



OPEN ACCESS

EDITED BY

Zodwa Dlamini,
Pan African Cancer Research Institute
(PACRI), South Africa

REVIEWED BY

Teresa Gledhill,
Central University of Venezuela, Venezuela
Jidong Zhang,
Zunyi Medical University, China

*CORRESPONDENCE

Haibo Lu

✉ luhaibo@hrbmu.edu.cn

Jie Lian

✉ lianjie@hrbmu.edu.cn

Shengjun Ji

✉ drshengjunji@jmu.edu.cn

[†]These authors have contributed
equally to this work and share
first authorship

[‡]These authors have contributed equally to
this work

RECEIVED 29 November 2024

ACCEPTED 10 February 2025

PUBLISHED 04 March 2025

CITATION

Zhu J, Xu B, Wu Z, Yu Z, Ji S, Lian J and Lu H
(2025) Integrative analysis of semaphorins
family genes in colorectal cancer:
implications for prognosis and
immunotherapy.
Front. Immunol. 16:1536545.
doi: 10.3389/fimmu.2025.1536545

COPYRIGHT

© 2025 Zhu, Xu, Wu, Yu, Ji, Lian and Lu. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Integrative analysis of semaphorins family genes in colorectal cancer: implications for prognosis and immunotherapy

Jiahao Zhu^{1†}, Benjie Xu^{1†}, Zhixing Wu^{2,3†}, Zhiwei Yu⁴,
Shengjun Ji^{5*†}, Jie Lian^{1*†} and Haibo Lu^{1*†}

¹Department of Outpatient Chemotherapy, Harbin Medical University Cancer Hospital, Harbin, Heilongjiang, China, ²Department of Computer Science, University of Liverpool, Liverpool, United Kingdom, ³Department of Biological Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, China, ⁴Department of Colorectal Surgery, Harbin Medical University Cancer Hospital, Harbin, Heilongjiang, China, ⁵Department of Radiotherapy and Oncology, Suzhou Municipal Hospital, The Affiliated Suzhou Hospital of Nanjing Medical University, Gusu School, Nanjing Medical University, Suzhou, Jiangsu, China

Background: Semaphorins (SEMs), originally identified as axon guidance factors, have been found to play crucial roles in tumor growth, invasiveness, neoangiogenesis, and the modulation of immune responses. However, the prognostic value of SEMA-related genes in colorectal cancer (CRC) remains unclear.

Methods: We applied a novel machine learning framework that incorporated 10 machine learning algorithms and their 101 combinations to construct a SEMA-related score (SRS). Multi-omics analysis was performed, including single-cell RNA sequencing (scRNA-seq), and spatial transcriptome (ST) to gain a more comprehensive understanding of the SRS. A series of cell experiments were conducted to prove the impact of key genes on CRC biological behavior.

Result: A consensus SRS was finally constructed based on a 101-combination machine learning computational framework, demonstrating outstanding performance in predicting overall survival. Moreover, distinct biological functions, mutation burden, immune cell infiltration, and immunotherapy response were observed between the high- and low-SRS groups. scRNA-seq and ST demonstrated unique cellular heterogeneity in CRC. We observed that SRS-high and SRS-low malignant epithelial cells exhibit different biological characteristics. High SRS malignant epithelial cells interact with myeloid and endothelial cells via SPP1 and COL4A2-ITGAV-ITGB8 pathways, respectively. Low SRS cells engage with myeloid and endothelial cells through MIF and JAG1-NOTCH4 pathways. Additionally, knocking down SEMA4C significantly inhibits the proliferation and invasion of CRC cells, while promoting apoptosis *in vitro*.

Conclusion: SRS could serve as an effective tool to predict survival and identify potential patients benefiting from immunotherapy in CRC. It also reveals tumor heterogeneity and provides valuable biological insights in CRC.

KEYWORDS

semaphorins, colorectal cancer, immunotherapy, single-cell sequencing, spatial transcriptome sequencing

Introduction

Colorectal cancer (CRC) is the third most common cancer globally, with 1.93 million cases reported in 2020. Additionally, it ranks as the second leading cause of global cancer-related deaths, accounting for around 916,000 fatalities that year (1). Despite reduced mortality rates due to screening, about 25% of patients are diagnosed with metastatic disease at initial diagnosis (2). Despite advancements in CRC patient survival through combined surgery, radiotherapy, and chemotherapy, challenges persist with disease recurrence and low survival rates. Recent advancements in immunotherapy, particularly with immune checkpoint inhibitors (ICIs), have demonstrated promising outcomes in the treatment of various cancers (3, 4). ICIs have been integrated into the standard treatment regimen for CRC, but only a limited subset of patients have experienced substantial and lasting benefits. Hence, it is crucial to discover biomarkers or signatures that could reliably predict treatment efficacy in CRC patients.

Semaphorins (SEMA3s), initially identified as axon guidance factors, are membrane-bound or secreted proteins involved in cell communication (5). Growing evidence indicates that SEMA3s expression is dysregulated in various cancers, contributing to angiogenesis (6), metastasis (7), and immune escape (8). They affect tumor progression by altering immune interactions between tumor cells and infiltrating immune cells within the tumor microenvironment (TME). This study focuses on SEMA3 to SEMA7, the semaphorins expressed in humans. Recent studies indicate that SEMA3s may function as attractants, influencing the recruitment of immune cells such as macrophages, natural killer cells, dendritic cells, and cytotoxic T lymphocytes to the TME (9). For instance, SEMA3A, SEMA3C, and SEMA3D promote tumor progression by attracting tumor-associated macrophages (10). SEMA3A, found on dendritic and B cells, facilitates T cell priming and differentiation of T helper cells in CD4⁺ T cells (11). It is associated with increased CD8⁺ T cell activation, playing a key role in IL-33-mediated anti-tumor immune responses (12).

This study utilized 10 machine learning algorithms to develop a consensus SEMA3s-related score (SRS) for the 20-member SEMA3s family in CRC. We also analyzed the differences in biological functions, mutation burden, immune cell infiltration, and immunotherapy response between the SRS-high and SRS-low groups. Single-cell RNA sequencing (scRNA-seq) and spatial transcriptomics (ST) analyses were performed to examine cellular heterogeneity in CRC tissue and to investigate the unique biological characteristics of SRS-high and SRS-low malignant epithelial cells. This study highlights the regulatory potential of SEMA3s, providing a theoretical basis for future targeted and immunotherapeutic approaches.

Materials and methods

Overview design of the study

Figure 1 shows the workflow of this study.

Landscape of SEMA family expression and mutagenesis in pan-cancer

Transcriptome data for 27 solid tumors were sourced from The Cancer Genome Atlas (TCGA) using the R package “TCGAbiolinks” (13). Supplementary Table S1 listed the 27 solid tumors used in this study. Transcriptome data from normal tissues were obtained from the Genotype-Tissue Expression (GTEx) project. The somatic mutation profiles from Genomic Data Commons (GDC) were compiled into a mutation annotation format (MAF) file. The trinucleotide W (TCW, W = A or T) mutation and tumor mutation burden (TMB) analysis were used the R package “maftools” (14).

Prognostic signature construction based on integrative machine learning in CRC

Twenty SEMA genes were selected to develop the prognostic signature. To ensure high accuracy and generalizability of the SRS, we employed an integration of 10 machine learning algorithms (Supplementary Table S2). The TCGA-CRC cohort, integrating TCGA-COAD and TCGA-READ datasets, was utilized for initial model training, with GSE17537 (15) and GSE39582 (16) cohorts employed for validation. To improve comparability across different cohorts, we applied Z-score normalization to all data in advance.

The development pipeline for the SRS involved constructing models using 101 combinations of these 10 machine learning algorithms to identify the optimal predictive SRS with the best concordance index (C-index) (17). The TCGA-CRC cohort served as the training set for model development. After establishing the models on the training set, they were tested on the validation cohorts (GSE17537 and GSE39582). The model with the highest average C-index, calculated across the validation datasets, was deemed optimal.

Comprehensive analysis of immune-omics molecular characterization and immunotherapy response based on SRS

The Immuno-Oncology Biological Research (IOBR) R package was utilized to gather diverse signatures associated with TME cell types, immunotherapy responses, immune suppression, and immune exclusion (18). We computed the enrichment score for each sample to thoroughly examine the immunological differences between patients with high and low SRS. We compared the distribution of tumor mutational burden (TMB) and tumor neoantigen burden (TNB) across the two groups. We utilized the tracking tumor immunophenotype (TIP) and tumor immune dysfunction and exclusion (TIDE) algorithms to explore biological mechanisms linked to SRS and assess the response to immunotherapy (19, 20). Predictive value of SRS was further verified in GSE126044 (21), GSE135222 (22), GSE91061 (23) and IMvigor210 cohort (24).

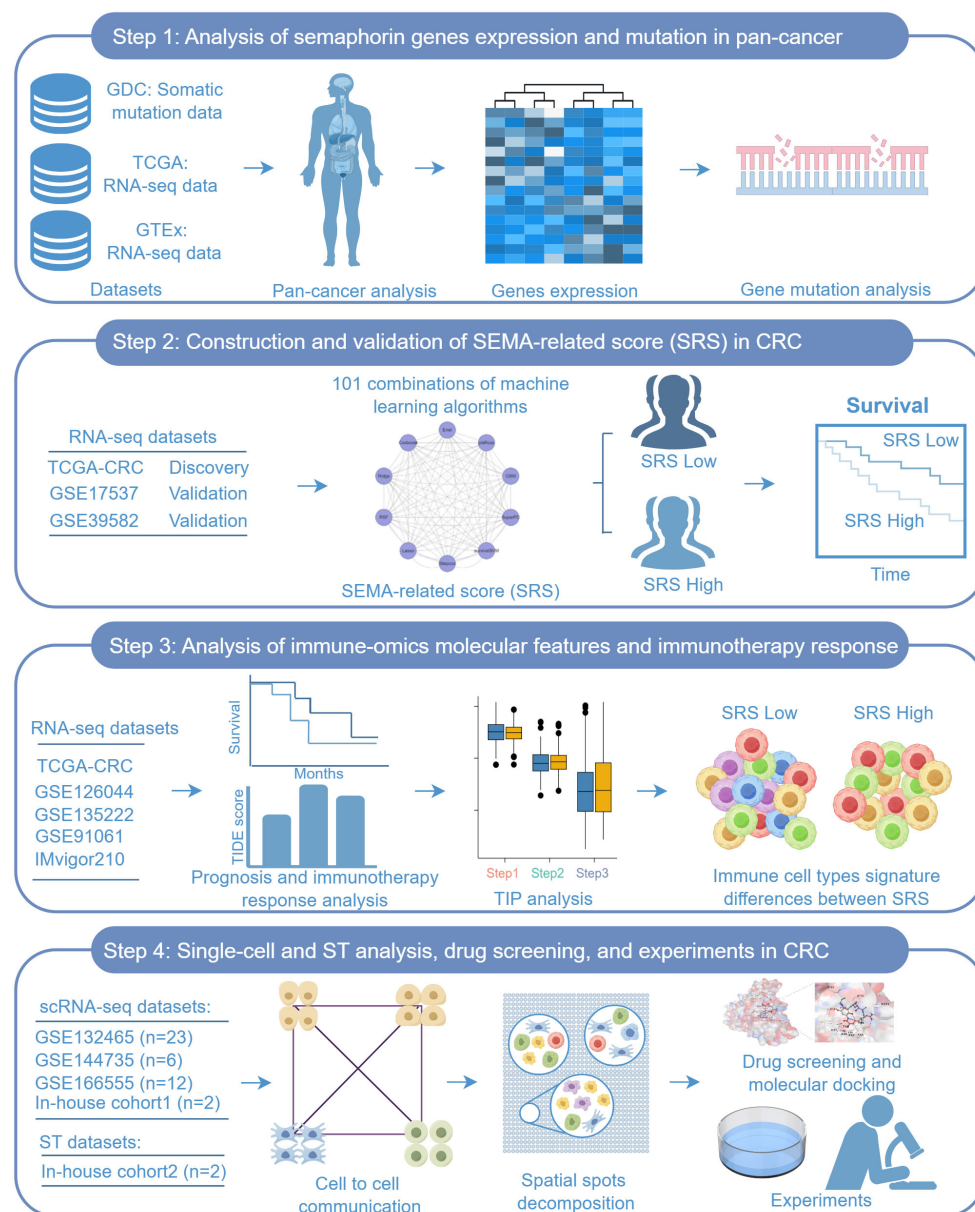


FIGURE 1
Workflow of the study.

scRNA-seq data collecting and pre-processing

To enhance the comprehensiveness of our study, we integrated our previous scRNA-seq cohort (In-house cohort1), which includes two rectal cancer tissue samples (25), with publicly available scRNA-seq datasets (GSE132465, GSE144735, and GSE166555) (26, 27). A total of 43 tumor tissue samples were included. We utilized Seurat to import the count data and create a Seurat object. After scaling the data as previously described (28), we selected 2000 highly variable genes and merged all single-cell data. Principal component analysis (PCA) was conducted using 50 principal components. To address batch effects present in both our study

and public data, we applied the Harmony package. The merged scRNA-seq data underwent clustering analysis at a resolution of 0.2, and was visualized using UMAP with 50 principal components. Clusters containing fewer than 100 cells were excluded to ensure robust results.

