Cancer Cell International

RESEARCH

Open Access

Construction of a risk scoring model based on machine learning and validation of the role of the key gene SERPINE1 in the progression of colon cancer

Xin Li[1†], Nana Li[2†], Yujie Wang[2†], Qixiang Han[2], Xiaodong Li[2], Qiushi Tu[2], Boshi Sun[2], Hao Yang[3*] and Yuan Gao[1*]

Abstract

Background and objective Colon cancer (CC) is a highly prevalent malignant tumor with a high mortality rate worldwide. Despite recent advancements in diagnosis and treatment, the overall prognosis for patients remains poor, especially for those with metastasis. Exploring key genes associated with the prognosis of patients with colon cancer, establishing effective molecular models, and validating their functions are necessary to optimize patient management and develop novel therapeutic strategies. This study aimed to reveal the role of the key gene SERPINE1 in the progression of colon cancer and its potential clinical application value through bioinformatics analysis and experimental validation.

Methods Multimodal data from colon cancer patients were obtained from The Cancer Genome Atlas (TCGA) database. Differentially expressed genes (DEGs) with a fold change ≥ 2 and an adjusted P value less than 0.05 were screened. A protein–protein interaction network was constructed using the STRING database, and 40 core genes were identified using Cytoscape software. Genes closely related to the prognosis of colon cancer patients were further selected using LASSO regression and Cox proportional hazards regression models to construct a risk scoring model for assessing patients' survival risk. The model performance was evaluated in combination with immune cell infiltration analysis and ROC curve assessment, and the associations between the target genes and the tumor immune microenvironment were explored. The expression and function of SERPINE1 were subsequently investigated. The expression level of SERPINE1 in colon cancer tissues was analyzed by Western blotting. The effects of SERPINE1 knockdown on the migration and invasion abilities of colon cancer cells were assessed by scratch wound healing and Transwell assays. The impact of SERPINE1 knockdown on tumor formation and metastasis in colon cancer cells was verified through mouse tumor and metastasis models.

†Xin Li, Nana Li and Yujie Wang contributed equally to this work.

*Correspondence:
Hao Yang
yanghao@hrbmu.edu.cn
Yuan Gao
yuan.gao@hrbmu.edu.cn

Full list of author information is available at the end of the article

Results The bioinformatics analysis identified seven target genes related to the prognosis of colon cancer, namely, CAMK2B, KIF1A, OPCML, SCN5A, SERPINE1, TH, and UCHL1, based on which a risk scoring model was constructed. Further analysis revealed that the risk score was not only associated with patients' survival prognosis but also significantly correlated with the infiltration of immune cells such as T cells and macrophages in the tumor immune microenvironment. Among these seven genes, SERPINE1 was highly expressed in colon cancer and was significantly positively correlated with a poor prognosis. In vitro experiments revealed that SERPINE1 knockdown inhibited the proliferation, migration, and invasion of colon cancer cells. Western blot experiments indicated that SERPINE1 knockdown upregulated the expression of epithelial cadherin (E-cadherin, E-CAD) and downregulated the expression of neural cadherin (N-cadherin, N-CAD) and vimentin (VIM), suggesting that SERPINE1 knockdown may inhibit the epithelial–mesenchymal transition (EMT). In a mouse subcutaneous tumor experiment, SERPINE1 knockdown significantly inhibited tumor growth. The liver metastasis model revealed that inhibiting SERPINE1 expression significantly reduced the number of liver metastases, further verifying the effect of high SERPINE1 expression on promoting the progression of colon cancer.

Conclusions A risk scoring model constructed based on the results of a bioinformatics analysis can effectively predict the survival prognosis of patients with colon cancer and reveals that SERPINE1 promotes the proliferation of colon cancer cells and accelerates the distant metastasis of tumor cells by inducing the EMT. These results provide an important reference for the prognostic assessment of colon cancer patients and pave the way for the development of therapeutic strategies targeting SERPINE1.

Keywords Colon cancer, SERPINE1, Risk score, Tumor immune microenvironment, Epithelial–mesenchymal transition

Introduction

Colon cancer (CC) is one of the most common malignant tumors and has high incidence and mortality rates globally. According to statistics from the International Agency for Research on Cancer, colon cancer ranks as the third most common cancer worldwide and the second leading cause of cancer-related death [1]. In recent years, due to the Westernization of dietary structures, changes in lifestyle, and population aging, the incidence of colon cancer has been continuously increasing in many countries [2]. In China, the incidence and mortality rates of colon cancer are also increasing, making it a significant public health concern [3].

Accumulating evidence suggests that colorectal cancer progression is driven by a complex interplay of epigenetic alterations, post-translational modifications, and tumor microenvironmental factors. DNA methylation–based biomarkers have shown promising diagnostic potential, although further validation in prospective and noninvasive settings is still required [4]. Mechanistically, multiple regulatory layers contribute to tumor development, including O-GlcNAcylation-mediated activation of NONO to promote tumor cell proliferation [5], as well as dysregulation of protein stability such as JOSD1-mediated YAP deubiquitination within the Hippo signaling pathway [6]. In addition, noncoding RNA networks and microbiota-related signaling, such as the HAS-CIRCpedia-5280/miR-4712-5p axis and Fusobacterium nucleatum-induced TLR4/NF-κB/PVT1 pathway, further modulate autophagy and chemoresistance [7, 8]. Moreover, tumor cells enhance their survival under stress conditions through coordinated regulation of DNA damage repair and cell cycle control. Mutant p53 facilitates resistance to endoplasmic reticulum stress by sustaining Ku70-dependent DNA repair via HDAC6, while p21 exhibits dose-dependent dual roles in apoptosis and genomic stability during chemotherapy [9–11]. From a clinical perspective, integrating molecular features such as RAS mutation status and tumor mutational burden (TMB) enables improved prognostic stratification, with RAS-mutant/TMB-low patients showing the poorest outcomes, whereas high TMB is associated with better survival even in microsatellite-stable cases [12–15]. Additionally, emerging therapeutic strategies targeting tumor suppressive pathways, such as TMEM52B-derived peptides that inhibit EGFR signaling, further highlight the potential for novel targeted interventions [16].

However, despite recent advancements in laparoscopic surgery, targeted therapy, and immunotherapy, the prognosis for patients with advanced colon cancer remains poor, especially for those with lymph node metastasis or distant metastasis, whose 5-year survival rates are significantly lower [1, 12]. Therefore, there is still a lack of sufficiently effective therapeutic strategies, highlighting an urgent need to identify novel targets to improve clinical outcomes.

In recent years, with the rapid development of high-throughput sequencing technologies and bioinformatics tools, researchers have been able to explore the molecular mechanisms of colon cancer at the genomic, transcriptomic, and proteomic levels [17, 18]. For example, public databases such as TCGA and GEO provide a vast amount of gene expression data. By integrating these data, DEGs related to the prognosis of colon cancer can

be identified [19–21]. The identification of these genes not only offers a new perspective for the molecular classification of colon cancer but also provides potential targets for personalized treatment [22]. Regrettably, the current exploitation of data is still insufficient, with most studies relying on single analytical methods, limiting the depth and breadth of the results. Machine learning-based analytical methods can more effectively identify key targets and more accurately predict patients' prognosis, providing a new approach for optimizing clinical treatment strategies.

In this study, we systematically analyzed the multimodal data of colon cancer patients from TCGA database to identify DEGs closely related to prognosis. By constructing protein–protein interaction networks, LASSO regression models, and Cox proportional hazards models, we identified seven key genes and constructed a risk scoring model. This study showed that this model not only can predict patient survival well but is also significantly correlated with immune cell infiltration in the tumor immune microenvironment. In further research, we found that SERPINE1, an important gene, is highly expressed in colon cancer and is significantly positively correlated with a poor prognosis. By combining in vitro and in vivo experiments, we verified that SERPINE1 promotes the proliferation, migration, and invasion of colon cancer cells by regulating the epithelial–mesenchymal transition (EMT) and accelerating distant metastasis.

This study aims to successfully construct a risk scoring model with good predictive ability by integrating bioinformatics and experimental validation, providing a new tool for the prognostic assessment of colon cancer patients. Moreover, this study reveals the role of SERPINE1 in the progression of colon cancer and clarifies its mechanism of influencing tumor metastasis by regulating the EMT and the immune microenvironment. These achievements not only provide theoretical support for precision medicine for colon cancer patients but also lay an important foundation for the development of targeted therapeutic strategies and new drug targets. Additionally, this study provides new potential biomarkers for the diagnosis and prognosis of colon cancer, which can help identify high-risk patients early, optimize treatment plans, and thereby improve patient survival rates and quality of life.

Materials and methods

Patient information and clinical tissues

Sixty patients with primary colon cancer diagnosed by pathology at the Second Affiliated Hospital of Harbin Medical University were included. We based our classification on the 8th edition of the American Joint Commission on Cancer's tumor, lymph node, and histological classification for colon cancer. The inclusion criteria were patients aged 18–65 years with no endocrine, cardiovascular, hematological, or infectious diseases. The exclusion criteria were as follows: patients younger than 18 years or older than 70 years and patients who had not received any antitumor treatment before surgery. This study was approved by the Scientific Ethics Committee of the Second Affiliated Hospital of Harbin Medical University in China. This study was conducted in accordance with the Declaration of Helsinki, and written informed consent was obtained from each participant before the study began (ethical approval number: KY2023-012).

Data collection and preprocessing

The data for this study were sourced from the public genomics database TCGA. Specifically, the transcriptomic data and clinical information from TCGA-COAD project were downloaded. This dataset includes the transcriptomic data and clinical data of 483 colon cancer tissues and 41 adjacent nontumor control tissues. Based on the clinical information, the clinical samples were divided into two groups: the N0 group (no lymph node metastasis) and the N1–3 group (with lymph node metastasis).

Differential expression analysis

The optimal regularization parameter (λ) was determined through 10-fold cross-validation to identify key genes and select them for inclusion in the multivariate Cox proportional hazards model using LASSO regression. The "survival" package in the R language was used to construct the model and calculate the risk score. The risk score was based on the Cox regression coefficients, and patients were divided into high- and low-risk groups according to the median value. Kaplan–Meier survival curves were plotted to compare differences. The "ggplot2" and "survminer" packages were used to visualize the survival analysis results, including survival curves, risk score distribution plots, and heatmaps. The predictive ability of the model was assessed using ROC curves, with higher AUC values indicating better performance. The C-index was calculated to evaluate the model's discriminatory ability. Univariate and multivariate Cox analyses were conducted to assess the independent diagnostic capability of the prognostic model [23].

Data collection and preprocessing

Gene expression profiles and corresponding clinical data for colorectal cancer were obtained from the TCGA-COAD cohort, which served as the primary dataset for model construction. Patients with incomplete survival information were excluded. Two independent external validation cohorts, GSE17536 and GSE271719, were downloaded from the Gene Expression Omnibus (GEO) database to evaluate the robustness and generalizability of the model.

For data preprocessing, TCGA RNA-seq count data were converted into transcripts per million (TPM) values and normalized using log2(TPM + 1). GEO datasets were processed using platform-specific normalization methods, followed by log2 transformation. Gene identifiers were annotated to official gene symbols, and when multiple probes corresponded to the same gene, their average expression values were used. To minimize technical bias and batch effects, each dataset was processed independently and not directly merged.

