the relationship between the risk score model and drug response variability.

(F-G) Drug sensitivity analysis for CTRP-derived compounds (F) and PRISM-derived compounds (G), demonstrating differential therapeutic responses between high-risk and low-risk groups. Lower y-axis values indicate greater drug sensitivity (*p < 0.05; **p < 0.01; ***p < 0.001).

(H) Estimated drug response differences between high-risk and low-risk groups for Cucurbitacin I, Oligomycin A, and ML21, highlighting potential therapeutic vulnerabilities.

Figure 13

Sodium butyrate (NaB) suppresses the proliferation, migration, and invasion of HCT116 cells.

(A) Cell proliferation assessed by CCK-8 assay was progressively inhibited with increasing doses of NaB (2, 4, 8 mM) and incubation time (24, 48, and 72 h).
(B) Quantitative analysis of the wound healing assay showed that treatment of HCT116 cells with NaB for 48 h dose-dependently reduced the relative migration area.
(C) Representative images of the wound healing assay.
(D) Representative images of the Transwell invasion assay.
(E) Quantitative analysis of invaded cells demonstrates that treatment of HCT116 cells with NaB for 48 h dose-dependently inhibited HCT116 cell invasion.

(F) Representative Western blot images and quantitative analysis showing that NaB treatment for 48 h dose-dependently reduced HOXB8 protein expression in HCT116 cells, with protein levels normalized to

GAPDH. Data are presented as the mean ± SEM from at least three independent experiments. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001. Full-length blots/gels are presented in additional file 2, Supplementary Figure S9

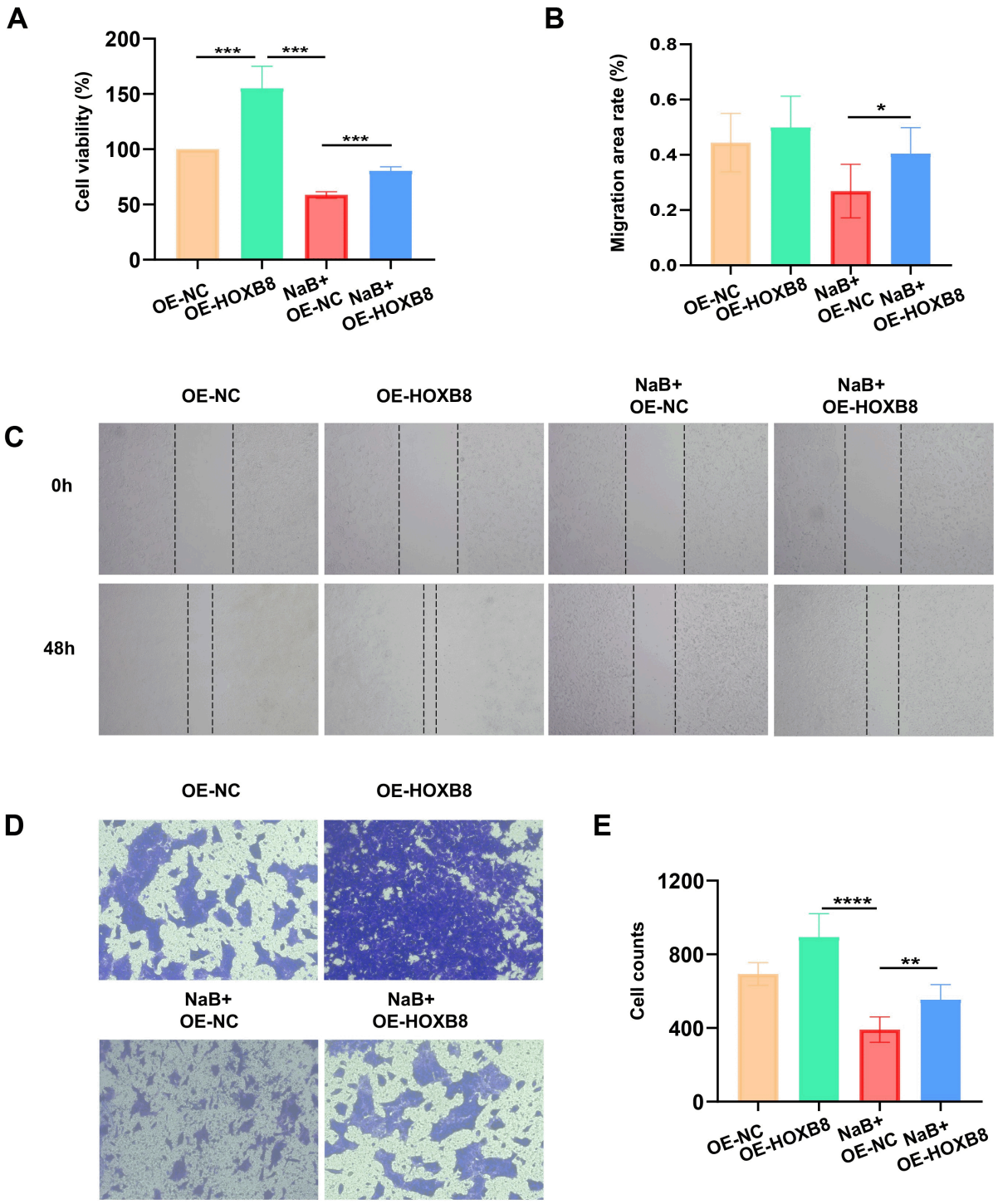


Figure 14

HOXB8 mediates the inhibitory effects of sodium butyrate on HCT116 malignant phenotypes. (A) Proliferation of HCT116 cells under different conditions: Control (untreated), OE-*HOXB8* (*HOXB8*-

overexpression), NaB (sodium butyrate treatment), and NaB + OE-*HOXB8*(*HOXB8*-overexpression with sodium butyrate treatment).
(B) Quantitative analysis of cell migration.
(C) Representative images of cell migration under different conditions.
(D) Representative images of cell invasion under different conditions.
(E) Quantitative analysis of cell invasion.
All data are presented as the mean ± SEM (n ≥ 3). One-way ANOVA was used for data analysis. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001.