Celltypes annotation and copy number variation inferring

For the first step, cells were primarily annotated according to marker genes (29), including epithelial cells (EPCAM, KRT8, KRT19), endothelial cells (VWF, COL1A2, CLDN5), myeloid

cells (CD68, CD14, C1QA), T/NK cells (CD3D, CD3E, CD7), B cells (MS4A1, CD79A, BANK1), plasma cells (JCHAIN, IGHA2, MZB1), fibroblast cells (COL1A1, COL3A1, DCN), mast cells (CPA3, MS4A2, TPSAB1). In the second step, we used the “FindNeighbors” and “FindClusters” functions in Seurat to identify cell subtypes. Specific cell types in CRC were defined using gene signatures and known lineage markers as auxiliary tools. To identify malignant cells from epithelial cells, the CopyKAT algorithm (version 0.1.0) (30) was employed to estimate CNVs. Aneuploid epithelial cells are considered malignant cells, whereas diploid epithelial cells are regarded as normal cells.

scRNA-seq data scores calculating

We applied five distinct algorithms, including “AddModuleScore” “AUCell” “ssGSEA” “singscore” and “UCell”, to calculate the SRS of malignant epithelial cells. The AddModuleScore (31) algorithm calculates enrichment scores by averaging the expression values of genes in a set, providing insights into biological processes. The ssGSEA algorithm evaluates the enrichment levels of specific gene sets within individual samples or cells. AUCell assesses gene set enrichment in scRNA-seq data by ranking gene expression and calculating the area under the curve. UCell (32) enables unsupervised cell type identification without the need for prior labels. SingScore quantifies the activity of biological functions or processes within individual cells, aiding in the assessment of cell states. By integrating the scores from these five algorithms, we derived a comprehensive score. Cells with a comprehensive score greater than 1 were classified as SRS high malignant tumor cells, whereas those with a score of 1 or less were categorized as SRS low malignant tumor cells.

Cell communication analysis

To infer cell-to-cell interactions, we employed two tools: “CellChat” (33) and “CellCall” (34). The CellChat R package employs an extensive signaling interaction database that includes various receptor-ligand interactions, such as multimeric complexes, soluble agonists and antagonists, and both stimulatory and inhibitory membrane-bound coreceptors. This enables detailed mapping of the cellular communication landscape. Conversely, CellCall aggregates ligand-receptor-transcription factor axis datasets using KEGG pathway information. By combining intracellular with intercellular signals, CellCall elucidates specific pathways involved in cellular interactions. A comprehensive understanding of specific cellular communication pathways was attained through CellCall, thereby enhancing the broader interaction network identified through CellChat.

ST data processing

In this study, we utilized the 10× spatial transcriptomic sequencing method to capture spatial data from two intestinal

cancer tissue samples (In-house cohort2). Detailed information on these samples can be found in our previous study (25). To process the ST data, we followed a series of steps. First, data standardization was performed using the “SCTransform” function within the “Seurat” package. Next, dimensionality reduction was achieved by applying the “RunPCA” function to conduct PCA, which reduces the dimensionality of the data and highlights the most significant variations. Following PCA, the “RunUMAP” function was employed to carry out UMAP for clustering the data.

Re-localization for ST data

The RCTD method was employed to assign cell types from the reference scRNA-seq dataset to spatial transcriptomic data (35). Cell type-specific marker genes were identified using the ‘FindAllMarkers’ function in Seurat. This step ensured the accurate identification of cell type-specific markers. Following this, the standard RCTD analysis protocol was meticulously followed, applying it to both the reference scRNA-seq data and the Visium spatial transcriptomics data.

ST heteromorphic cell spatial network analysis

Heterotypic score algorithm was used according to the previous study (36). The kNN function in the dbscan package was used to calculate the distances between spots containing SRS high malignant tumor cells and spots containing myeloid or endothelial cells. Specifically, if all six neighboring spots around a SRS high malignant tumor cell contained myeloid or endothelial cells, these spots were selected for further analysis. The threshold for a spot to be considered as containing tumor cells was set at 0.2, while the threshold for myeloid or endothelial cells was set at 0.1. This analysis uncovered numerous associations between specific cell types, providing insight into the intricate cellular interactions within the tissue microenvironment.

Spatial map of cell dependencies

MISTy is an interpretable multi-view machine learning framework designed to analyze highly multiplexed intercellular relationships within data and thereby facilitate the understanding of cell interaction patterns and mechanisms (37). Cell-type RCTD estimates were modeled using three spatial contexts: (1) intrinsic correlations within a locale, (2) juxta estimations for nearby neighbors (maximum distance of 5), and (3) para estimations for distant neighbors (radius of 15 spots). The median standardized importance of each context were interpreted as cell-type dependencies, indicating colocalization or mutual exclusion, without implying causation.

Spatial signaling analysis

The “COMMOT” package (38) was employed to analyze signaling directions within spatial transcriptome data. COMMOT,

short for COMMunication analysis by Optimal Transport, was designed to infer cell-cell communication by concurrently evaluating multiple ligand-receptor pairs. This package facilitates the summarization and comparison of spatial signaling directions and identifies the downstream impacts of cell-cell communication on gene expression through ensemble of trees models.

Potential therapy agents for patients with high SRS

Gene Set Enrichment Analysis (GSEA) (39) was employed to compare oncogenic pathway activation between high- and low-SRS patients. Drug sensitivity data for cancer cell lines were obtained using the Cancer Therapeutics Response Portal (CTRP) and Profiling Relative Inhibition Simultaneously in Mixtures (PRISM). The AUC was utilized to assess drug sensitivity. Potential drugs for high-SRS patients were explored by referencing to previous studies (40), and the area under the AUC were used as a indicators for evaluating drug sensitivity. Agents with lower AUC in the high-SRS group ($\log_{FC} > 0.2$) and those with negative correlation coefficients ($r < -0.15$) were selected as the final potential therapeutic agents. Additionally, molecular docking was also performed. Protein structures were obtained from the Protein Data Bank, and small molecule information was retrieved from PubChem. Molecular docking was carried out using the online tool CB-Dock2.

Cell culture and siRNA transfections

HT29 and SW480 cells, obtained from ATCC, were maintained in Dulbecco's Modified Eagle Medium (Gibco, USA) with 10% fetal bovine serum (CellMax, China) and 1% penicillin-streptomycin. The cells were maintained at 37°C in a 5% CO₂ environment, with routine testing to ensure they were free of mycoplasma contamination. Lipofectamine 3000 (Invitrogen, USA) was used for siRNA transfection according to the manufacturer's guidelines. The siRNA, sourced from Ribio, China, is detailed in [Supplementary Table S3](#) with the specific sequences used in the study.

Western blotting

Total proteins were extracted using radio immunoprecipitation assay lysis (RIPA) buffer supplemented with phenylmethylsulfonyl fluoride (PMSF) and a Protease Inhibitor Cocktail (APEX BIO, USA). Protein concentrations were measured using a BCA kit from Beyotime, China. Protein lysates (30 µg) were separated on a 10% SDS-PAGE gel and transferred to polyvinylidene fluoride (PVDF) membranes. Membranes were blocked using 5% skim milk in TBST and incubated with primary antibodies against SEMA4C (1:3000, Cat. no: 28402-1-AP; Proteintech) and β-actin (1:1000, Cat. no: TA09; OriGene Technologies, Inc.). Following the washing step, membranes were treated with HRP-conjugated secondary antibodies (1:10,000, Cat. no: 92632210/92632211; LI-COR

Biosciences). Protein bands were detected with an ECL kit (Beyotime, China).

CCK-8 assay and colony formation assays

To evaluate cell proliferation, CRC cells transfected with either sh-NC or si-SEMA4C lentivirus were seeded into 96-well plates. Cell proliferation was assessed with the Cell Counting Kit-8 (CCK-8, Beyotime, China) following the manufacturer's guidelines. Optical density (OD) readings at 450 nm were recorded to determine the cell proliferation rate.

Transfected CRC cell lines were plated at 1,000 cells per well in 6-well plates and cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum for 10 days to facilitate colony formation. Colonies were then fixed with methanol and stained with 1% crystal violet (Beyotime, China). Stained colonies were imaged for analysis.

Transwell assay

Cell migration was evaluated using a Transwell assay. Cells (1×10^5 cells/mL) were suspended in serum-free DMEM and seeded into the upper chamber of an 8-µm pore insert (Corning, NY, USA), while the lower chamber contained DMEM with 10% FBS. After 24 hours of incubation, cotton swabs were used to remove non-migrated cells, and migrated cells were imaged at 200× magnification in five random fields. Migration was measured as the average cell count per field, expressed with the standard deviation. For the invasion assay, the Transwell membrane was pre-coated with 1 mg/mL Matrigel (BD Biosciences, NJ, USA) before following the same protocol as the migration assay.

Cell apoptosis assay

Cell apoptosis was evaluated with an Annexin V-FITC/PI Detection Kit. Four hours after transfection, cells were exposed to 50 µmol/L 6-OHDA and incubated for 24 hours at 37°C with 5% CO₂. Following incubation, cells were harvested, rinsed twice with sterile PBS, and labeled with Annexin V-FITC/PI. Apoptosis levels were then analyzed by flow cytometry.

Statistical analysis

Unpaired Student's t-test was used for normally distributed variables, and the Wilcoxon rank-sum test was applied for non-normally distributed variables when comparing two groups. For datasets with more than two groups, one-way ANOVA was employed for parametric data, and the Kruskal-Wallis test was utilized for non-parametric data. Two-sided Fisher's exact test was used to analyze contingency tables. The SRS threshold value was set based on the mean. Statistical analyses were conducted

using R (v4.2.1), GraphPad Prism (v9.3.1), and Python (v3.8). (ns: $P > 0.05$, *: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$, ****: $P < 0.0001$).

Results

Landscape of SEMA family expression and SEMA mutagenesis in pan-cancer

To investigate the genomic characteristics of the 20 SEMA family genes across multiple cancers, we generated a comprehensive heatmap. The heatmap illustrated expression profiles across 9,765 samples, comprising 9,398 primary tumors and 367 metastatic tumors, spanning 27 solid cancer types from TCGA (Figure 2A). In the pan-cancer analysis, SEMA4B showed the highest expression levels among all the SEMAs. Additionally, certain SEMAs showed notably high expression in specific cancer types, such as SEMA5B in KIRC. A Pearson correlation analysis of all 20 SEMA family members showed that most SEMAs were significantly positively correlated (Figure 2B). We also assessed the expression of the 20 SEMA family members by comparing their levels in primary tumors and adjacent normal tissues (ANTs) across 14 cancer types. Figure 2C illustrates significant dysregulation of SEMAs in these cancer types relative to their corresponding ANTs. SEMA3E was notably downregulated in 13 cancer types, except for uterine corpus endometrial carcinoma (UCEC), where no significant downregulation was observed. The detailed expression of SEMA3E in the 14 cancer types and their corresponding ANTs is exhibited in Figure 2D.