Model construction and risk stratification

The TCGA cohort was randomly divided into a training set and an internal validation set at a ratio of 1:1 using the "caret" R package. Based on 20 hub genes identified from prior network analysis, Least Absolute Shrinkage and Selection Operator (LASSO) regression was first applied in the training set using the glmnet package to reduce dimensionality and prevent overfitting, with the optimal penalty parameter determined via ten-fold cross-validation. Genes with non-zero coefficients were retained and subsequently subjected to univariate Cox proportional hazards regression analysis, with $P < 0.05$ considered statistically significant.

The selected genes were then included in multivariate Cox regression analysis to identify independent prognostic factors, resulting in four high-risk genes and their corresponding coefficients. A prognostic risk score model was constructed as a linear combination of gene expression levels weighted by their Cox regression coefficients: Risk score = Σ (Expression of gene_i×Coefficient_i). Patients were stratified into high- and low-risk groups based on the median risk score in the training cohort, and the same cutoff value was applied to the internal validation set and external cohorts.

Model evaluation and statistical analysis

The predictive performance of the prognostic model was evaluated in the training set, internal validation set, and external validation cohorts (GSE17536 and GSE271719). Kaplan–Meier survival analysis and log-rank tests were performed using the survival and survminer R packages to compare overall survival between high- and low-risk groups.

Time-dependent receiver operating characteristic (ROC) curves were generated using the timeROC R package to assess model performance at different time points, and the area under the curve (AUC) was calculated accordingly.

Protein–protein interaction (PPI) network

The STRING database (https://string-db.org/) was used to perform the PPI analysis and visualize the proteins encoded by the differentially expressed mRNAs, with an interaction score threshold set at 0.4. The results were then imported into Cytoscape software, and the cytoHubba plugin was used to identify core genes within the network. The top 40 proteins were selected for PPI network construction to reveal their potential interactions in colon cancer.

Prognostic model construction

In this study, we combined least absolute shrinkage and selection operator (LASSO) regression with multivariate Cox proportional hazards models to screen and establish a prognostic molecular model for colon cancer. First, the expression levels of the 40 core genes identified in the PPI network were extracted and integrated with patients' clinical prognostic information (including the survival time and status) to form a comprehensive dataset containing gene expression and clinical features. The "glmnet" package in R was used for the LASSO regression analysis. By introducing the L1 regularization term, LASSO regression can effectively handle multicollinearity in high-dimensional data and perform feature selection and model fitting simultaneously.

Analysis of immune cell infiltration

In this study, immune cell infiltration was analyzed using the "IOBR" package in R software, which integrates multiple algorithms. Normalized gene expression data and reference sets were input, and after default parameters were set, the infiltration scores of various immune cells in colon cancer samples were obtained. The "limma" package in R was used for the differential expression analysis to identify differentially expressed mRNAs between the groups. The screening threshold for differentially expressed genes was set at a fold change greater than 2 and a P value less than 0.05. The results were visualized in the form of volcano plots and heatmaps using the "ggplot" package, which intuitively displays the statistical significance and clustering of gene expression.

Single-sample gene set enrichment analysis (ssGSEA)

We employed the "GSVA" (Gene Set Variation Analysis) package in R software to systematically evaluate the activity of the biological pathways in the samples. This package can score gene expression profiles for HALLMARK pathways and KEGG pathways, revealing differences in pathway activity among different samples. Specifically, GSVA calculates the enrichment score of each gene set in a sample to assess changes in the activity of specific pathways. This method not only avoids the loss of information associated with a traditional binary classification (i.e., whether a sample contains genes of a particular pathway) but also provides a more continuous and comprehensive profile of pathway activity and expression.

Drug treatment sensitivity

We used the "pRRophetic" package in R software to predict the sensitivity of high-risk and low-risk CC patients to clinically relevant chemotherapeutic and targeted drugs. The pRRophetic package, which is based on gene expression data, uses machine learning algorithms to construct drug sensitivity prediction models that can estimate the response of tumor cells to various chemotherapeutic drugs. Specifically, the package uses pre-trained models derived from the Cancer Cell Line Encyclopedia (CCLE) and Genomics of Drug Sensitivity in Cancer (GDSC) database—two authoritative resources containing gene expression data and corresponding drug response profiles of thousands of cancer cell lines, including CRC cell lines. These pre-trained models establish a robust correlation between gene expression patterns and drug sensitivity, which is then applied to the gene expression profiles of our high-risk and low-risk CRC patient groups to estimate their drug response. For the analysis, we first input the normalized gene expression data (log2-transformed TPM values for TCGA-COAD cohort and log2-transformed normalized values for GEO cohorts) of high-risk and low-risk patients into the pRRophetic package. The package then calculates the half-maximal inhibitory concentration (IC50) for each selected drug in both groups; a lower IC50 value indicates higher sensitivity of tumor cells to the corresponding drug. Quality control of prediction results: After the prediction was completed, the distribution of IC50 values of each drug in the two groups was checked to exclude abnormal prediction values caused by data errors or algorithm deviation.

Cell culture

A 15 ml centrifuge tube with 10 ml of DMEM-F12(Gibco) was prepared in advance. Human colon cancer cell lines HCT116 and Lovo cells (purchased from Procell, Wuhan, China) remove from liquid nitrogen, quickly placed in a 37 °C water bath to thaw thoroughly, and then mixed in a centrifuge tube. After centrifugation at 1000 rpm for 5 min, the supernatant was aspirated, and the cells were resuspended and added to DMEM-F12 containing 10% fetal (Gibco) bovine serum. After mixing thoroughly, the cells were cultured in a 37 °C incubator with 5% CO2.

CCK-8 detection

HCT-116 cells in the logarithmic growth phase from the SERPINE1-KD group and the control group were seeded into 96-well plates at a density of 5×10^3 cells per well, with 100μL of culture medium added to each well, and cultured for 24 h to allow cell adhesion. Subsequently, cytarabine(HY-13605,MCE) at different concentration gradients (0, 0.625, 1.25, 2.5, 5, 10μmol/L) was added respectively, with 3 replicate wells set for each concentration, and a blank control group (only culture medium added, no cells) was set at the same time. Following this, the cells were treated with CCK-8 solution (Vazyme, Nanjing, China) for 2 h, and the optical density at 450 nm (OD450) was measured using a microplate reader (Bio-Rad, USA).

Transwell assay

The Transwell assay was conducted using 24-well plates and polycarbonate filters (8 μm pores, Corning). Transwell inserts without Matrigel were used to observe cell migration. A total of 2×10^4 cells were suspended in serum-free medium and plated in the upper chamber, while 600 μL of complete medium was added to the lower chamber. After incubation for 24 h, the cells in the Transwell chamber were fixed with paraformaldehyde and stained with 0.25% crystal violet. The cells that migrated to the lower side of the membrane were imaged and counted. Cell migration and invasion abilities were statistically analyzed using ImageJ software.

Wound healing assay

A six-well plate was prepared, and three evenly spaced horizontal lines were drawn on the bottom with a marker as a reference line for subsequent imaging. Cell suspensions were prepared at a density of 5×10^5 cells per well and seeded into six-well plates. An optical microscope was used to capture images of three random fields of view as the 0-hour control. Images were captured again after 24 and 48 h. Before each image was captured, the cells were rinsed with PBS to remove detached cells, and the culture medium containing 1% fetal bovine serum was replaced with fresh culture medium [24].

Tissue immunofluorescence staining

Frozen tissue sections were removed from the − 20 °C freezer and brought to room temperature. The sections were rinsed three times with PBS. They were then fixed with cold acetone for 20 min, followed by aspiration of the supernatant and washes with PBS. After permeabilization with 0.5% Triton X-100 for 20 min and washes with PBS, the sections were blocked with 5% BSA and goat serum for 1 h without washing. The primary anti-PAI-1 (1:200, Abcam) was added, and the samples were incubated overnight at 4 °C. The primary antibody was aspirated, and the sections were washed three times with PBS before the addition of the anti-AF-594 secondary antibody and incubation for 1 h at 37 °C in the dark. After washes with PBS, DAPI was added and incubated for 10 min to stain the nuclei. Finally, an antiquenching agent was added, and the sections were sealed with nail polish, air-dried in the dark, and observed and photographed under a confocal microscope.

Immunohistochemical (IHC) staining

Colon cancer tissues were fixed, dehydrated, and embedded in paraffin before sectioning (4 μm thick). The sections were dewaxed and rehydrated and then pretreated with 3% H_2O_2 for 20 min. The sections were placed in citrate buffer and heated in a pressure cooker until boiling, followed by a 2-minute steam timer; then, they were naturally cooled and washed three times with PBST (each for 5 min). Blocking was performed with 10% goat serum, and anti-PAI-1(1:200, Abcam) was added and incubated overnight at 4 °C. The next day, the sections were washed at room temperature, incubated with the secondary antibody for 30 min, and then washed three times with PBS. DAB was incubated with the sections for 10 min for color development, and the sections were rinsed with PBS and observed under a microscope. The nuclei were stained with hematoxylin, dehydrated with a gradient of alcohol solutions, treated with xylene, sealed with neutral resin, air-dried, and photographed [25].

Western blot

After trypsin digestion, the cells were washed three times with PBS, lysed with RIPA lysis buffer containing PMSF on ice for 30 min, sonicated for 1 min (13000 r), and centrifuged for 15 min to collect the supernatant for the protein concentration measurement. The samples were mixed with 5× loading buffer, boiled for 10–20 min at 100 °C, and stored at −20 °C. After the gel was prepared, protein samples were loaded into the wells and electrophoresed at 80 V until reaching the upper gel and at 120 V until reaching the bottom, after which electrophoresis was terminated. The proteins were transferred to a PVDF membrane at 250 mA on ice for 1–2 h. The membrane was blocked with 5% skim milk at 37 °C for 1 h, and then the primary anti-PAI-1(1:1000,Abcam), N-cadherin(1:10000,Abcam), E-cadherin(1:5000,Abcam), Vimentin(1:20000,Proteintech)was added and incubated overnight at 4 °C. After the membrane was washed three times with TBST, the secondary goat anti-mouse or goat anti-rabbit antibody (1:5000 Affinity) was added, and the mixture was incubated at room temperature for 1 h. The membrane was washed again three times with TBST, and the chemiluminescent substrate was added for detection using an ECL imaging system [26].

Lentivirus transduction

The shRNA sequences used were as follows:

SERPINE1-RNAi (60421-1)–CAGACAGTTTCAGGCTGACTT;

SERPINE1-RNAi (60423-1)–CATCATCAATGACTGGGTGAA;

SERPINE1-RNAi (60424-1)–GCATCTGTACAAGGAGCTCAT.

A scrambled shRNA was used as a negative control. Briefly, 1.5×10^5 tumor cells were seeded into 24-well plates in 500 μl of culture medium and cultured overnight at 37 °C. When the cell density reached 30%–50%, the cells were infected with the lentivirus in the presence of polybrene. The medium was replaced with 500 μl of fresh medium per well after 1 day, and the infection status was observed after 2–3 days. Stable cell lines were selected with puromycin for 3–4 weeks, and SERPINE1 knockdown was confirmed by Western blotting before use in subsequent experiments.

Subcutaneous tumor model

A total of 10 SPF BALB/c nude mice (male, 6 weeks old, 18–20 g) were randomly assigned to two groups with five mice per group via a computer-generated random number table by an independent investigator after 7 days of acclimatization under standard laboratory conditions. Confluent normal colon cancer cells and SERPINE1-knockdown colon cancer cells were digested with trypsin, washed with PBS, and adjusted to a density of 2 × 10^6 cells/ml. In a laminar flow hood, the BALB/c nude mice (male,6weeks old,18–20 g) were fixed according to the group, and 0.2 ml of the cell suspension was subcutaneously injected 0.3 cm into the back from the right front limb armpit. The length and width of the tumors were measured weekly with calipers, and the tumor volume was calculated using the formula V = 1/2*W^2*L, with the data recorded.