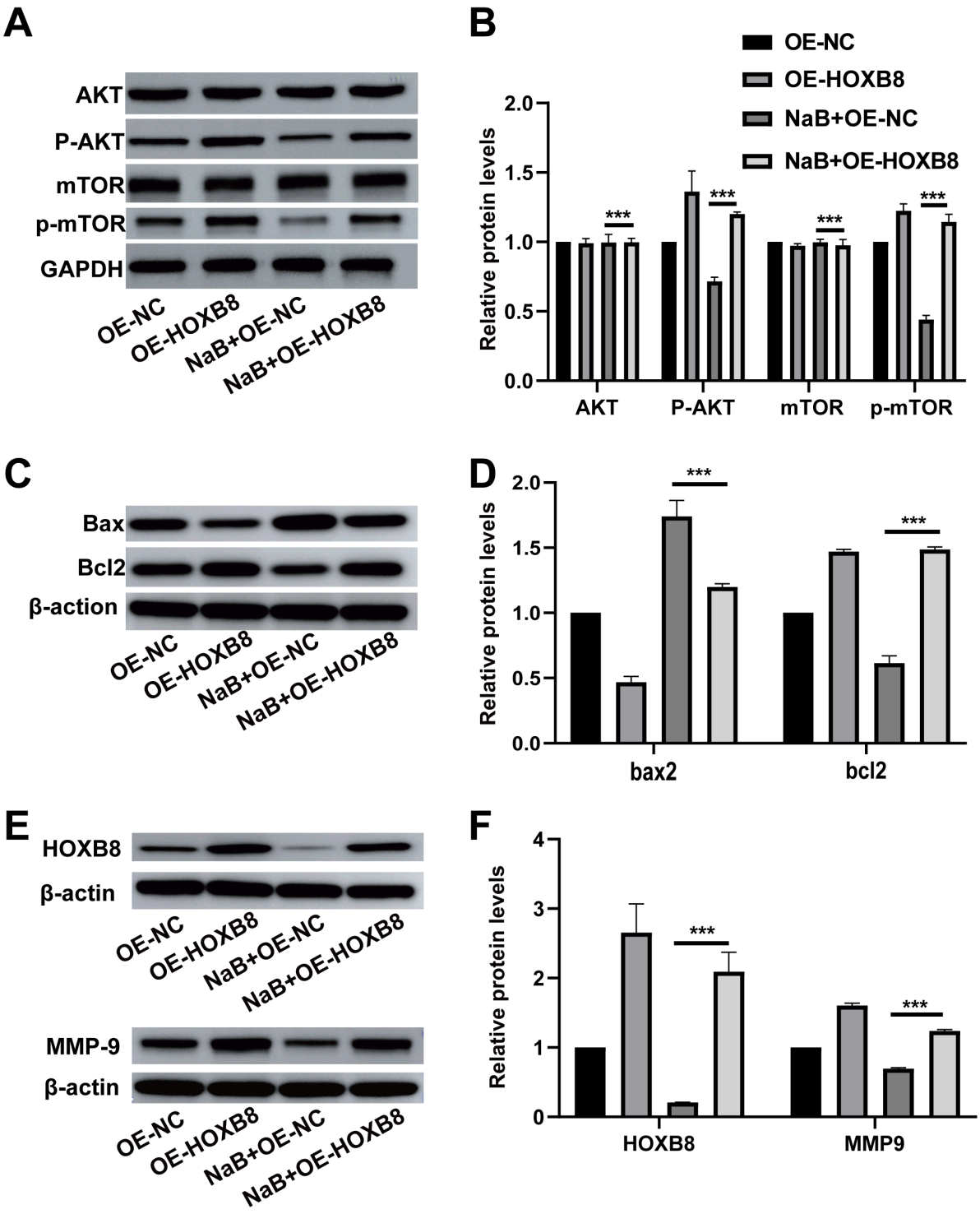


Figure 15

NaB inhibits AKT/mTOR signaling in HCT116 cells in association with *HOXB8* suppression

(A) Western blot analysis of phosphorylated AKT (p-AKT Ser473), total AKT, phosphorylated mTOR (p-mTOR), total mTOR.

(B) Quantitative analysis of the gray value ratio of p-AKT/AKT and p-mTOR/mTOR,

(C) Western blot analysis of apoptosis-related proteins (Bcl-2/Bax),

(D) Quantitative analysis of the gray value ratio of apoptosis-related proteins (Bcl-2/Bax),

(E) Western blot analysis of HOXB8 and metastasis-related protein MMP9.

(F) Quantitative analysis of the gray value ratio of HOXB8 and metastasis-related proteins MMP9.

Data were analyzed using one-way ANOVA. *p < 0.05, **p < 0.01, ***p < 0.001. Full-length blots/gels are presented in additional file 2, Supplementary Figure S9.

Supplementary Files

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

- Additionalfile1.zip
- Additionalfile2.zip
- SupplementaryInformation.docx