We calculated the TCW mutation count for each tumor sample across all cancer types, considering changes from TCW to TTW and TCW to TGW, where W represents A or T. UCEC, skin cutaneous melanoma, and colon cancer exhibited the highest TCW mutation counts (Figure 2E). Using a random forest algorithm, we identified SEMA3E as the most significant contributor to TCW mutations in pan-cancer analysis (Figure 2F). Additionally, we evaluated the associations between TMB and SEMA family members across different cancer types. The comprehensive heatmap (Figure 2G) illustrates these correlations, where the color intensity reflecting the magnitude of the correlation and point size indicating the level of statistical significance. Notably, we observed that all SEMAs, except SEMA3G, had significant correlations with TMB in thymoma.

SRS signature predicts prognosis and immunotherapy response in CRC

All 20 SEMA family members were used to develop a SRS using a machine learning-based technique within a Leave-One-Out Cross-Validation framework to fit 101 models. The TCGA-CRC cohort was utilized as the training dataset, with the GSE39582 and GSE17537 datasets employed for validation. The optimal model, which integrated Lasso regression and RSF, achieved an average C-index of 0.729 (Figure 3A). To determine the optimal model parameters, we selected the λ value that minimized the penalized likelihood deviation in the TCGA-CRC training cohort (Figure 3B,

left). For the RSF component, the error rate was assessed as a function of the classification tree (Figure 3B, middle), and out-of-bag importance values were used to evaluate gene contributions (Figure 3B, right). The SRS for each patient was determined by applying the regression coefficient-weighted expression of SEMA3E, SEMA4C, and SEMA6C within the RSF model. Patients in the training and validation cohorts were divided into SRS-high and SRS-low groups based on the median SRS value. It is noteworthy that the overall survival (OS) of patients with high SRS levels was significantly lower compared to those with low SRS levels (all P -values < 0.05 , Figures 3C–E).

The TIDE algorithm assessed patient responses to immunotherapy. The analysis revealed that SRS was also positively associated with the TIDE score (Figure 3F) and patients in the low SRS group exhibited significantly better responsiveness to immunotherapy in comparison to those in the high SRS group ($P < 0.001$; Figure 3G for Fisher's exact test results). Chi-square tests showed that the SRS-low group had a lower proportion of patients with III/IV T stage ($P = 0.038$), lower mortality ($P = 0.001$), and lower recurrence ($P = 0.012$, Figure 3H).

We calculated the TIP to explore potential biological mechanisms associated with the SRS. The SRS-low group exhibited significant differences primarily at step 4, which involves the recruitment of tumor immune-infiltrating cells (Figure 3I). The SRS-low group exhibited significantly higher proportions of T cells, dendritic cells, macrophages, and Th17 cells, whereas the SRS-high group showed significantly higher proportions of neutrophils, eosinophils, B cells, Th2 cells, and regulatory T cells (Tregs).

The validation of SRS predictive power for immunotherapy response

We further substantiated our conclusions by examining multiple immunotherapy validation cohorts. A lower SRS was linked to better OS and progression-free survival (PFS), with statistical significance observed in GSE126044 for OS ($P = 0.008$; Figure 4A) and PFS ($P = 0.041$; Figure 4B), and in GSE135222 for PFS ($P = 0.004$; Figure 4C). Furthermore, a lower SRS was generally correlated with a better response to immunotherapy (GSE91061, $P = 0.043$; Figure 4D). Post-immunotherapy patients with progressive disease had higher SRS compared to those with partial response ($P = 0.034$) and stable disease ($P = 0.007$) in IMvigor 210 cohort (Figure 4E).

Immune characteristics related to SRS

We utilized the IOBR R package to examine the tumor microenvironment in colorectal cancer. As expected, signatures linked to favorable immunotherapy outcomes were significantly enriched in the SRS-low group (Figure 4F). CRCs with low SRS levels were more likely to exhibit characteristics of hot tumors with heightened immune responses. In contrast, fibroblasts and Tregs were predominantly enriched in the SRS-high group. The SRS-high

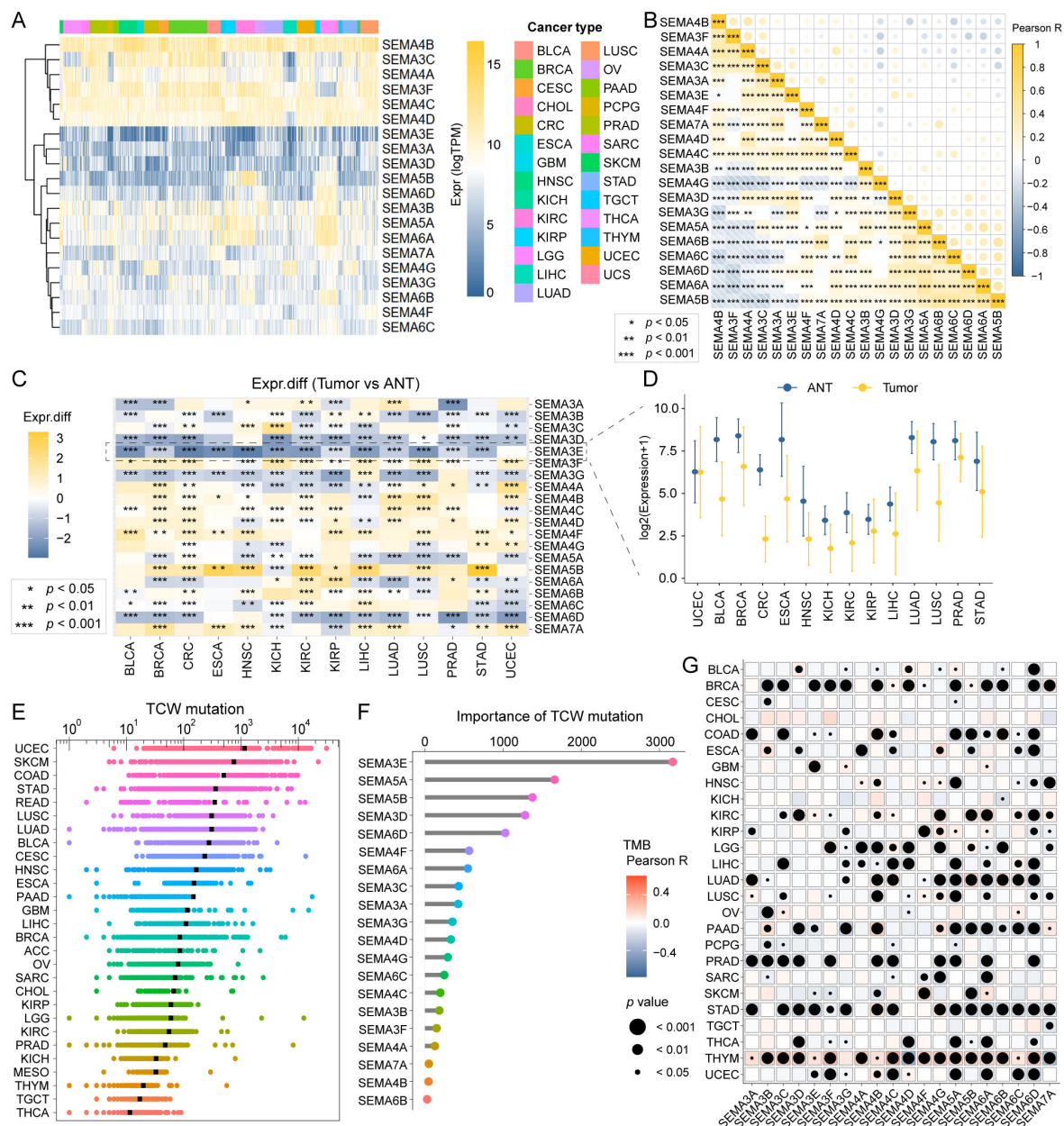


FIGURE 2

Landscape of SEMAs family expression and mutagenesis in pan-cancer. (A) A comprehensive heatmap illustrates the expression pattern of SEMAs across solid cancer types from TCGA. (B) The Pearson correlation analysis was performed among all 20 SEMAs family members in pan-cancer. (C) All 20 SEMAs family members were compared in primary tumors versus adjacent normal tissues (ANT). (D) SEMA3E was significantly downregulated in 13 of 14 cancer types compared with paracancerous tissue. (E) TCW mutation count in pan-cancer. (F) SEMA3E emerged as the most important contributor to TCW mutation in pan-cancer. (G) The correlations among TMB and SEMAs were analyzed in pan-cancer. All SEMAs, except SEMA3G, had significant correlations with TMB in THYM.

group exhibited enrichment of immunosuppressive and immune-exclusion markers, including the epithelial-mesenchymal transition (EMT) pathway, suggesting an immunosuppressive state (Figures 4G, H). These findings suggest that SRS-high CRCs are more likely to be cold tumors, which typically respond poorly to immunotherapy.

TMB and TNB are well-recognized biomarkers for assessing patient responses to immunotherapy. To investigate the differences in these biomarkers between the SRS-high and SRS-low groups, we conducted a comprehensive analysis. Our findings revealed that the SRS-low group demonstrated significantly elevated levels of TMB (P

= 0.041) and TNB (P = 0.015), indicating that this group may possess greater immunogenicity (Figures 4I, J).

Heterogeneity of malignant tumor epithelial cells in CRC

scRNA-seq data from CRC tissues were extracted from four datasets comprising 43 CRC patients and 84,589 cells (Figure 5A). Major cell clusters were annotated using marker genes, identifying