Liver metastasis model

Confluent normal colon cancer cells and SERPINE1-knockdown colon cancer cells were digested with trypsin, washed with PBS, and adjusted to a density of 2 × 10^6 cells/ml. Six- to eight-week-old nude mice($n = 3$) were administered analgesics and anesthetized in accordance with the experimental animal rights protocol. An 8-millimeter incision was made in the skin directly above the spleen. Within 30 to 60 s, 100 μl of cells were injected into the exposed tip of the spleen. After 5 min, the abdominal cavity was closed. The mice were euthanized after 6 weeks to obtain the liver.

Statistical methods

In the bioinformatics study, for comparisons between two groups of data, t tests or Mann–Whitney U tests were selected based on whether the data followed a normal distribution. For multiple comparison analyses, the false discovery rate (FDR) method was used to adjust P values to avoid type I errors. Correlations were determined using Spearman's correlation analysis. The survival analysis was conducted using the Kaplan–Meier survival analysis and log-rank tests; for Cox proportional hazards regression analysis (used to calculate hazard ratios [HR]),

95% confidence intervals (CIs) were provided for all HR values to ensure the completeness of result descriptions. All the statistical analyses were performed in R software, with $P < 0.05$ indicating statistical significance.

For experimental validation, all experiments were independently repeated three times. A single-blind procedure was strictly implemented throughout the experiment: mice were labeled with anonymous digital codes, the investigator responsible for modeling, breeding and outcome assessment was blinded to group allocation, and code decoding was only performed during data statistical analysis. The sample size was determined based on previous similar studies and the 3R principles to ensure statistical power. The statistical analysis and visualization of the data were performed using ImageJ and GraphPad Prism 9.0 software. The experimental data are presented as the means ± standard deviations. Independent samples following a normal distribution were analyzed using parametric t tests, whereas independent samples not following a normal distribution were analyzed using nonparametric t tests. $P < 0.05$ was considered statistically significant ($^*P < 0.05$, $P < 0.01$; * $P < 0.001$, $P < 0.0001$; ns: $P > 0.05$).

Results

Identification of core genes associated with the lymph node metastasis of colon cancer

Colon cancer samples were divided into N0 (no metastasis) and N1–3 (with metastasis) groups based on the lymph node metastasis status. This staging method not only clearly distinguishes samples with different metastatic statuses but also helps us gain a deeper understanding of the changes in gene expression during tumor progression and their impacts on patients' prognosis. We employed rigorous bioinformatics analysis techniques to screen for differentially expressed genes (DEGs). Specifically, we set the screening thresholds at a fold change ≥ 2 and an adjusted P value less than 0.05 to ensure that the selected genes had high statistical significance and biological relevance. Compared with the N0 group (no lymph node metastasis), 83 significantly differentially expressed genes (DEGs) were identified in the N1–3 group (with lymph node metastasis) (Fig. 1A and B),

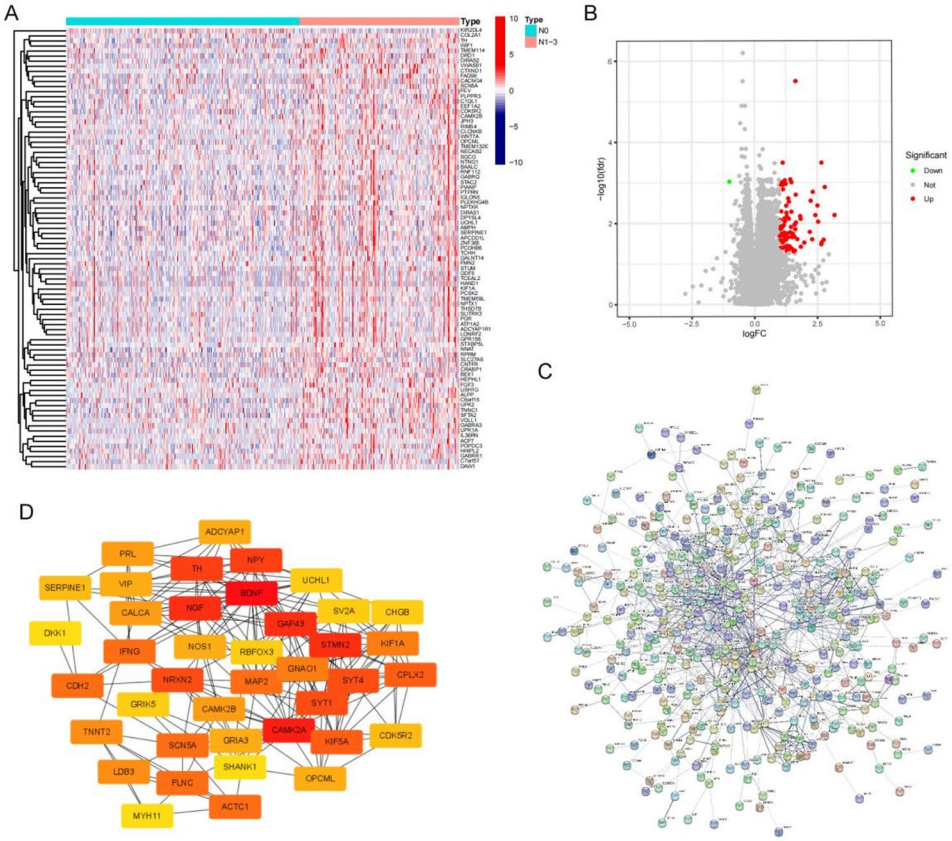

Fig. 1 Screening of differentially expressed genes (DEGs) and analysis of protein–protein interaction networks.(**A**) Hierarchical clustering heatmap of DEGs between N0 (non-metastatic) and N1–3 (lymph node metastatic) colon cancer samples (red: upregulated; blue: downregulated).(**B**) Volcano plot of DEGs filtered by $|\log_2 FC| \geq 1$ and adjusted $P < 0.05$ (red: upregulated; green: downregulated; gray: non-significant).(**C**) PPI network of DEGs constructed via STRING (confidence score ≥ 0.400).(**D**) Top 40 hub genes identified by cytoHubba, with TH, NGF, BDNF, etc., as core hub genes (red/orange nodes indicate higher centrality)

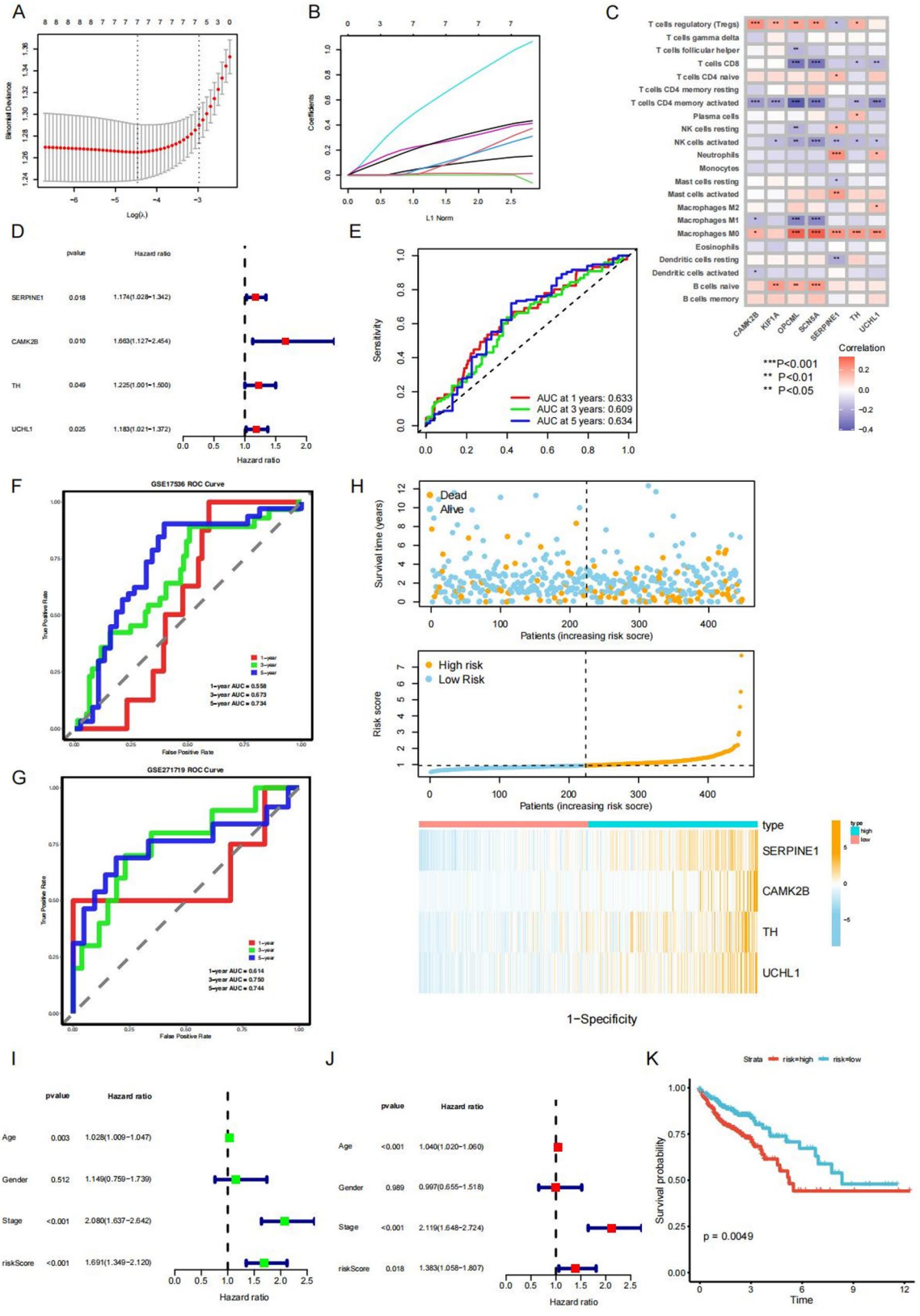

Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Construction and characterization of a colon cancer prognostic model via LASSO and Cox regression analyses. (**A**) LASSO regression with 10-fold cross-validation for optimal λ selection. (**B**) LASSO coefficient profiles identified seven prognostic genes: CAMK2B, KIF1A, OPCML, SCN5A, SERPINE1, TH, and UCHL1.(**C**) Correlation heatmap of seven core genes and immune infiltration. Red represents positive correlation and blue represents negative correlation. SERPINE1 was negatively correlated with Tregs, while CAMK2B and KIF1A showed positive correlations. (**D**) Multivariate Cox forest plot indicated that SERPINE1, CAMK2B, TH, and UCHL1 were independent prognostic predictors. (**E**) ROC curves revealed moderate predictive efficiency, with AUC values of 0.633 (1-year), 0.609 (3-year), and 0.634 (5-year). (**F**) Time-dependent ROC analysis of the prognostic model in the GSE17536 external validation cohort. (**G**) Time-dependent ROC analysis of the prognostic model in the GSE271719 external validation cohort.(**H**) Risk score distribution, survival status, and expression heatmap of model genes in high-risk (orange) and low-risk (blue) groups. (**I**) Univariate Cox analysis showed that age, gender, tumor stage and risk score significantly affected prognosis. (**J**) Multivariate Cox analysis confirmed risk score and stage as independent prognostic factors. (**K**) Kaplan–Meier survival analysis indicated poorer overall survival in the high-risk group ($p = 0.0049$)

providing an important foundation for further exploration of immune cell infiltration in the tumor microenvironment and potential biomarkers. After screening the DEGs, we used the STRING online analysis tool to construct a protein–protein interaction network containing 87 DEGs and set the confidence level at 0.400 to ensure the reliability and accuracy of the constructed network (Fig. 1C). The PPI results were subsequently imported into the cytoHubba plugin in Cytoscape software, and 40 core hub genes were screened from the constructed protein–protein interaction network and visualized (Fig. 1D). Among them, TH, NGF, BDNF, NPY, GAP43, STMN2, and CAMK2A were located at more central positions in the network.

Construction of a prognostic model to predict the prognosis of colon cancer patients

Through the construction of a LASSO regression model based on the 40 core genes described above, genes related to the prognosis of colon cancer patients were screened. The results showed that 7 genes, namely, CAMK2B, KIF1A, OPCML, SCN5A, SERPINE1, TH, and UCHL1, were closely related to the prognosis of patients with colon cancer (Fig. 2A-B). The analysis of immune cell infiltration revealed that the expression levels of CAMK2B, KIF1A, OPCML, SCN5A, and TH were significantly positively correlated with the infiltration score of Treg cells, whereas SERPINE1 expression was significantly negatively correlated with the infiltration score of Treg cells; all genes except SERPINE1 were negatively correlated with activated memory $CD4^+$ T cells; all genes except CAMK2B were negatively correlated with NK cells; and all genes except KIF1A were positively correlated with M0 macrophages (Fig. 2C). The multivariate Cox analysis indicated that among these 7 genes, the expression levels of SERPINE1, CAMK2B, TH, and UCHL1 had independent predictive power for the prognosis of patients with colon cancer, and their expression was associated with a poor prognosis (Fig. 2D). A prognostic model was constructed using the multivariate Cox algorithm, as described in the Materials and Methods section. Using this formula, we calculated the risk scores for colon cancer patients with prognostic information in TCGA database. The ROC curves revealed that the AUC value of the prognostic model was 0.633 for predicting the 1-year survival rate of patients with colon cancer, 0.609 for the 3-year survival rate, and 0.634 for the 5-year survival rate (Fig. 2E). To further evaluate the robustness and generalizability of the prognostic model, two independent GEO cohorts (GSE17536 and GSE271719) were used for external validation. The same risk score formula derived from the TCGA cohort was applied to these datasets. Time-dependent ROC analysis demonstrated improved predictive performance in the external cohorts, with AUC values reaching up to approximately 0.75 (Fig. 2F, G), indicating good generalizability of the model across independent datasets. Figure 2H shows the risk scores, survival status, and expression levels of the four genes constituting the model in patients with colon cancer. The survival analysis revealed that patients in the high-risk group had a worse prognosis than those in the low-risk group did (Fig. 2I). We included the clinical factors of colon cancer patients in TCGA database, such as age, sex, and stage, along with the prognostic model, in univariate and multivariate Cox regression analyses to investigate whether the prognostic model could serve as an independent predictor of the prognosis of colon cancer patients. The results showed that both the prognostic model and stage could serve as independent predictors of the prognosis of patients with colon cancer (Fig. 2J, K).

Differences in immune features between different risk score groups

To explore the association between the risk score and the tumor immune microenvironment (TIME) in colon cancer, we systematically analyzed the correlation between the risk score and immune cell infiltration, immune-related signatures, immune cell marker gene expression, and cancer-related pathway activity across different risk score groups. The analysis of immune cell infiltration revealed that the risk score was positively correlated with several immune cell types and scores, including activated myeloid dendritic cells, B cells, and $CD4^+$ naive T cells, but negatively correlated with common lymphoid progenitor cells and natural killer T cells (Fig. 3A).The correlation analysis between the risk score and immune response revealed that the risk score was positively correlated with the IFN-γ response ($r = 0.17$, $P < 0.001$),

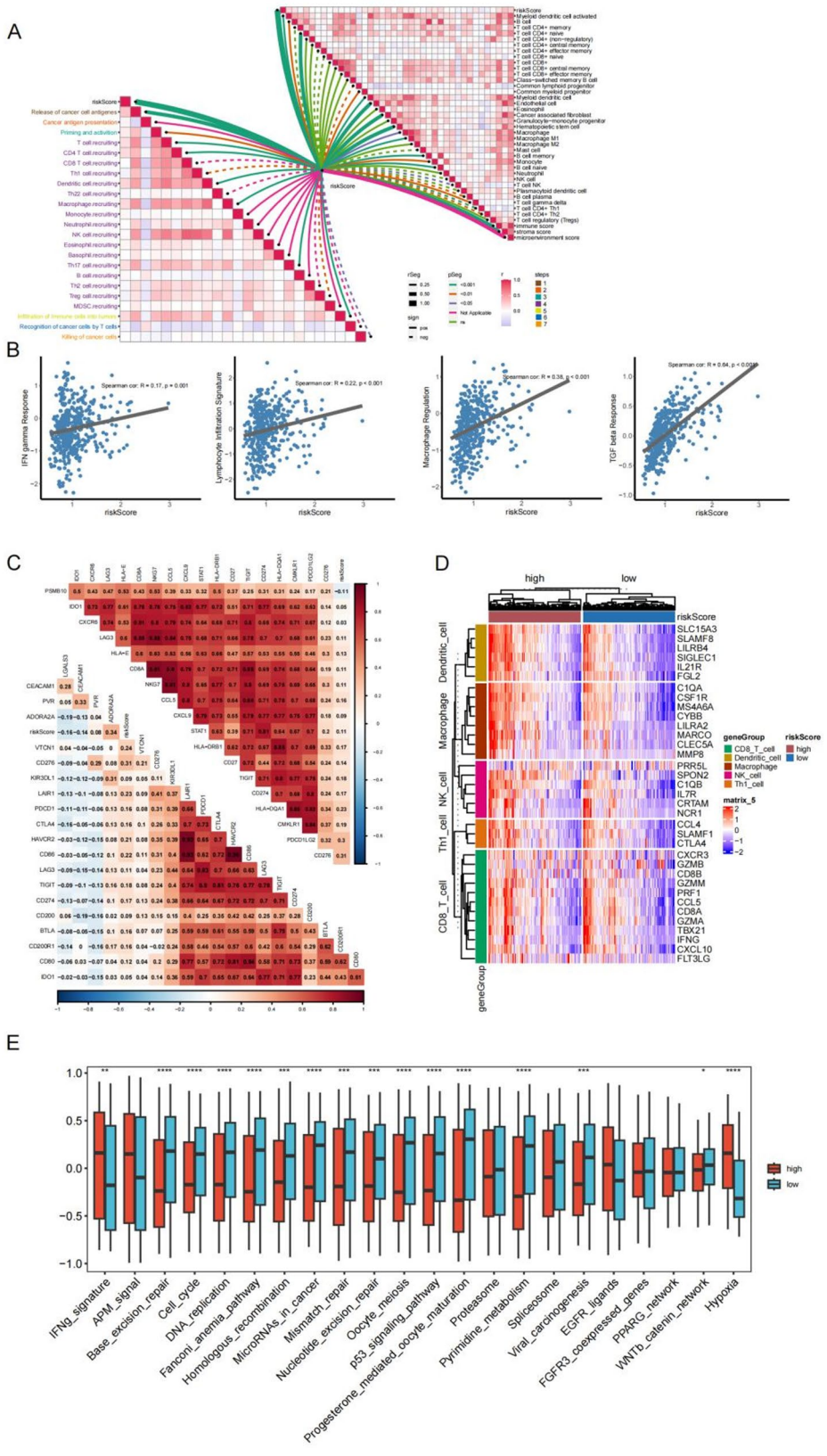

Fig. 3 (See legend on next page.)

(See figure on previous page.)

Fig. 3 Analysis of the correlation between the colon cancer risk score and immune profile. (**A**) Correlation network of risk score with immune cell infiltration and immune scores. Green lines represent positive correlations, and purple lines represent negative correlations. The risk score was positively correlated with activated myeloid dendritic cells, B cells and naive CD4⁺ T cells, while negatively correlated with common lymphoid progenitors and NKT cells. (**B**) Scatter plots showing correlations between risk score and multiple immune responses. Risk score was positively associated with IFN-γ response ($r = 0.17$, $P = 0.001$), lymphocyte infiltration ($r = 0.22$, $P < 0.001$), macrophage regulation ($r = 0.38$, $P < 0.001$), and TGF-β response ($r = 0.64$, $P < 0.001$). Dashed lines indicate correlation trends. (**C**) Heatmap illustrating expression correlations of immune cell signature genes in colon cancer. Color gradients from blue (low) to red (high) represent expression levels; correlation coefficients reflect the strength of pairwise associations. (**D**) Expression heatmap of immune signature genes between high- and low-risk groups. Immune cell-related genes (dendritic cells, macrophages, NK cells, Th1 and CD8⁺ T cells) were differentially expressed between the two groups with the same color scale as in (**C**). (**E**) Comparison of IFN-γ signature, hypoxia score and cancer pathway scores between risk subgroups. Red represents the high-risk group and blue represents the low-risk group. ****$P < 0.0001$, ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$

lymphocyte infiltration ($r = 0.22$, $P < 0.001$), macrophage regulation ($r = 0.38$, $P < 0.001$), and the TGF-β response ($r = 0.64$, $P < 0.001$) (Fig. 3B). These results indicate that the risk score is closely associated with the functional status of the TIME, and changes in the risk score may be accompanied by alterations in immune response activity and lymphocyte infiltration intensity.

We subsequently explored the differential expression of characteristic genes of various immune cells between the high- and low-scoring groups. Figure 3C shows the correlation between the expression of each gene in colon cancer patients from the two groups. Compared with that in the low-scoring group, the expression of characteristic genes related to dendritic cells, macrophages, NK cells, and Th1 cells was increased in the high-scoring group, whereas the expression of characteristic genes related to $CD8^+$ T cells was decreased (Fig. 3D). In addition, we analyzed the differences in the activity of cancer-related pathways and specific signatures between the high-risk and low-risk groups.Compared with those in the low-scoring group, the IFN-γ signature score and hypoxia score were higher in the high-scoring group, whereas the scores of multiple cancer-related pathways were lower (Fig. 3E).These results suggest that the risk score is associated with the functional status of cancer-related pathways and the hypoxic microenvironment.

Collectively, these findings demonstrate that the risk score is closely associated with the immune landscape of colon cancer, including immune cell infiltration, immune-related signature activity, immune cell marker gene expression, and cancer-related pathway activity. Future studies can further verify the regulatory mechanisms underlying these associations, such as exploring the role of risk score-related factors in regulating immune cell function and pathway activity, which may provide new insights for the immunotherapeutic management of colon cancer.

Correlation analysis between the risk score and immune features

We subsequently analyzed immune cell infiltration using various algorithms (such as xCell, TIMER, quanTIseq, MCP-counter, EPIC, CIBERSORT-ABS, and CIBERSORT) to explore the correlation between the risk score and immune cell infiltration (Fig. 4A). The results from multiple algorithms all indicated that the risk score was significantly positively correlated with tumor-associated fibroblasts and M2 macrophages (Fig. 4B), suggesting that the higher the risk score was, the greater the infiltration of these two immune cell types. The relationships between the risk score calculated by the ESTIMATE algorithm and the immune score, stromal score, and ESTIMATE score were subsequently explored. The results showed that the immune score, stromal score, and ESTIMATE score of the high-risk-score group were significantly higher than those of the low-risk-score group ($p < 0.05$, Fig. 4C). Subsequently, we further revealed the correlation between the model score and the expression of four genes in the model with Hallmark pathways and KEGG pathways. The results revealed that both the risk score and the expression of the four genes were related to various oncogenic pathways (Fig. 4D and E).