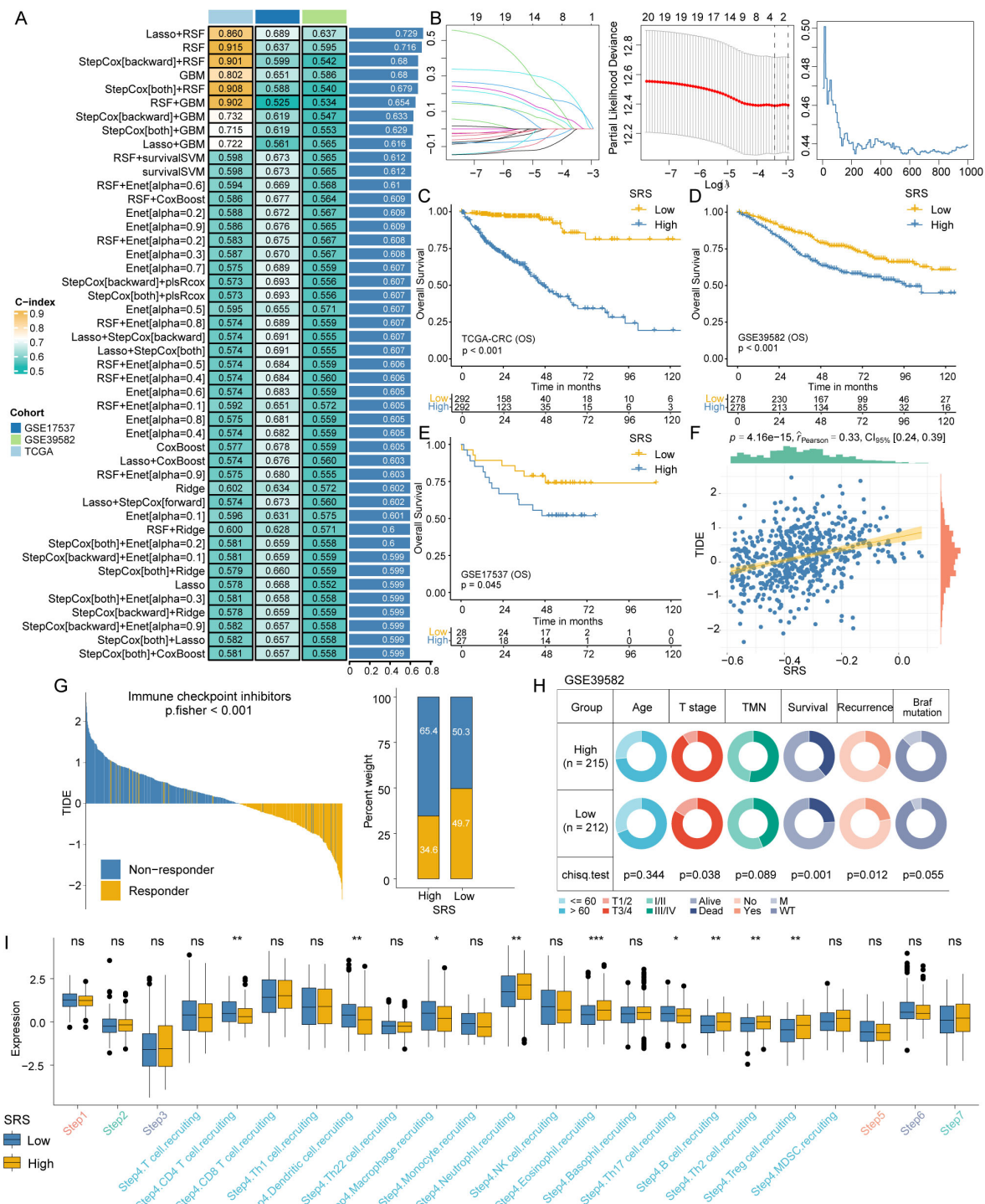


FIGURE 3
SRS construction and differences analysis between SRS-high and SRS-low groups. **(A)** A combination of 101 machine learning algorithms was generated through a comprehensive computational framework. The C-index of each model was calculated from the TCGA-CRC, GSE17537, and GSE39582 cohorts and sorted by the average C-index. **(B)** Determination of the best λ when the penalized likelihood deviation reached its minimum value in the TCGA-CRC training cohort (left); error rate in the random survival forest as a function of the classification tree (middle); and calculation of the out-of-bag importance values for the predicted genes (right). **(C–E)** Kaplan-Meier survival curves for SRS in the TCGA-CRC training cohort and the GSE17537 and GSE39582 validation cohorts, illustrating overall survival differences. **(F)** A significant positive association exists between SRS and the TIDE score. **(G)** The TIDE algorithm predicts response to immunotherapy between SRS-high and SRS-low groups. **(H)** Pie charts showing the chi-squared test results between SRS and clinical characteristics in CRC patients from the GSE39582 cohort. **(I)** Differences in the degree of activation between SRS-high and SRS-low groups at each step of TIP. *: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$. ns: not significant.

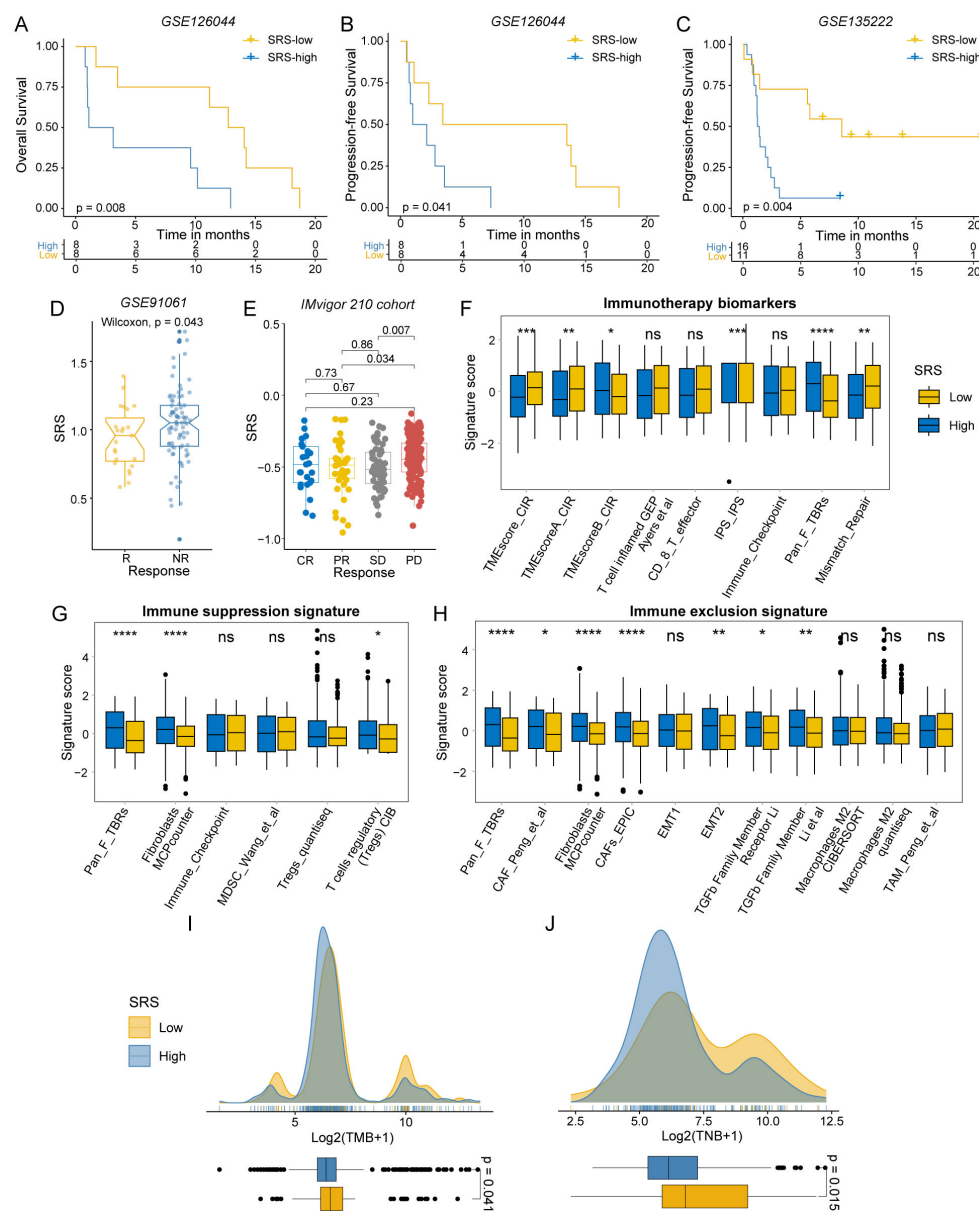


FIGURE 4

Immunotherapy response differences between SRS-high and SRS-low groups. (A, B) Survival analysis of SRS-high and SRS-low groups in GSE126044. (C) Survival analysis of SRS-high and SRS-low groups in GSE135222. (D) Distribution of SRS in different immunotherapy response groups of GSE91061. (E) Distribution of SRS in different immunotherapy response groups of IMvigor 210. (F) The distribution of immunotherapy biomarkers between SRS-high and SRS-low groups. (G) The distribution of immune suppression signatures between SRS-high and SRS-low groups. (H) The distribution of immune exclusion signatures between SRS-high and SRS-low groups. (I) The distribution of TMB between SRS-high and SRS-low groups. (J) The distribution of TNB between SRS-high and SRS-low groups. (ns: $P > 0.05$, *: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$, ****: $P < 0.0001$).

T/NK cells, B cells, plasma cells, myeloid cells, mast cells, fibroblasts, endothelial cells, and epithelial cells (Figure 5B). The proportional composition of these nine cell types across these samples is shown in Figure 5C.

Malignant tumor epithelial cells were identified by using the CopyKAT algorithm based on CNV. Among the 14,195 epithelial cells, 9,426 (66.4%) were predicted to be aneuploid, while the remaining 4,769 (33.6%) cells were predicted to be diploid (Figures 5D, E). We re-clustered 9,426 malignant epithelial cells to differentiate those with high and low SRS. Eight clusters of malignant epithelial cells were identified, with clusters 7 classified as SRS high

cluster (EpiH) and the remaining clusters as SRS low cluster (EpiL) (Figures 5F, G). Cell-to-cell interaction analyses were conducted to investigate ligand-receptor interactions between malignant epithelial cells with high or low SRS and other cell types. Results from Cellchat demonstrated varying interactive strengths among the 10 cell subtypes, with the greatest interactions between EpiH and myeloid cells, followed by endothelial cells (Figure 5H).

Then, cell-to-cell communication between malignant epithelial cells with different SRS and myeloid cells or endothelial cells was analyzed. The results demonstrate significant ligand-receptor interactions among different cell types (Figures 5I, J). In the

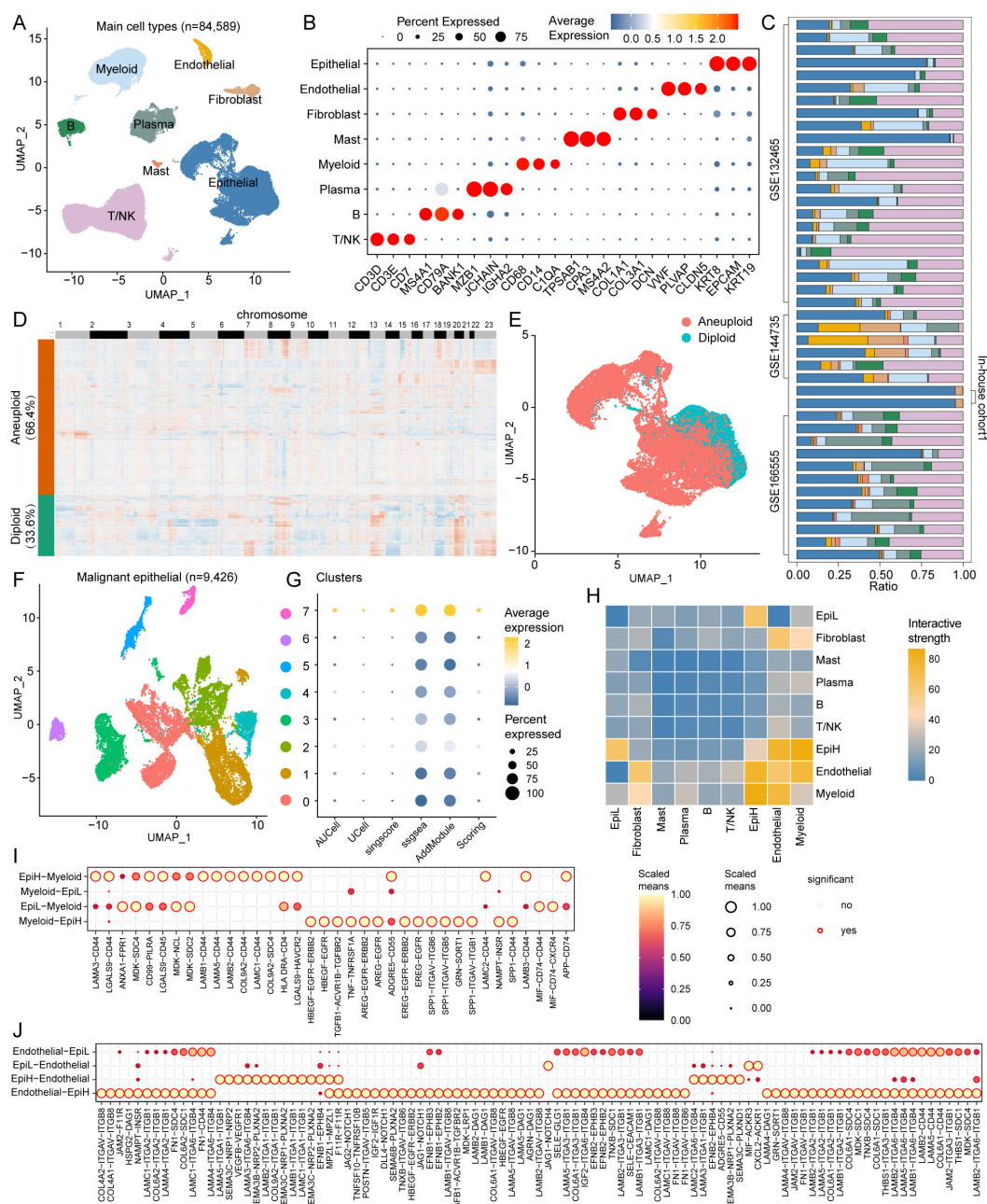


FIGURE 5

Single-cell RNA sequencing analysis in CRC. **(A)** UMAP plots of 84,589 cells from tumor tissues of 43 CRC patients in the 4 cohorts (GSE132465, GSE144735, GSE166555, and In-house cohort1), showing 8 major cell types. **(B)** Dot plots displaying marker genes for each major cell type in the 4 cohorts. **(C)** Proportional distribution of the nine major cell types across the 43 samples. **(D)** Heatmap of inferred copy number variations in epithelial cells with CopyKat algorithm, showing aneuploid (66.4%) and diploid (33.6%) epithelial cells. **(E)** UMAP plots of aneuploid and diploid epithelial cells. **(F)** UMAP plots of Reclustered 9,426 malignant epithelial cells, showing 8 cell types. **(G)** Bubble plots demonstrating the enrichment scores of SRS for each reclustered malignant epithelial cell types. **(H)** Interactive strength of cell-to-cell interactions among the 9 major cell types. **(I)** Potential ligand-receptor interactions between malignant epithelial cells and myeloid cells. **(J)** Potential ligand-receptor interactions between malignant epithelial cells and endothelial cells.