Prediction of drug treatment sensitivity in colon cancer patients using the prognostic model

The "prophetic" package was used to explore the differences in sensitivity to commonly used chemotherapeutic drugs between high-risk and low-risk groups of colon cancer patients. Drug candidates were selected according to their clinical relevance in colorectal cancer: they include agents either approved for CRC treatment, under clinical evaluation in advanced CRC, or widely reported to regulate proliferation, survival, or metabolism in colorectal cancer cells. The results of this study showed that, compared with the low-risk group, the high-risk group was less sensitive to bexarotene, cyclopamine, dasatinib, DMOG, imatinib, midostaurin, nilotinib, pazopanib, pyrimethamine, rapamycin, shikonin, and temsirolimus (Fig. 5A-L)but more sensitive to cytarabine and metformin(Fig. 5M, N).To validate these in silico findings, we further performed CCK-8 assays to evaluate the sensitivity of cytarabine in HCT-116 colon cancer cells with SERPINE1 knockdown (SERPINE1-KD) and corresponding control cells. Consistent with the bioinformatics prediction, the CCK-8 results demonstrated that the SERPINE1-KD group exhibited significantly enhanced sensitivity to cytarabine compared with the control group (Supplementary Fig. 2A, B). Specifically,

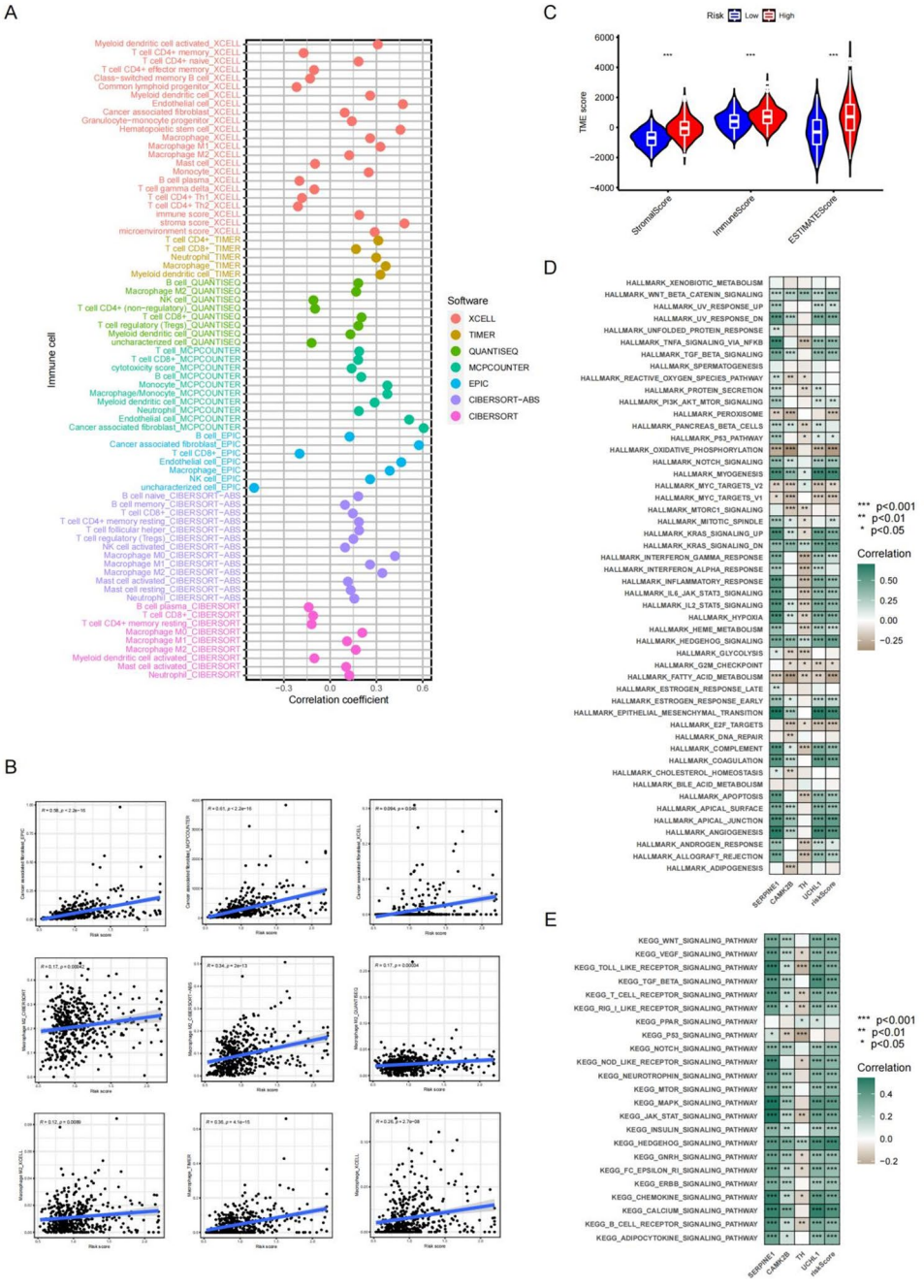

Fig. 4 Correlation analysis between the colon cancer risk score and immune cell invasion and pathways. (**A**) Analysis of immune cell infiltration. Immune cell infiltration was analyzed using a variety of algorithms (xCell, Timer, quanTIseq, MCP-counter, EPIC, CIBERSORT-ABS, and CIBERSORT). The dots in the graph represent the correlation scores of immune cells obtained using different algorithms; the color varies from red (positive correlation) to blue (negative correlation), and the size of the dots indicates the strength of the correlation. (**B**) Correlations of risk scores with specific immune cells. The scatter plot shows a positive correlation between risk scores and tumor-associated fibroblasts (CAFs) and M2 macrophages. The dashed line in each scatter plot represents the correlation trend, and the Spearman correlation coefficient (R) and P value represent the strength and statistical significance of the correlation. (**C**) The relationship between the risk score and the ESTIMATE algorithm score. The violin plot shows the differences in the distributions of immune scores, interstitial scores, and ESTIMATE scores between the high- and low-risk score groups. The white dots represent the median, the black lines represent the interquartile ranges, and the whiskers represent the minimum and maximum values of the data. An asterisk indicates statistical significance. (**D**) Correlations of the risk score with Hallmark pathways. The heatmap shows the correlations between the risk score and multiple Hallmark pathways. The color changes from green (positive correlation) to brown (negative correlation), with an asterisk indicating statistical significance. (**E**) Correlations of the risk score with KEGG pathways. The heatmap shows the correlations between the risk score and multiple KEGG pathways. The color coding is the same as that in Figure D, with an asterisk indicating statistical significance. ****$P < 0.0001$, ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$

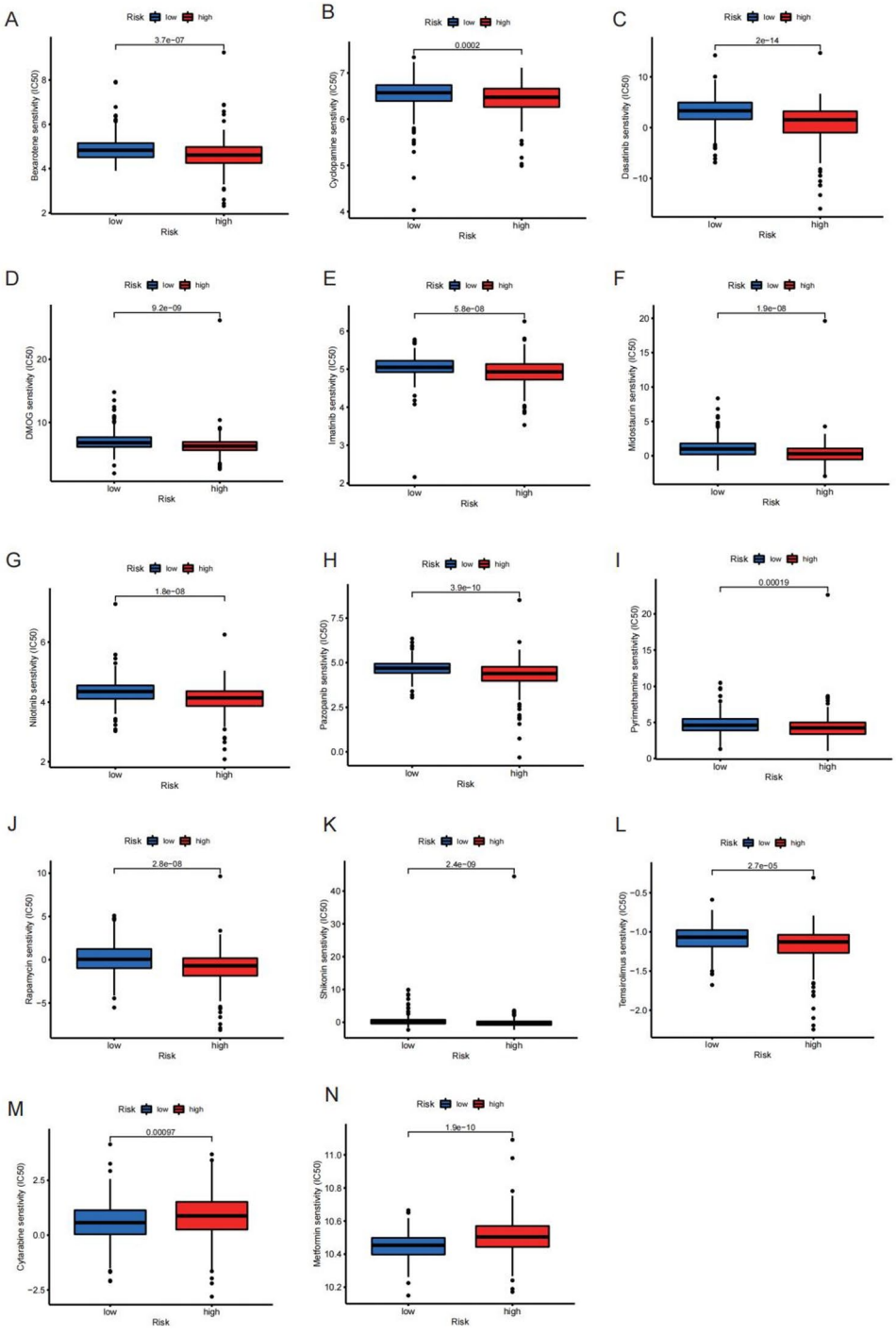

Fig. 5 Differences in drug sensitivity between different risk groups in terms of colon cancer scores. (**A**–**L**) Comparative analysis of predicted drug sensitivity between high-risk and low-risk groups using the pRRopheticalgorithm. Boxplots show the estimated IC50 values of bexarotene (**A**), cyclopamine (**B**), dasatinib (**C**), DMOG (**D**), imatinib (**E**), midostaurin (**F**), nilotinib (**G**), pazopanib (**H**), pyrimethamine (**I**), rapamycin (**J**), shikonin (**K**), and temsirolimus (**L**). The high-risk group exhibited significantly higher predicted IC50 values, indicating reduced sensitivity to these agents compared with the low-risk group. (M, N) Predicted IC50 values of cytarabine (**M**) and metformin (**N**) in high- and low-risk groups. The high-risk group showed significantly lower IC50 values, suggesting increased sensitivity to these drugs