interactions between EpiH and myeloid cells, SPP1 signaling pathways and several ligands such as LAMB1-CD44 and LAMA5-CD44, among others, show notable signaling capabilities, indicating that EpiH may play a crucial role in regulating myeloid cell functions. Similarly, the communication between EpiL and myeloid cells also exhibits significant interactions, particularly through MIF signaling

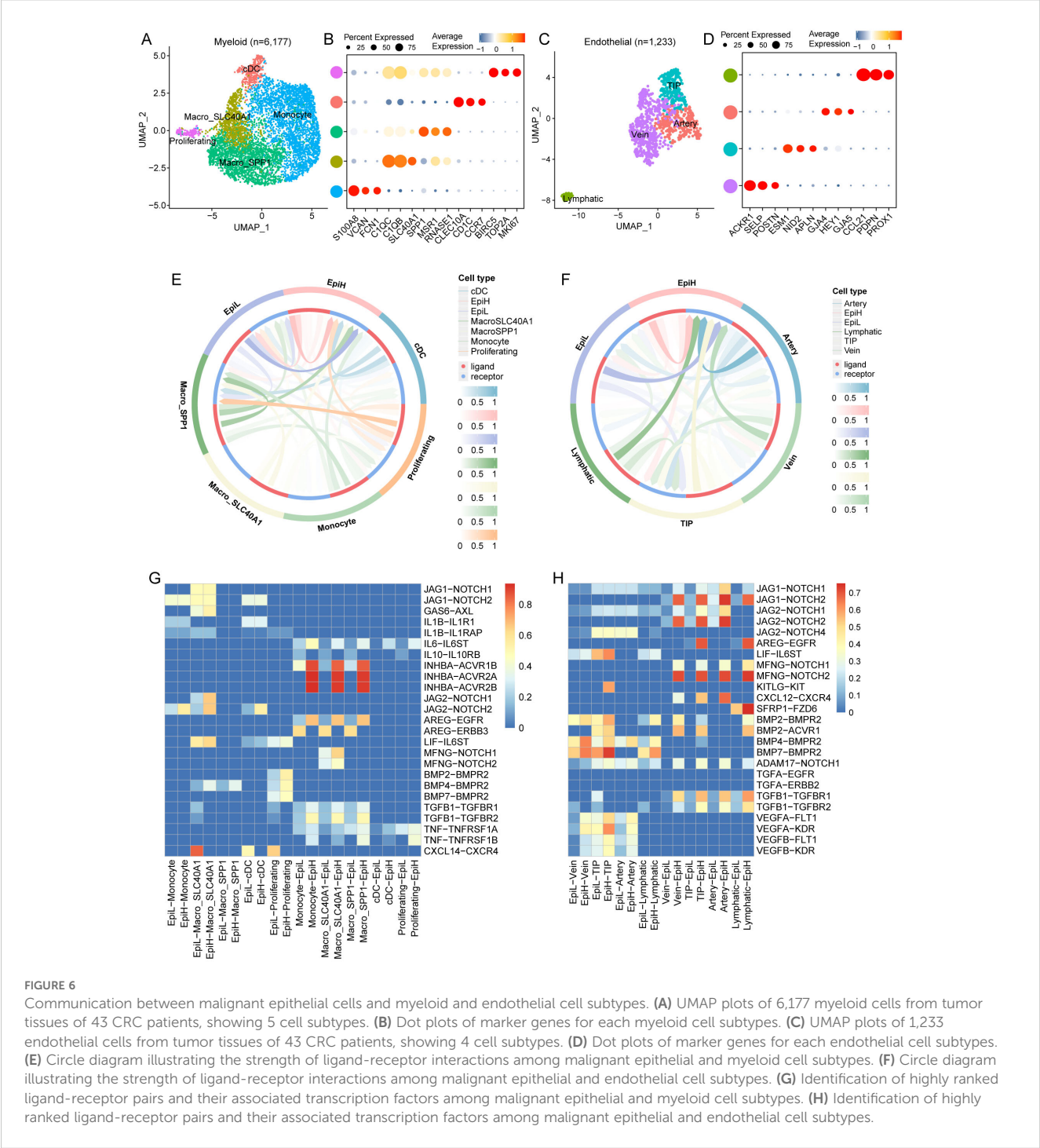
pathways. Regarding the endothelial cell analysis, EpiH demonstrates strong signaling interactions with endothelial cells, especially with the COL4A2-ITGAV-ITGB8 and LAMA5-ITGA1-ITGB1 ligand-receptor pairs. The interactions between EpiL and endothelial cells also show significance, especially with the JAG1-NOTCH4 ligand-receptor pairs.

Cell-to-cell communication between malignant epithelial cells and subtypes of myeloid or endothelial cells

In this study, myeloid cells were re-clustered into five distinct subtypes in accordance with a previous study (41): monocytes, conventional DCs (cDCs), SPP1-positive macrophages (Macro_SPP1), SLC40A1-positive macrophages (Macro_SLC40A1), and proliferating cells (Figures 6A, B; Supplementary Table S4). Endothelial cells were re-clustered into four subtypes—tip cells, vein endothelial cells, artery

endothelial cells, and lymphatic endothelial cells—according to another study (42) (Figures 6C, D; Supplementary Table S5). The CellCall analysis demonstrated significant interactions between malignant epithelial cells with different SRS and the various subtypes of myeloid and endothelial cells (Figures 6E, F).

Notably, the interaction between EpiL and monocytes displays a strong signaling potential, particularly through the JAG1-NOTCH1 and JAG2-NOTCH2 pathways. These interactions suggest that EpiL may actively engage with monocytes, influencing their differentiation and function. Moreover, the presence of AREG-



EGFR and AREG-ERBB3 interactions indicates potential regulatory mechanisms in epithelial-myeloid communication. The identified ligand-receptor pairs, such as INHBA-ACVR1B and INHBA-ACVR2A/B, further emphasize the functional cross-talk between epithelial and myeloid cells, potentially affecting cellular proliferation and signaling cascades. The heatmap highlights significant interactions between IL-10 and TNF pathways, indicating a balance of pro-inflammatory and anti-inflammatory signals between epithelial and myeloid cells (Figure 6G).

Noteworthy interactions include the JAG1-NOTCH1 and JAG2-NOTCH pathways, which are prominently involved in mediating communication between EpiH and various endothelial cell types. These interactions suggest a role for EpiH in modulating endothelial cell fate and responsiveness. Additionally, the presence of AREG-EGFR and LIF-IL6ST interactions indicates potential pathways by which EpiL may regulate endothelial cell activities, including survival and proliferation. The involvement of BMP and TGFA signaling—particularly BMP2-BMP2R and TGFA-EGFR—further highlights the importance of these pathways in endothelial function and tissue repair processes. Furthermore, the heatmap reveals interactions involving chemokines, such as CXCL12-CXCR4, which may be critical for endothelial cell migration and recruitment in response to inflammatory stimuli (Figure 6H).

Cell spatial network analysis and spatial signaling analysis with ST

To further assess the spatial organization of EpiH, EpiL, myeloid, and endothelial cells, we performed ST analysis on tumor tissue sections from our previous ST dataset, which includes two intestinal cancer tissue examples (25). We utilized the RCTD methodology to deconvolute annotated cell types from scRNA-seq data and integrate them with ST data. This approach enabled us to infer the predominant cell types present at each spatial location (Figures 7A–H).

The spatial network analysis of EpiH in relation to myeloid and endothelial cells is illustrated in the Figures 7I–L, where purple dots represent spots containing EpiH and yellow dots denote the presence of myeloid or endothelial cells. In the EpiH_Myeloid panel, the clustering of yellow spots in proximity to EpiH regions suggests potential spatial interactions that may facilitate communication and functional modulation between these cell types. This spatial arrangement likely reflects the dynamic role of EpiH in influencing myeloid cell behavior, including recruitment and activation within the local microenvironment. Similarly, the EpiH_Endothelial panel reveals a comparable pattern of endothelial cells adjacent to EpiH spots, with the presence of yellow dots near EpiH indicating potential interactions critical for angiogenesis, tissue repair, and overall vascular integrity. Based on the results from MISTy, epithelial cells with high SRS demonstrated a congruence in clustering patterns and a heightened correlation in spatial interactions with myeloid and endothelial cells within the internal space (Figures 7M, N). Considering the significance of the INHBA-ACVR1B and SEMA3E-PLXND1 signaling pathways in the interaction between EpiH and myeloid or endothelial cells, COMMOT software was utilized to infer the spatial signaling

directionality of these two pathways. Figure 7O illustrates the INHBA-ACVR1B signaling pathway, and Figure 7P depicts the SEMA3E-PLXND1 signaling pathway.

Potential therapeutic drugs screening

Notable prognostic disparities were observed between populations characterized by high and low SRS levels. GSEA indicated that pathways related to EMT, angiogenesis, hypoxia, and TNF-alpha signaling via the NF-kB pathway were markedly activated in SRS-high patients (Figure 8A). Due to the inadequate immunotherapy response in patients with high SRS, we utilized the CTRP and PRISM platforms to identify potential therapeutic drugs for these individuals. We employed Oxaliplatin, a standard CRC treatment, as a benchmark to confirm the robustness of our methodology. Our algorithm revealed that patients exhibiting low ERCC1 expression levels responded more effectively to cisplatin therapy, corroborating established clinical findings and indicating possible advantages for chemotherapy (Figure 8B). Finally, we screened two CTRP-derived agent (tanespimycin and clofarabine; Figure 8C) and four PRISM-derived agents (nutlin-3, LGX818, CGM097, and GDC-0152; Figure 8D). We further explored the molecular docking interactions of the target gene SEMA4C with tanespimycin, clofarabine, nutlin-3, and LGX818. The simulations revealed that the binding energy of SEMA4C with tanespimycin was -8.0 kcal/mol, with clofarabine was -6.5 kcal/mol, with nutlin-3 was -6.3 kcal/mol, and with LGX818 was -6.6 kcal/mol (Figures 8E–H). These results indicate that their potential as therapeutic agents for targeting SEMA4C.

Cell function experiments

Drug screening analysis revealed that SEMA4C could be targeted by multiple drugs, whereas no potential intervention drugs were identified for the other two genes. Additionally, research on the biological role of SEMA4C in colorectal cancer remains limited. Given this, SEMA4C was selected for further investigation to assess its function in CRC progression at the cellular level. To demonstrate the tumor-promoting role of SEMA4C in CRC, we developed knockdown vectors for SEMA4C (si-SEMA4C#1, si-SEMA4C#2) alongside RNAi negative control vectors (si-NC) and transfected them into HT29 and SW480 cell lines. Western blot analysis confirmed the successful transfection and a significant reduction in SEMA4C expression in both cell lines when using si-SEMA4C#1 and si-SEMA4C#2 (Figure 9A). CCK-8 and colony formation assays demonstrated significantly reduced cell viability and proliferation in the knockdown groups relative to the controls (Figures 9B, C). Transwell assays demonstrated a marked decrease in the migration and invasion abilities of HT29 and SW480 cells in the knockdown groups compared to the control group (Figures 9D–F). Additionally, flow cytometry results indicated an increase in cell apoptosis in the knockdown groups (Figure 9G). These findings suggest that SEMA4C contributes to CRC cell growth and migration, highlighting its potential role as a tumor promoter.