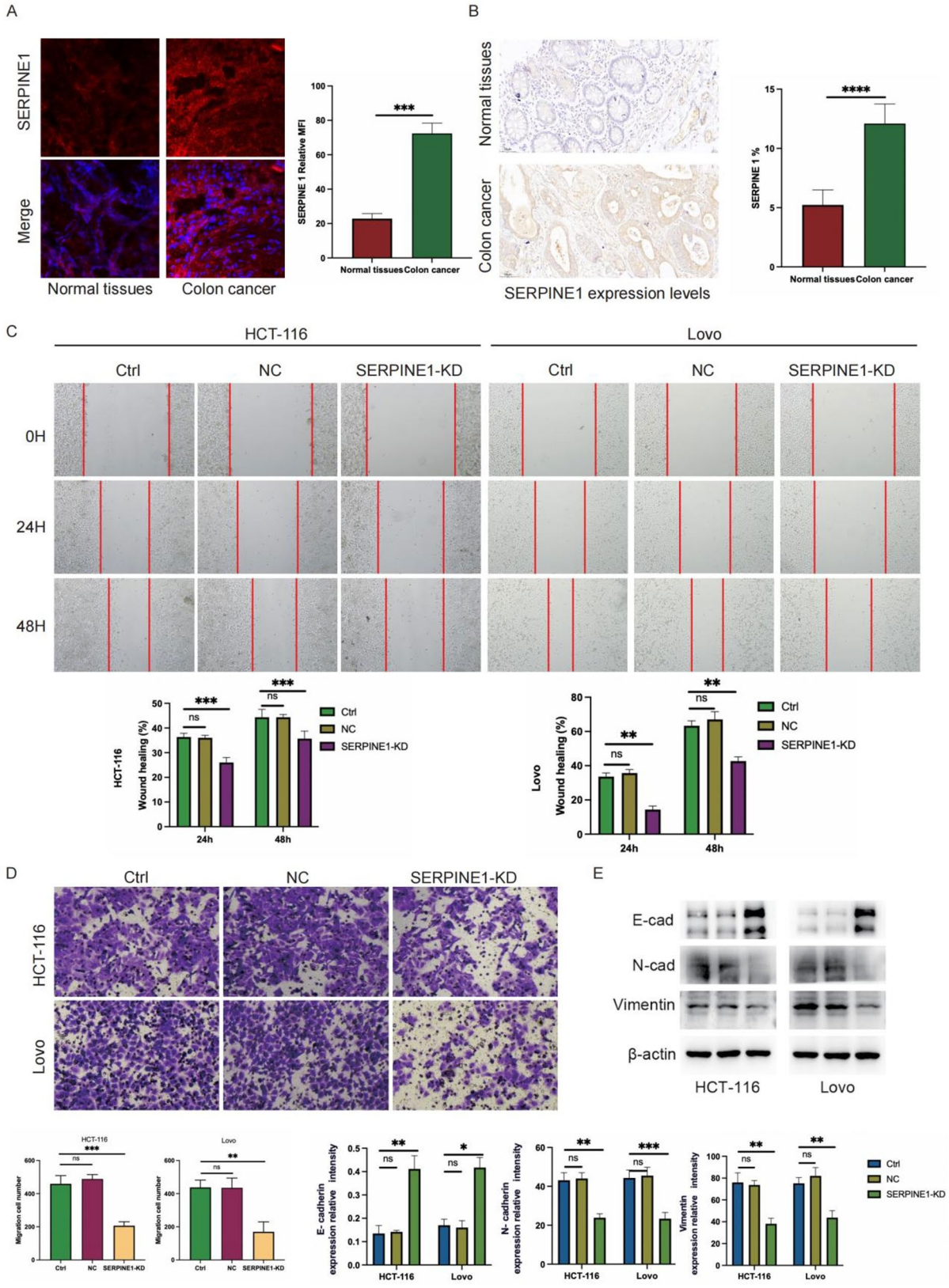

Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 Results of the quantitative analysis of SERPINE1 expression levels in colon cancer tissues. (**A**) Immunofluorescence staining of SERPINE1 in normal and colon cancer tissues. SERPINE1 fluorescence intensity was significantly higher in cancer tissues, and statistical comparison of mean fluorescence intensity (MFI) is shown in the histogram. (**B**) Immunohistochemical detection further confirmed elevated SERPINE1 protein levels in colon cancer tissues compared with normal controls, with quantitative analysis presented in the bar graph. (**C**) Scratch wound-healing assay for cell migration. In HCT116 and Lovo cells, the migration ability was markedly suppressed in the SERPINE1 knockdown (SERPINE1-KD) group at 24 h and 48 h compared with the control (Ctrl) and negative control (NC) groups. (**D**) Transwell migration assay. Fewer migrated cells were observed in the SERPINE1-KD group than in Ctrl and NC groups in both cell lines. (**E**) Western blot detection of EMT-related proteins. Knockdown of SERPINE1 significantly increased the epithelial marker E-cadherin, while decreasing the mesenchymal markers N-cadherin and vimentin. All values are mean ± SD. $^{****}P < 0.0001$, $^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$

the half-maximal inhibitory concentration (IC50) of cytarabine in SERPINE1-KD HCT-116 cells was significantly lower than that in the control cells, indicating that the downregulation of SERPINE1 can effectively improve the responsiveness of colon cancer cells to cytarabine treatment. This experimental evidence not only confirms the reliability of our drug sensitivity prediction results but also reveals a potential regulatory role of SERPINE1 in modulating the efficacy of cytarabine in colon cancer.

Expression and biological functions of SERPINE1 in colon cancer

In this study, we first quantitatively analyzed the expression of SERPINE1 in colon cancer tissues using immunofluorescence and immunohistochemical techniques. We observed a significant increase in the expression of SERPINE1 in colon cancer tissues compared with normal tissues using immunofluorescence staining (Fig. 6A). A subsequent immunohistochemical analysis confirmed this finding, showing that the expression level of SERPINE1 in colon cancer tissues was significantly higher than that in normal tissues (Fig. 6B). We further investigated the role of SERPINE1 in colon cancer by selecting two cell lines, HCT-116 and Lovo, for experimental validation. The experimental design included three groups: an untreated group (Ctrl), a negative control group (NC), and a SERPINE1 knockdown group (KD). Using shRNA technology to knock down SERPINE1, we found that SERPINE1 was expressed in the untreated group (Ctrl) and negative control group (NC), while its expression was reduced in the KD groups (60421, 60423, and 60424), as confirmed by Western blotting. The cell line with the lowest expression was selected for subsequent experiments (Supplementary Fig. 1 A). We conducted a plate cloning experiment to further verify the function of SERPINE1, which revealed a significant decrease in the proliferation capacity of the SERPINE1-knockdown group (Supplementary Fig. 1B). We also observed a significant reduction in the migration capacity of colon cancer cells in Transwell and scratch wound healing assays. The Transwell assay revealed that the number of HCT-116 and Lovo cells that penetrated the chamber membrane in the SERPINE1-KD group was significantly lower than that in the Ctrl and NC groups (Fig. 6C). The scratch wound healing assay further confirmed these results, showing that the degree of wound healing in the SERPINE1-KD group was significantly lower than that in the NC and Ctrl groups at 24 and 48 h (Fig. 6D). The proliferation and invasion of tumor cells are highly complex processes regulated by various mechanisms and factors. We detected changes in the expression of proteins related to the epithelial–mesenchymal transition (EMT) using Western blotting and confocal microscopy of immunofluorescence staining to further explore the role of SERPINE1 in colon cancer. The experimental results revealed that in the HCT-116 and Lovo cell lines, SERPINE1 knockdown affected the expression levels of the key EMT markers E-cadherin (E-CAD), N-cadherin (N-CAD), and vimentin (VIM). Specifically, compared with that in the Ctrl and NC groups, the expression of E-CAD was significantly upregulated in the SERPINE1 knockdown (KD) group, whereas the expression of N-CAD and VIM was significantly downregulated in the KD group. These findings suggest that SERPINE1 knockdown may be associated with a reduced transition toward a mesenchymal phenotype, potentially by promoting the expression of the epithelial marker E-CAD while inhibiting the expression of N-CAD and VIM (Fig. 6E, Supplementary Fig. 1 C).

SERPINE1 knockdown inhibits tumor growth and metastasis in colon cancer animal models

In animal experiments, we systematically explored the effects of SERPINE1 knockdown on the tumorigenic and metastatic capabilities of colon cancer cells and comprehensively validated the important role of SERPINE1 knockdown in inhibiting tumorigenesis and metastasis through the construction of subcutaneous tumor and liver metastasis models. The results of the subcutaneous tumor experiments revealed that the tumor growth curve of the KD group was slower than that of the NC group (Fig. 7B), and the tumor volume and weight of the KD group were less than those of the NC group (Fig. 7A and C). Immunofluorescence staining of subcutaneous tumors showed that the expression of SERPINE1 was reduced in the KD group compared with the NC group (Fig. 7D). In the liver metastasis model, the number of liver metastases in the KD group was significantly lower than that in the NC group (Fig. 7E). These results indicate that SERPINE1 knockdown effectively inhibits tumor growth and metastasis, suggesting that SERPINE1 may

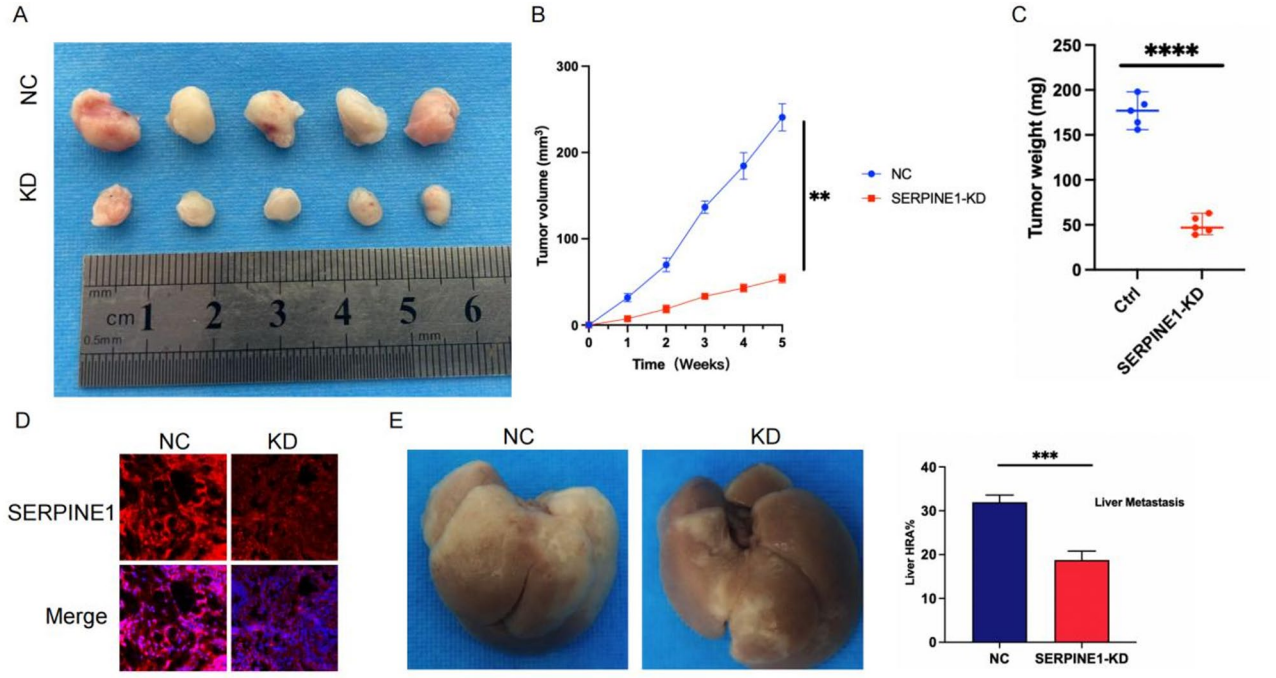

Fig. 7 Effects of SERPINE1 knockdown on colon cancer tumor growth and metastasis. (**A**) Volume of the subcutaneous tumors in the knockdown group (KD) compared with the control group (NC). The tumor volume in the knockdown group was significantly smaller than that in the control group. (**B**) Tumor growth plot depicting the change in the tumor volume over time in the control and knockdown groups. The curve shows that the tumor growth rate was slower in the knockdown group than in the control group. (**C**) Comparison of the actual masses of tumors from the control and knockdown groups in the subcutaneous tumor experiment. The tumor mass in the knockdown group was significantly smaller than that in the control group. (**D**) Immunofluorescence staining showed that the expression of SERPINE1 was lower in the KD group than in the NC group. (E) A comparison of liver metastasis in the control and knockdown groups revealed a significant reduction in the number of liver metastases in the knockdown group.All values are mean ± SD. ****$P < 0.0001$,***$P < 0.001$,**$P < 0.01$, *$P < 0.05$

promote colon cancer progression and confirming its critical role in the development of colon cancer.