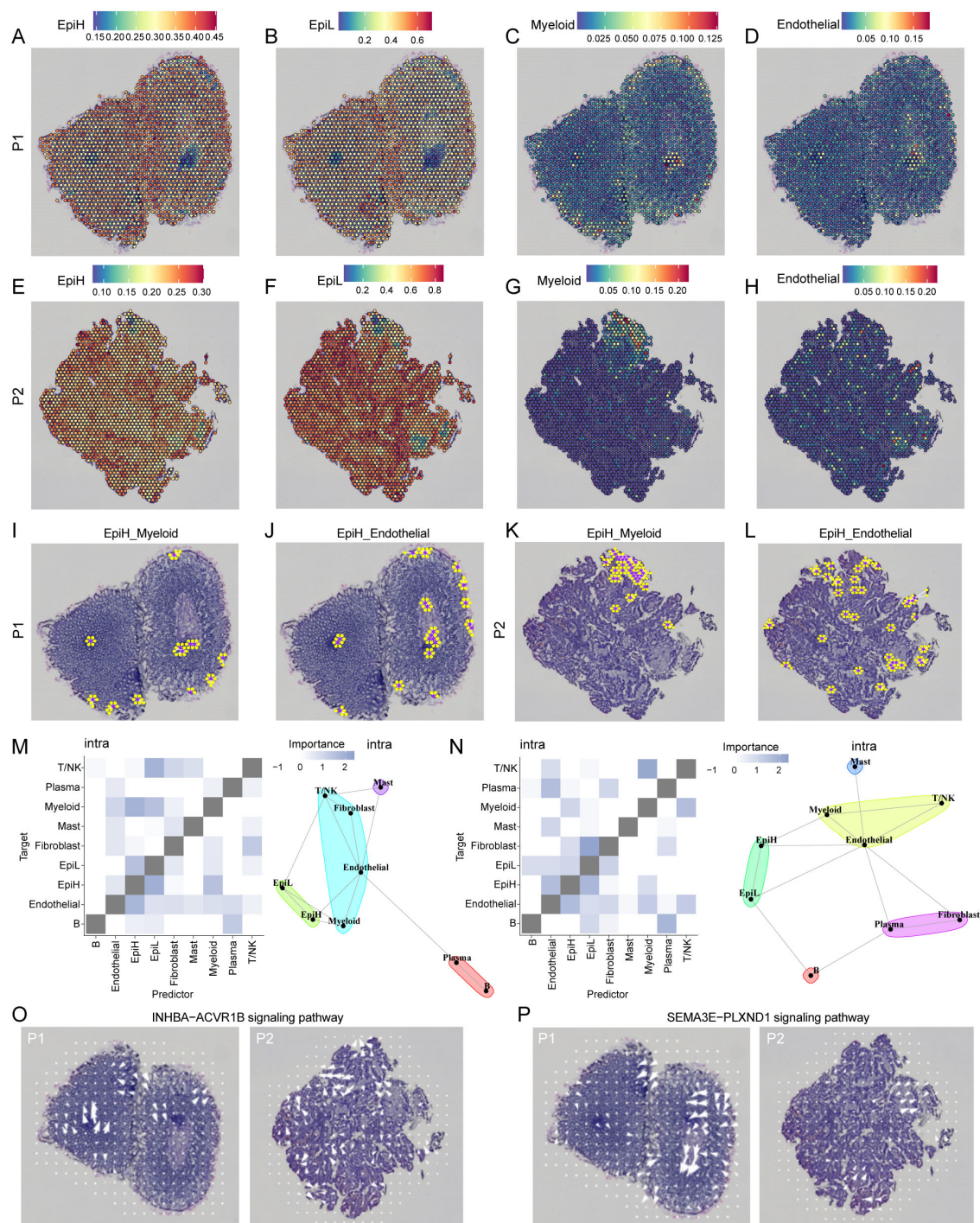


FIGURE 7
ST analysis of malignant epithelial, myeloid and endothelial cells in two CRC examples. **(A–H)** Co-localization of malignant epithelial cells with SRS-high (EpiH), SRS-low (EpiL), myeloid and endothelial cells. **(I–L)** The spatial network analysis of EpiH in relation to myeloid and endothelial cells. Purple dots represent spots containing EpiH and yellow dots denote the presence of myeloid or endothelial cells. **(M, N)** Extrapolation of spatial clustering and projection of spatial correlations among different cell types based on the MISTy algorithm. **(O)** Spatial signaling directionality analysis of the INHBA-ACVR1B pathways. **(P)** Spatial signaling directionality analysis of the SEMA3E-PLXND1 pathways.

Discussion

This study presents a novel computational framework that incorporates ten unique machine learning algorithms and their 101 potential combinations. This extensive analysis led to the

identification of the SRS, which exhibits high predictive accuracy for the prognosis of colorectal cancer CRC. Moreover, we utilized the SRS for risk stratification of CRC patients, evaluating their response to immunotherapy. Our findings suggest that CRC patients exhibiting a low SRS demonstrate enhanced survival rates

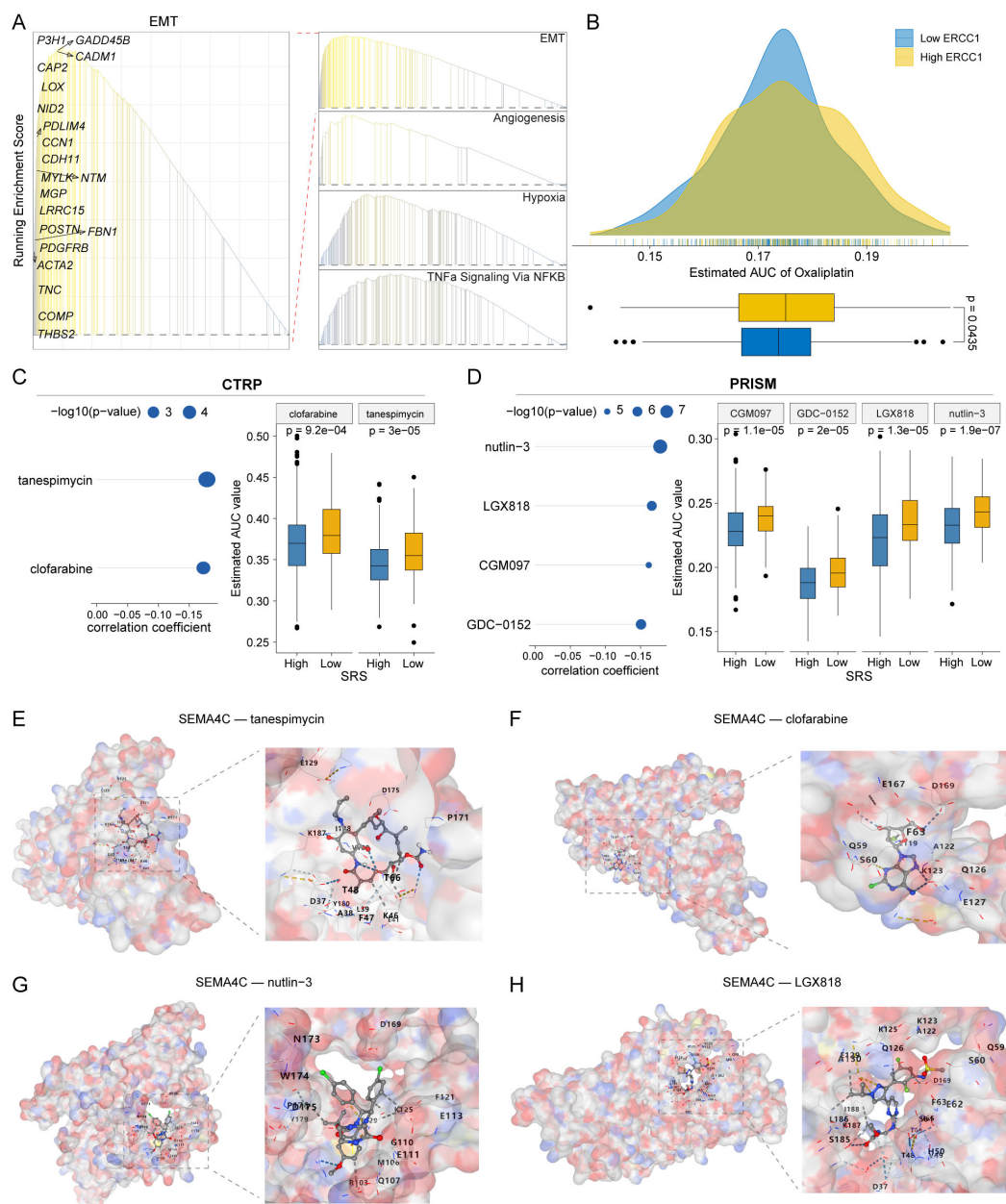


FIGURE 8

Potential agents for patients with high SRS. **(A)** Identification of significantly activated pathways in the SRS-high group using the GSEA algorithm. **(B)** Predicting oxaliplatin sensitivity to assess the computational algorithm's feasibility. **(C, D)** Analyzing correlation and differences in drug sensitivity for potential candidates from the CTRP and PRISM datasets. **(E–H)** Conducting molecular docking studies between SEMA4C and the compounds tanespimycin, clofarabine, nutlin-3, and LGX818.

and exhibit a greater likelihood of benefiting from immunotherapy. These insights offer a rational framework for the administration of immunotherapy in clinical practice, marking a substantial advancement toward more effective personalized medicine strategies. Furthermore, we identified potential pharmacological agents that may inhibit the progression of CRC to a high SRS phenotype, thereby presenting novel opportunities for CRC preventive strategies. In contrast to previous research, which primarily focused on the prognostic implications of the signature, this study incorporated bulk transcriptome analysis, scRNA-seq, and ST to attain a more comprehensive understanding of the SRS.

These findings provide robust biological evidence and explanations for the SRS, emphasizing its potential in guiding personalized medicine approaches.

Our study utilized an innovative computational framework to identify a strong prognostic signature, SRS. Kaplan-Meier curve analysis demonstrated that the SRS effectively stratifies CRC patients' risk concerning OS. Additionally, SRS is associated with more advanced CRC stages and metastasis occurrence, linked to adverse clinical outcomes. Approximately 30% of patients diagnosed with locally advanced CRC eventually progress to metastatic disease, underscoring the critical need for early