Discussion

Using large-scale multimodal data from TCGA database, this study screened differentially expressed genes (DEGs) closely related to the prognosis of patients with colon cancer and identified seven prognosis-related genes for the construction of a Cox proportional hazards model through protein–protein interaction (PPI) network construction and a LASSO regression model. The constructed risk scoring model can effectively predict the 1-, 3-, and 5-year survival rates of colon cancer patients (AUC > 0.6), indicating good predictive performance. In recent years, the importance of risk scoring models based on multimodal data in evaluating the prognosis of patients with cancer has been widely validated. Studies have successfully developed multigene models for predicting the prognosis of patients with cancer through the integration of multimodal data, significantly improving the prediction accuracy [27, 28]. This finding suggests that the strategy of combining multimodal data and machine learning algorithms is an effective approach for identifying key tumor genes and using them for clinical risk predictions.

Further analysis revealed that the risk score not only predicts patient survival but also reflects differences in the tumor immune microenvironment. The CIBERSORT analysis revealed that the expression levels of some genes in the model was closely related to the infiltration of various immune cell types (such as macrophages and regulatory T cells). In patients with high-risk scores, the infiltration levels of immunosuppressive cells (such as M2 macrophages and fibroblasts) were elevated, whereas the levels of tumor-suppressing cells (such as $CD8^+$ T cells and NK cells) was reduced. Studies have indicated that immunosuppressive cells promote tumor immune escape and malignant progression by secreting various factors [29, 30]. Therefore, this risk scoring model not only provides a precise tool for predicting the prognosis but also offers a basis for exploring the potential mechanisms of the tumor immune microenvironment.

Additionally, the clinical application value of the model was demonstrated via a drug sensitivity analysis. This study revealed differences in the responses of CC patients to chemotherapy and targeted drugs based on the risk score. The results showed that the low-risk group was

more sensitive to bexarotene, cyclopamine, dasatinib, DMOG, imatinib, midostaurin, nilotinib, pazopanib, pyrimethamine, rapamycin, shikonin, and temsirolimus, whereas the high-risk group was more sensitive to cytarabine and metformin. These findings provide important guidance for the development of personalized treatment strategies. For example, for high-risk CRC patients (with high-risk scores), our prediction shows they are more sensitive to cytarabine and metformin—cytarabine is a first-line adjuvant chemotherapy drug for advanced CRC, and metformin has been shown to improve the efficacy of chemotherapy in CRC patients with diabetes, suggesting that combining these two drugs may be a more effective treatment strategy for high-risk patients. For low-risk CRC patients, their higher sensitivity to targeted drugs (e.g., dasatinib, pazopanib) and mTOR inhibitors (rapamycin, temsirolimus) provides an opportunity for less toxic targeted therapy instead of conventional chemotherapy, which can reduce treatment-related adverse events (e.g., myelosuppression, gastrointestinal toxicity) and improve the quality of life of low-risk patients [31, 32]. This treatment optimization strategy, which is based on the molecular classification, is expected to further improve the survival rate and quality of life of CC patients.

Through the construction of a LASSO regression model, we identified seven genes closely related to the prognosis of patients with colon cancer, namely, CAMK2B, KIF1A, OPCML, SCN5A, and SERPINE1, among others. The multivariate Cox analysis revealed that among these seven genes, the expression levels of SERPINE1, CAMK2B, TH, and UCHL1 have independent predictive power for the prognosis of colon cancer patients. Further analysis revealed that the expression level of SERPINE1 was significantly correlated with the infiltration pattern of immune cells in the tumor immune microenvironment. SERPINE1 is closely related to tumor invasion, metastasis, and immune regulation in various cancers (such as breast cancer and gastric cancer) [33, 34]. In addition, SERPINE1 has potential applications in cancer genesis and treatment by affecting the tumor microenvironment and immune regulation [35].

The serine protease inhibitor SERPINE1 orchestrates the degradation and remodeling of the extracellular matrix (ECM) through its regulatory effects on pivotal mediators within the fibrinolytic system, thereby establishing a permissive tumor microenvironment that potentiates migratory and invasive capacities of breast carcinoma cells [36].Furthermore, SERPINE1 promotes distant tumor metastasis by stimulating the secretion of angiogenic factors, such as vascular endothelial growth factor (VEGF), thereby establishing a pro-angiogenic vascular microenvironment that supports metastatic cell survival and dissemination [37]. In the context of therapeutic resistance, elevated SERPINE1 expression is associated with chemoresistance in lung adenocarcinoma, particularly to paclitaxel and gemcitabine, potentially through activation of drug efflux pumps or anti-apoptotic signaling pathways [38]. Notably, preclinical studies reveal that co-administration of SERPINE1 inhibitors with immune checkpoint inhibitors (e.g., anti-PD-1 antibodies) synergistically enhances the efficacy of immunotherapy in gastric cancer, offering a novel strategy for precision oncology [39].

Accumulating evidence demonstrates its significant overexpression in CRC tissues compared to adjacent normal mucosa, which correlates with advanced tumor stage and poor prognosis [40, 41]. Mechanistically, SERPINE1-driven dysregulation of extracellular matrix (ECM) remodeling facilitates cancer progression by enhancing matrix metalloproteinase (MMP) activity and disrupting tissue architecture [42]. Additionally, SERPINE1 contributes to the tumor microenvironment (TME) by promoting cancer cell adhesion, migration, and invasion through integrin-mediated signaling cascades, thereby accelerating metastatic spread [43].Although SERPINE1 has been studied in other cancers, its specific mechanisms of action and associations with the immune microenvironment in colon cancer have not been fully elucidated. Therefore, we selected SERPINE1 as the focus of further research.

Based on the results of the bioinformatics analysis, this study further validated the function of SERPINE1 through in vitro and in vivo experiments and explored its molecular mechanisms in the progression of colon cancer. SERPINE1 was highly expressed in CC tissues compared with normal tissues. The experiments further revealed that SERPINE1 knockdown regulates the epithelial–mesenchymal transition by increasing E-CAD expression and decreasing N-CAD and VIM expression, reducing the potential of tumor cells to transition to a mesenchymal phenotype. These findings confirm the key role of SERPINE1 in shaping the malignant phenotype of CC cells and suggest that it may accelerate tumor progression by promoting the EMT.

In addition, animal experiments further validated the function of SERPINE1 at the systemic level. The results from the mouse subcutaneous tumor model showed that SERPINE1 knockdown significantly reduced the tumor volume and mass. The liver metastasis model indicated that the number and volume of liver metastases in the SERPINE1-knockdown group were significantly decreased. These in vivo experimental results are highly consistent with the in vitro data and further confirm the key role of SERPINE1 in tumor invasion and metastasis.

To further explore the association between SERPINE1 and metastasis-related molecules, we focused on its intimate link with the epithelial-mesenchymal transition

(EMT) process. Accumulating evidence suggests that SERPINE1 interacts with core metastatic mediators, including matrix metalloproteinase 1 (MMP1) [44, 45], matrix metalloproteinase 2 (MMP2) [46, 47], and transforming growth factor-β (TGF-β) [48, 49], as well as with metastatic structures and the tumor microenvironment (TME) [50, 51].As a downstream effector of TGF-β-induced EMT, SERPINE1 forms a positive feedback loop with TGF-β signaling to sustain mesenchymal phenotypes and enhance cell motility. Meanwhile, SERPINE1 cooperates with MMP1 and MMP2 to promote extracellular matrix (ECM) degradation and basement membrane disruption, which are critical steps for cancer cell invasion and distant metastasis.These findings collectively position SERPINE1 as a central hub linking EMT, matrix remodeling, and immune suppression, highlighting its pivotal role in driving colon cancer metastasis.

This study revealed that high SERPINE1 expression not only regulates the malignant behavior of tumor cells themselves but also may exacerbate tumor progression by altering the infiltration characteristics of immune cells in the immune microenvironment. These findings provide an important reference for the future development of therapeutic strategies targeting SERPINE1. Mechanistically, SERPINE1 may mediate the formation of an immunosuppressive TME through multiple pathways: it can activate the TGF-β/Smad signaling pathway to promote M2-type macrophage polarization, which in turn secretes IL-10 and TGF-β to suppress CD8 + T cell activity—a regulatory pattern verified in triple-negative breast cancer [52].Meanwhile, it modulates extracellular matrix remodeling to enhance the recruitment and accumulation of regulatory T cells, and increases matrix stiffness and collagen deposition to form a physical barrier that impairs the infiltration of anti-tumor immune cells into the tumor parenchyma, a mechanism analogous to that by which SERPINE1 shapes an immune-exempt niche in pancreatic cancer [53]. Collectively, our study is the first to elucidate the dual pro-tumor roles of SERPINE1 in colon cancer, namely regulating the malignant phenotype of tumor cells and reshaping the immunosuppressive TME, identifying it as a core molecular hub linking tumor proliferation, metastasis and TME immune imbalance. This finding lays a theoretical foundation for the development of SERPINE1-targeted therapies, as targeted inhibition of SERPINE1 can not only directly suppress the growth and metastasis of colon cancer cells but also reverse the tumor immunosuppressive state and restore the body's anti-tumor immune function, achieving a dual anti-tumor effect of targeted therapy combined with immune activation.

Overall, this study, which combines bioinformatics analysis and experimental validation, not only constructed a precise risk scoring model, providing a scientific basis for the prognostic assessment and personalized treatment of CC patients, but also revealed the multiple roles of SERPINE1 as a key gene in CC. It regulates the EMT and the tumor immune microenvironment by modulating the expression of the related proteins E-CAD, N-CAD, and VIM, promoting tumor progression and exacerbating the immunosuppressive state. These findings not only suggest potential directions for future targeted therapeutic strategies but also provide a reference for further research on the role of SERPINE1 in other tumors. As related research progresses, the outcomes of this study are expected to accelerate the development of precision medicine for CC and ultimately provide new solutions to improve patients' prognosis and quality of life.