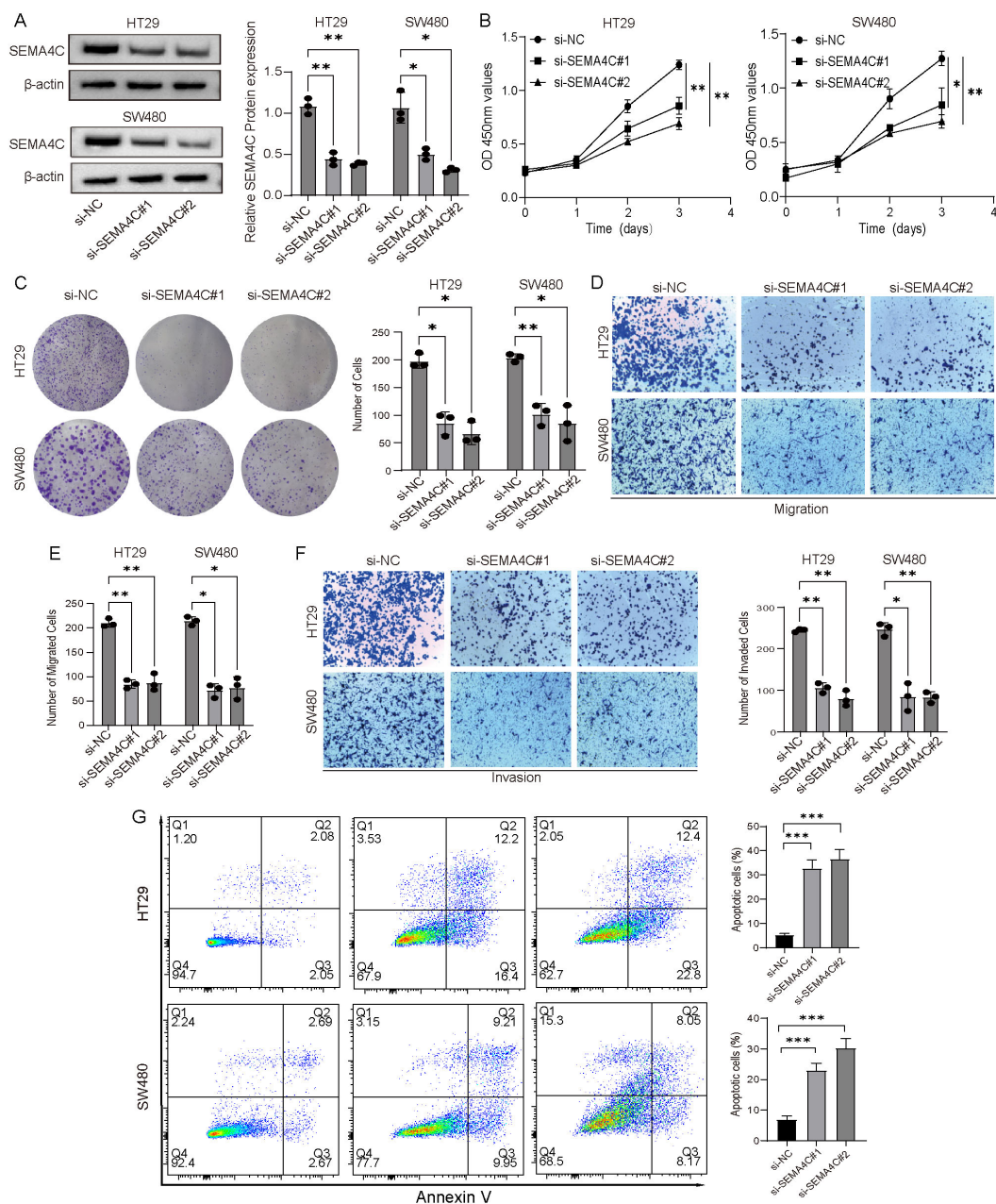


FIGURE 9

Verification of the tumor-promoting effect of SEMA4C in CRC. (A) Western blot analysis of SEMA4C expression in HT29 and SW480 cell lines. (B, C) CCK-8 and colony formation assays were conducted to evaluate the impact of SEMA4C downregulation on the proliferation and viability of HT29 and SW480 cell lines. (D–F) Transwell assays were performed to evaluate the influence of SEMA4C downregulation on the cell migration and invasion. (G) Impact of SEMA4C downregulation on the cell apoptosis was analyzed by flow cytometry assay. (*: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$).

predictive measures to enhance prognosis and therapeutic outcomes (43). Notably, our study revealed that the SRS predicts not only prognosis but also the occurrence and progression of CRC. It demonstrated outstanding diagnostic performance across various datasets, affirming its broad applicability. SRS is a crucial tool for assessing CRC patient survival rates and enhancing personalized medicine approaches.

In constructing the SRS, three SEMAs family genes were ultimately included: SEMA3E, SEMA4C, and SEMA6C. Class-3 SEMAs are a subfamily of seven vertebrate SEMAs known for being

the only secreted type (44). Initially identified as axon guidance factors, they also play roles in immune responses, angiogenesis, lymphangiogenesis, and other physiological and developmental functions. Notably, SEMA3E is a particularly intriguing gene due to its dual role in cancer biology. Research has demonstrated that SEMA3E could inhibit tumor angiogenesis (45). However, SEMA3E also has an important role in promoting tumor progression. This pro-tumorigenic effect is mediated by activating the oncogenic tyrosine kinase ERBB2 and the PI3K/AKT signaling pathways (46, 47). Furin-like pro-protein convertases, commonly found in

advanced invasive and metastatic cancers, regulate SEMA3E (48). These enzymes cleave SEMA3E into the p61 fragment, revealing a new signaling function. This fragment allows PLEXIND1 to interact with the oncogenic kinase ERBB2, initiating its signaling cascade. Although both SEMA3E and p61 bind to PlexinD1, only p61 can complex with ERBB2 and activate downstream pathways. Consequently, cancer cells may increase the conversion of precursor-SEMA3E into the pro-metastatic p61, promoting tumor progression (46). Additionally, A recent study revealed that SEMA3E knockout inhibits dendritic cell (DC) migration by modulating CCR7 expression and increasing programmed death ligand 2 levels, compared to wild-type mice (49). These findings illustrate that SEMA3E plays a role in regulating DC migration and function during inflammation. The transmembrane protein SEMA4C is overexpressed in several malignant tumors, including breast, esophageal, gastric, and colorectal cancers. The biological function of SEMA4C in macrophage recruitment was reported to contribute to the malignant nature of tumors. It could interact with the PLEXINB2 receptor to activate the NF- κ B pathway, which induced the production of colony-stimulating factor 1, thereby promoting tumor growth and progression (50). Research on the role of SEMA6C in tumors is scarce. SEMA6C has recently been identified as a novel activator of focal adhesion kinase and Yes-associated protein that operates independently of cell adhesion mechanisms. This activation is essential for the viability and growth of cancer cells (51). By influencing key signaling pathways, SEMA6C plays a significant role in promoting tumor progression and survival, making it a potential target for therapeutic interventions in cancer treatment.

We analyzed immune-related signatures between the two groups using the GSEA algorithm and the IOBR R package. The SRS-high group showed significant activation of various oncogenic pathways and exhibited a greater propensity for a cold tumor phenotype (52). In contrast, the SRS-low group exhibited elevated TMB and TNB, meaning a richer variety of immune cell types and an increased presence of tumor neoantigens. These factors may enhance the immune response, making the tumor more easily recognized by the immune system, thereby improving the response to immunotherapy (53). Survival analysis indicated better prognostic outcomes for the SRS-low group, validated across multiple immunotherapy cohorts. Furthermore, TIDE analysis revealed an enhanced response to immunotherapy within the SRS-low group, indicating that SRS may serve as a valuable tool for the early identification of populations sensitive to immunotherapy.

scRNA-seq and ST were employed to explore the heterogeneity of malignant epithelial cells for a comprehensive understanding of the SRS in CRC. This study sought to elucidate the influence of SEMAs on the progression of CRC and its implications for immunotherapeutic strategies. Malignant epithelial cells with high SRS interacts strongly with myeloid and endothelial cells through SPP1 and COL4A2-ITGAV-ITGB8 signaling pathways, respectively. Similarly, Malignant epithelial cells with low SRS shows significant interactions with myeloid cells via MIF signaling and with endothelial cells through JAG1-NOTCH4 signaling pathways. ST analysis provided insights into the interaction patterns between various cell types within the tumor

microenvironment. Malignant epithelial cells exhibiting high SRS demonstrated stronger communication with myeloid and endothelial cells compared to those with low SRS, aligning with scRNA-seq analysis findings. The heterogeneity observed indicates that malignant epithelial cells with varying SRS levels possess unique biological traits, highlighting SEMAs as a potential therapeutic target for CRC.

Our study conducted a comprehensive analysis of SEMAs gene expressions in CRC using bulk RNA, scRNA-seq and ST, revealing expression disparities among cellular subpopulations. The study highlights SEMAs' involvement in tumor immune evasion and drug resistance. Nonetheless, we recognize various limitations within this study. Initially, we assessed and confirmed the SRS signature in both the training and external validation sets. These cohorts differed in size and sequencing platforms, although we utilized Z-score normalization to mitigate these differences. Validating our findings requires a large-scale, multi-center prospective study. Additionally, *in vivo* studies are required to elucidate the biological functions of SRS-related genes in CRC. In conclusion, although we have assessed the sensitivity of SRS-high subgroups to various small molecule drugs, further validation through *in vitro* drug assays and clinical trials is necessary.

Conclusion

We developed the SRS using a machine learning framework, which improved performance in predicting patient prognosis across different cohorts and showed significant correlations with immunotherapy response. The SRS has the potential to be an effective tool for predicting prognosis and tailoring personalized treatment for CRC patients. Our study provides new insights into the molecular mechanisms of CRC development and progression through bulk RNA sequencing, scRNA-seq, and ST.

Data availability statement

The data used in this study are publicly available. The TCGA RNA-seq data and corresponding clinical data were downloaded from the "TCGAbiolinks" R package. The GSE17537, GSE39582, GSE126044, GSE135222, GSE91061, GSE132465, GSE144735, and GSE166555 were downloaded from the GEO (<https://www.ncbi.nlm.nih.gov/geo/>). IMvigor210 cohort was downloaded from the "IMvigor210CoreBiologies" R package. In-house cohort1 and In-house cohort2 is available from our previous study (<https://data.mendeley.com/drafts/mdr6mgmvz3>). Further inquiries can be directed to the corresponding authors.

Ethics statement

This study was approved by the Ethics Committee of the Harbin Medical University Cancer Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the

participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

JZ: Conceptualization, Formal analysis, Methodology, Project administration, Writing – original draft, Writing – review & editing. BX: Investigation, Visualization, Writing – original draft. ZW: Formal analysis, Methodology, Writing – original draft. ZY: Resources, Writing – original draft. SJ: Resources, Writing – original draft. JL: Formal analysis, Writing – original draft. HL: Conceptualization, Funding acquisition, Supervision, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Natural Science Foundation of China (grant no. 82303742 and 62372141), Key Research and Development Program of Heilongjiang Province (2023ZX06C09), China Postdoctoral Science Foundation (grant no. 2024T170205), Haiyan Foundation of Harbin Medical University Cancer Hospital (grant no. 04000480) and China Postdoctoral Science Foundation (grant no. 2024MD763971).

Acknowledgments

We would like to extend our sincere gratitude to the reviewers for their invaluable and constructive comments provided during the

manuscript revision process. We thank Figdraw (figdraw.com) for providing a great plotting platform.

Conflict of interest

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

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2025.1536545/full#supplementary-material>

References

1. WHO. Cancer. Available online at: www.who.int/news-room/fact-sheets/detail/cancer (Accessed October 8, 2024).
2. Siegel RL, Wagle NS, Cercak A, Smith RA, Jemal A. Colorectal cancer statistics, 2023. *CA Cancer J Clin.* (2023) 73:233–54. doi: 10.3322/caac.21772
3. Jiang J, Zhang F, Wan Y, Fang K, Yan ZD, Ren XL, et al. Semaphorins as potential immune therapeutic targets for cancer. *Front Oncol.* (2022) 12:793805. doi: 10.3389/fonc.2022.793805
4. Al Zein M, Boukhoud M, Shammaa H, Mouslem H, El Ayoubi LM, Iratni R, et al. Immunotherapy and immunoevasion of colorectal cancer. *Drug Discovery Today.* (2023) 28:103669. doi: 10.1016/j.drudis.2023.103669
5. Huber AB, Kolodkin AL, Ginty DD, Cloutier JF. Signaling at the growth cone: ligand-receptor complexes and the control of axon growth and guidance. *Annu Rev Neurosci.* (2003) 26:509–63. doi: 10.1146/annurev.neuro.26.010302.081139
6. Neufeld G, Kessler O. The semaphorins: versatile regulators of tumour progression and tumour angiogenesis. *Nat Rev Cancer.* (2008) 8:632–45. doi: 10.1038/nrc2404
7. Klotz R, Thomas A, Teng T, Han SM, Iriando O, Li L, et al. Circulating tumor cells exhibit metastatic tropism and reveal brain metastasis drivers. *Cancer Discovery.* (2020) 10:86–103. doi: 10.1158/2159-8290.CD-19-0384
8. Mastrantonio R, You H, Tamagnone L. Semaphorins as emerging clinical biomarkers and therapeutic targets in cancer. *Theranostics.* (2021) 11:3262–77. doi: 10.7150/thno.54023
9. Chen LH, Cuang EY. Importance of semaphorins in cancer immunity. *Transl Lung Cancer Res.* (2019) 8:S468–70. doi: 10.21037/tlcr.2019.12.22
10. Li H, Wang JS, Mu LJ, Shan KS, Li LP, Zhou YB. Promotion of Sema4D expression by tumor-associated macrophages: Significance in gastric carcinoma. *World J Gastroenterol.* (2018) 24:593–601. doi: 10.3748/wjg.v24.i5.593
11. Kumanogoh A, Shikina T, Suzuki K, Uematsu S, Yukawa K, Kashiwamura S, et al. Nonredundant roles of Sema4A in the immune system: defective T cell priming and Th1/Th2 regulation in Sema4A-deficient mice. *Immunity.* (2005) 22:305–16. doi: 10.1016/j.immuni.2005.01.014
12. Suga Y, Nagatomo I, Kinehara Y, Koyama S, Okuzaki D, Osa A, et al. IL-33 induces sema4A expression in dendritic cells and exerts antitumor immunity. *J Immunol.* (2021) 207:1456–67. doi: 10.4049/jimmunol.2100076
13. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* (2016) 44:e71. doi: 10.1093/nar/gkv1507
14. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* (2018) 28:1747–56. doi: 10.1101/gr.239244.118
15. Smith JJ, Deane NG, Wu F, Merchant NB, Zhang B, Jiang A, et al. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. *Gastroenterology.* (2010) 138:958–68. doi: 10.1053/j.gastro.2009.11.005
16. Marisa L, de Reyniès A, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PloS Med.* (2013) 10:e1001453. doi: 10.1371/journal.pmed.1001453

17. Liu Z, Liu L, Weng S, Guo C, Dang Q, Xu H, et al. Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. *Nat Commun.* (2022) 13:816. doi: 10.1038/s41467-022-28421-6
18. Zeng D, Ye Z, Shen R, Yu G, Wu J, Xiong Y, et al. IOBR: multi-omics immunology biological research to decode tumor microenvironment and signatures. *Front Immunol.* (2021) 12:687975. doi: 10.3389/fimmu.2021.687975
19. Xu L, Deng C, Pang B, Zhang X, Liu W, Liao G, et al. TIP: A web server for resolving tumor immunophenotype profiling. *Cancer Res.* (2018) 78:6575–80. doi: 10.1158/0008-5472.CAN-18-0689
20. 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) 24:1550–8. doi: 10.1038/s41591-018-0136-1
21. Cho JW, Hong MH, Ha SJ, Kim YJ, Cho BC, Lee I, et al. Genome-wide identification of differentially methylated promoters and enhancers associated with response to anti-PD-1 therapy in non-small cell lung cancer. *Exp Mol Med.* (2020) 52:1550–63. doi: 10.1038/s12276-020-00493-8
22. Jung H, Kim HS, Kim JY, Sun JM, Ahn JS, Ahn MJ, et al. DNA methylation loss promotes immune evasion of tumours with high mutation and copy number load. *Nat Commun.* (2019) 10:4278. doi: 10.1038/s41467-019-12159-9
23. Riaz N, Havel JJ, Makarov V, Desrichard A, Urba WJ, Sims JS, et al. Tumor and microenvironment evolution during immunotherapy with nivolumab. *Cell.* (2017) 171:934–949.e16. doi: 10.1016/j.cell.2017.09.028
24. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. TGF β attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature.* (2018) 554:544–8. doi: 10.1038/nature25501
25. Wang M, Deng C, Yang C, Yan M, Lu H, Zhang Y, et al. Unraveling temporal and spatial biomarkers of epithelial-mesenchymal transition in colorectal cancer: insights into the crucial role of immunosuppressive cells. *J Transl Med.* (2023) 21:794. doi: 10.1186/s12967-023-04600-x
26. Lee HO, Hong Y, Etioglu 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) 52:594–603. doi: 10.1038/s41588-020-0636-z
27. Uhlitz F, Bischoff P, Peidli S, Sieber A, Trinks A, Lüthen M, et al. Mitogen-activated protein kinase activity drives cell trajectories in colorectal cancer. *EMBO Mol Med.* (2021) 13:e14123. doi: 10.15252/emmm.202114123
28. 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) 184:3573–3587.e29. doi: 10.1016/j.cell.2021.04.048
29. Qi J, Sun H, Zhang Y, Wang Z, Xun Z, Li Z, et al. Single-cell and spatial analysis reveal interaction of FAP+ fibroblasts and SPP1+ macrophages in colorectal cancer. *Nat Commun.* (2022) 13:1742. doi: 10.1038/s41467-022-29366-6
30. Gao R, Bai S, Henderson YC, Lin Y, Schalck A, Yan Y, et al. Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat Biotechnol.* (2021) 39:599–608. doi: 10.1038/s41587-020-00795-2
31. Biermann J, Melms JC, Amin AD, Wang Y, Caprio LA, Karz A, et al. Dissecting the treatment-naïve ecosystem of human melanoma brain metastasis. *Cell.* (2022) 185:2591–2608.e30. doi: 10.1016/j.cell.2022.06.007
32. Andreatta M, Carmona SJ. UCell: Robust and scalable single-cell gene signature scoring. *Comput Struct Biotechnol J.* (2021) 19:3796–8. doi: 10.1016/j.csbj.2021.06.043
33. 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) 12:1088. doi: 10.1038/s41467-021-21246-9
34. Zhang Y, Liu T, Hu X, Wang M, Wang J, Zou B, et al. CellCall: integrating paired ligand-receptor and transcription factor activities for cell-cell communication. *Nucleic Acids Res.* (2021) 49:8520–34. doi: 10.1093/nar/gkab638
35. Cable DM, Murray E, Zou LS, Goeva A, Macosko EZ, Chen F, et al. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat Biotechnol.* (2022) 40:517–26. doi: 10.1038/s41587-021-00830-w
36. Bäckdahl J, Franzén L, Massier L, Li Q, Jalkanen J, Gao H, et al. Spatial mapping reveals human adipocyte subpopulations with distinct sensitivities to insulin. *Cell Metab.* (2021) 33:1869–1882.e6. doi: 10.1016/j.cmet.2021.07.018
37. 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) 23:97. doi: 10.1186/s13059-022-02663-5
38. Cang Z, Zhao Y, Almet AA, Stabell A, Ramos R, Plikus MV, et al. Screening cell-cell communication in spatial transcriptomics via collective optimal transport. *Nat Methods.* (2023) 20:218–28. doi: 10.1038/s41592-022-01728-4
39. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* (2005) 102:15545–50. doi: 10.1073/pnas.0506580102
40. Chu G, Ji X, Wang Y, Niu H. Integrated multiomics analysis and machine learning refine molecular subtypes and prognosis for muscle-invasive urothelial cancer. *Mol Ther Nucleic Acids.* (2023) 33:110–26. doi: 10.1016/j.omtn.2023.06.001
41. Cheng S, Li Z, Gao R, Xing B, Gao Y, Yang Y, et al. A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell.* (2021) 184:792–809.e23. doi: 10.1016/j.cell.2021.01.010
42. Li J, Wang D, Tang F, Ling X, Zhang W, Zhang Z. Pan-cancer integrative analyses dissect the remodeling of endothelial cells in human cancers. *Natl Sci Rev.* (2024) 11:nwae231. doi: 10.1093/nsr/nwae231
43. Sauer R, Liersch T, Merkel S, Fietkau R, Hohenberger W, Hess C, et al. Preoperative versus postoperative chemoradiotherapy for locally advanced rectal cancer: results of the German CAO/ARO/AIO-94 randomized phase III trial after a median follow-up of 11 years. *J Clin Oncol.* (2012) 30:1926–33. doi: 10.1200/JCO.2011.40.1836
44. Toledano S, Nir-Zvi I, Engelman R, Kessler O, Neufeld G. Class-3 semaphorins and their receptors: potent multifunctional modulators of tumor progression. *Int J Mol Sci.* (2019) 20:556. doi: 10.3390/ijms20030556
45. Fukushima Y, Okada M, Kataoka H, Hirashima M, Yoshida Y, Mann F, et al. Sema3E-PlaxinD1 signaling selectively suppresses disoriented angiogenesis in ischemic retinopathy in mice. *J Clin Invest.* (2011) 121(5):1974–85. doi: 10.1172/JCI44900
46. Casazza A, Kigel B, Maione F, Capparuccia L, Kessler O, Giraudo E, et al. Tumour growth inhibition and anti-metastatic activity of a mutated furin-resistant Semaphorin 3E isoform. *EMBO Mol Med.* (2012) 4:234–50. doi: 10.1002/emmm.201100205
47. Hagihara K, Haraguchi N, Nishimura J, Yasueda A, Fujino S, Ogino T, et al. PLXND1/SEMA3E promotes epithelial-mesenchymal transition partly via the PI3K/AKT-signaling pathway and induces heterogeneity in colorectal cancer. *Ann Surg Oncol.* (2022) 29:7435–45. doi: 10.1245/s10434-022-11945-y
48. Bassi DE, Fu J, Lopez de Cicco R, Klein-Szanto AJP. Proprotein convertases: “master switches” in the regulation of tumor growth and progression. *Mol Carcinog.* (2005) 44:151–61. doi: 10.1002/mc.20134
49. Movassagh H, Shan L, Koussih L, Alamri A, Ariaee N, Kung SKP, et al. Semaphorin 3E deficiency dysregulates dendritic cell functions: *In vitro* and *in vivo* evidence. *PloS One.* (2021) 16:e0252868. doi: 10.1371/journal.pone.0252868
50. Yang J, Zeng Z, Qiao L, Jiang X, Ma J, Wang J, et al. Semaphorin 4C promotes macrophage recruitment and angiogenesis in breast cancer. *Mol Cancer Res: MCR.* (2019) 17(10):2015–28. doi: 10.1158/1541-7786.MCR-18-0933
51. Fard D, Testa E, Panzeri V, Rizzolio S, Bianchetti G, Napolitano V, et al. SEMA6C: a novel adhesion-independent FAK and YAP activator, required for cancer cell viability and growth. *Cell Mol Life Sci.* (2023) 80:111. doi: 10.1007/s00018-023-04756-1
52. Galon J, Bruni D. Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. *Nat Rev Drug Discovery.* (2019) 18:197–218. doi: 10.1038/s41573-018-0007-y
53. Borst J, Ahrends T, Băbala N, Melief CJM, Kastenmüller W. CD4+ T cell help in cancer immunology and immunotherapy. *Nat Rev Immunol.* (2018) 18:635–47. doi: 10.1038/s41577-018-0044-0

Glossary

ANT	adjacent normal tissue	Macro_SPP1	SPP1-positive macrophages
AUC	area under the dose-response curve	MAF	mutation annotation format
CCK8	cell counting kit-8	OS	overall survival
CCL	broad institute cancer cell line encyclopedia	PCA	principal component analysis
CCLs	cancer cell lines	PCs	principal components
cDCs	conventional DCs	PMSF	phenylmethylsulfonyl fluoride
C-index	concordance index	PRISM	profiling relative inhibition simultaneously in mixtures
CNV	copy number variation	PVDF	polyvinylidene fluoride
COMMOT	COMMunication analysis by Optimal Transport	RCTD	robust cell type decomposition
CRC	colorectal cancer	RIPA	radio immunoprecipitation assay lysis
CTRP	cancer therapeutics response portal	scRNA-seq	single-cell RNA sequencing
DCs	dendritic cells	SEMA5	semaphorins
EMT	epithelial-mesenchymal transition	ST	spatial transcriptomic
EpiH	malignant epithelial cells with SRS-high	TCGA	the cancer genome atlas
EpiL	malignant epithelial cells with SRS-low	TIDE	tumor immune dysfunction and exclusion
GDC	genomic data commons	TIP	tracking tumor immunophenotype
GSEA	gene set enrichment analysis	TMB	tumor mutation burden
GTE	genotype-tissue expression	TME	tumor microenvironment
ICIs	immune checkpoint inhibitors	TNB	tumor neoantigen burden
IOBR	Immuno-Oncology Biological Research	Tregs	regulatory T cells
KEGG	Kyoto Encyclopedia of Genes and Genomes	UCEC	uterine corpus endometrial carcinoma
Macro_SLC40A1	SLC40A1-positive macrophages	UMAP	uniform manifold approximation and projection