Despite identifying the important role of SERPINE1 in colon cancer and constructing a prognostic risk model with potential clinical applicability, several limitations should be acknowledged. Although external validation was performed using independent GEO cohorts, the model was primarily developed based on retrospective public datasets such as TCGA, which may introduce selection bias and limit its generalizability across diverse populations; moreover, cross-platform heterogeneity among transcriptomic datasets may also influence model performance. In addition, the predictive performance of the model in the TCGA cohort was moderate, suggesting that it may be more suitable for patient risk stratification rather than precise individual outcome prediction, and future studies integrating multi-omics data and comprehensive clinical variables may further improve its predictive accuracy. Furthermore, while a Cox regression–based model was adopted due to its interpretability and clinical relevance, more advanced machine learning approaches may further enhance predictive performance and will be explored in future studies. Finally, although our experimental results confirmed the involvement of SERPINE1 in epithelial–mesenchymal transition (EMT) and the tumor microenvironment, the detailed molecular mechanisms underlying its function, as well as its potential role in modulating drug response, remain to be fully elucidated and warrant further investigation.

Conclusions

This study identified genes related to lymph node metastasis in colon cancer through a bioinformatics analysis and experimental validation, constructed a prognostic model, and revealed correlations between the risk score and immune features. Additionally, this study explored the expression and biological function of SERPINE1 in colon cancer in detail, revealing its important roles in regulating tumor progression and the immune microenvironment. These findings provide new targets and ideas for the diagnosis, prognostic assessment, and treatment

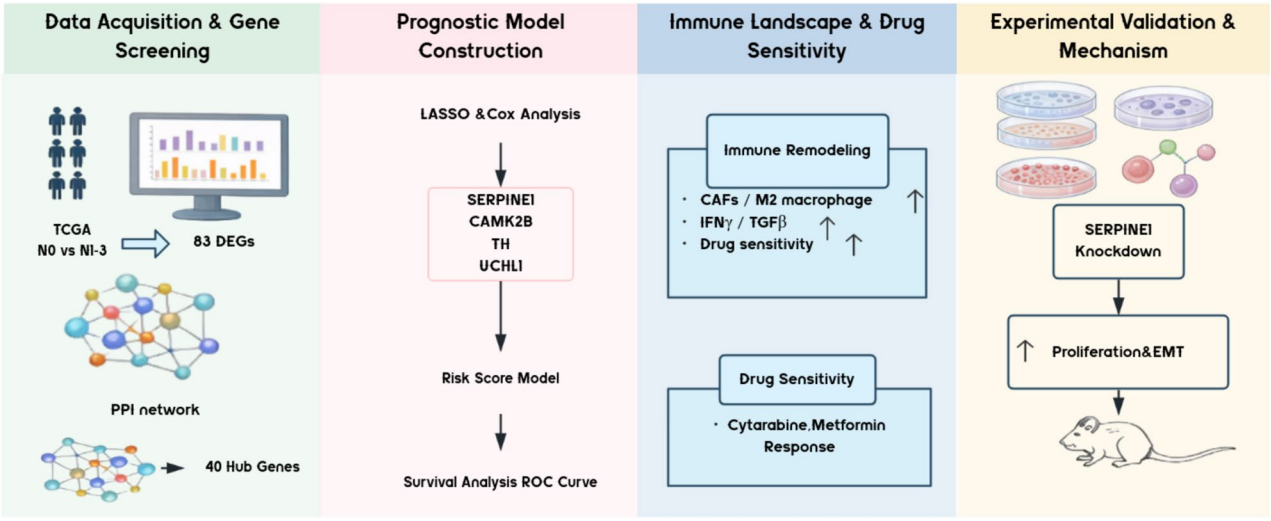

Fig. 8 Schematic illustration of the overall workflow of this study. Public datasets (e.g., TCGA-COAD) were used for differential expression analysis and identification of metastasis-related genes. Candidate genes were further screened using machine learning methods to construct a prognostic model. The model was evaluated by survival analysis and ROC curves. Subsequently, immune infiltration, functional enrichment, and drug sensitivity analyses were performed to explore biological and clinical relevance. Finally, key findings were validated through in vitro experiments

of colon cancer. Future research can explore the specific molecular mechanisms of these genes in the occurrence and development of colon cancer and evaluate their potential for clinical application (Fig. 8)

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12935-026-04322-2.

Supplementary Material 1

Supplementary Material 2

Acknowledgements
Not applicable..

Author contributions
Xin Li: bioinformatics analysis, manuscript writing, vitro assays. Nana Li and Yujie Wang: prepare tissue and conduct immunohistochemistry, animal experiment. Qixiang Han and Xiaodong Li: Conduct immunohistochemistry, manuscript editing. Qiushi Tu: bioinformatics analysis supervision. Boshi Sun: manuscript revision, providing fund. Hao Yang: performed initial bioinformatics analysis. Yuan Gao: experimetal design, manuscript editing.

Funding
This work was supported by Heilongjiang Province Postdoctoral Research Startup Fund [Grant No. LBH-Q21033] and National Natural Science General Program Cultivation Project Fund of the Second Affiliated Hospital of Harbin Medical University [Grant No. 20230125].

Data availability
All the data and material could be traced from the paper or can be requested from the corresponding author.

Declarations

Ethics approval and consent to participate
Our study was approved by the Ethics Committee of the second affiliated hospital of Harbin Medical University (KY2023-012).

Consent for publication
All the listed authors have participated in the study, and have seen and approved the submited manuscript.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Head and Neck Radiotherapy, Harbin Medical University Cancer Hospital, Harbin 150001, Heilongjiang Province, China
[2]Department of Colorectal Surgery, The Second Affliated Hospital of Harbin Medical University, Harbin 150001, Heilongjiang Province, China
[3]Department of General Surgery, Harbin Medical University Cancer Hospital, Harbin, Heilongjiang150001, China

Received: 7 April 2025 / Accepted: 22 April 2026
Published online: 05 May 2026

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229–63. https://doi.org/10.3322/caac.21834.
2. Matsuda T, Fujimoto A, Igarashi Y. Colorectal cancer: epidemiology, risk factors, and public health strategies. Digestion. 2025;106(2):91–9. https://doi.org/10.1159/000543921.
3. Huang J, Li H, Yan H, et al. The comparative burden of brain and central nervous system cancers from 1990 to 2019 between China and the United States and predicting the future burden. Front Public Health. 2022;10:1018836. https://doi.org/10.3389/fpubh.2022.1018836.
4. Chang R, Li G, Jin GN, Zhang B, Ding ZY. Identification of hypermethylated CpG sites mapped to LIFR as specific diagnostic biomarkers of colon cancer. Genes Dis. 2024;12(2):101334. https://doi.org/10.1016/j.gendis.2024.101334.
5. Kim Y, Lee KT, Kim HB, Jung H, Ko JY, Kweon TH, et al. O-GlcNAcylation of NONO regulates paraspeckle component assembly and contributes to colon cancer cell proliferation. Cell Death Discov. 2025;11(1):234. https://doi.org/10.1038/s41420-025-02405-z.
6. Sun Y, Liu D, Zhang X, Su P, Li X, Li Z, et al. Regulation of Hippo/YAP axis in colon cancer progression by the deubiquitinase JOSD1. Cell Death Discov. 2024;10(1):365. https://doi.org/10.1038/s41420-024-02136-7.

7. Gao K, Zhang J, Liu C, Yang Y, Wang J, Zhang J, et al. *Fusobacterium nucleatum* enhances oxaliplatin resistance in colon cancer by increasing PVT1 expression. J Transl Med. 2025;23(1):1112. https://doi.org/10.1186/s12967-025-07226-3.
8. An S, Cui J, Yang W, Zhang M, Yu H, Lu J, Tian Y, Qiao L, Wang X, Bao L, Zhao P. HAS-CIRCpedia-5280 sponges miR-4712-5p inhibited colon cancer autophagyinduced by human beta-defensin-1. J Transl Med. 2025;23(1):281. https://doi.org/10.1186/s12967-024-05860-x.
9. Benedetti R, Di Crosta M, Gilardini Montani MS, D'Orazi G, Cirone M. Mutant p53 upregulates HDAC6 to resist ER stress and facilitates Ku70 deacetylation, which prevents its degradation and mitigates DNA damage in colon cancer cells. Cell Death Discov. 2025;11(1):162. https://doi.org/10.1038/s41420-025-02433-9.
10. Kim H, Kim H, Jang E, Eom YW, Yoon G, Choi KS, et al. Dual role of p21 in regulating apoptosis and mitotic integrity in response to doxorubicin in colon cancer cells. Cell Death Discov. 2025;11(1):133. https://doi.org/10.1038/s41420-025-02416-w.
11. Tang Y, Yao F. Combined KRAS and TP53 mutation in patients with colorectal cancer enhance chemoresistance to promote postoperative recurrence and metastasis. BMC Cancer. 2024. https://doi.org/10.1186/s12885-024-12776-8.
12. Ianniello M, Ottaiano A, Bocchetti M, Ruggiero R, Santorsola M, Sirica R, et al. Tumor mutational burden modulates the prognostic effect of RAS mutations in metastatic colon cancer: mechanistic insights and genotype-phenotype correlations. J Transl Med. 2025;23(1):1226. https://doi.org/10.1186/s12967-025-07273-w.
13. Di Mauro A, Santorsola M, Savarese G, Sirica R, Ianniello M, Cossu AM, et al. Correction to: high tumor mutational burden assessed through next-generation sequencing predicts favorable survival in microsatellite stable metastatic colon cancer patients. J Transl Med. 2025;23(1):28. https://doi.org/10.1186/s12967-024-05971-5.
14. Di Mauro A, Santorsola M, Savarese G, Sirica R, Ianniello M, Cossu AM, et al. High tumor mutational burden assessed through next-generation sequencing predicts favorable survival in microsatellite stable metastatic colon cancer patients. J Transl Med. 2024;22(1):1107. https://doi.org/10.1186/s12967-024-05927-9.
15. Mondal K, Posa MK, Shenoy RP, Roychoudhury S. KRAS mutation subtypes and their association with other driver mutations in oncogenic pathways. Cells. 2024;13(14):1221. https://doi.org/10.3390/cells13141221.
16. Lee Y, Ko D, Yoon J, Kim S. TMEM52B-derived peptides inhibit generation of soluble E-cadherin and EGFR activity to suppress colon cancer growth and early metastasis. J Transl Med. 2025;23(1):146. https://doi.org/10.1186/s12967-025-06075-4.
17. Liu J, Jin X, Qiu C, Han P, Wang Y, Zhao J, et al. Integrated transcriptomics-proteomics analysis identifies molecular phenotypic alterations associated with colorectal cancer. J Proteome Res. 2023. https://doi.org/10.1021/acs.jproteome.3c00526.
18. Ullah I, Yang L, Yin FT, Sun Y, Li XH, Li J, et al. Multi-Omics approaches in colorectal cancer screening and diagnosis, recent updates and future perspectives. Cancers (Basel). 2022;14(22):5545. https://doi.org/10.3390/cancers14225545.
19. Baird T, Roychoudhuri R. GS-TCGA: gene set-based analysis of The Cancer Genome Atlas. J Comput Biol. 2024;31(3):229–40. https://doi.org/10.1089/cmb.2023.0278.
20. Vaziri-Moghadam A, Foroughmand-Araabi M-H. Integrating machine learning and bioinformatics approaches for identifying novel diagnostic gene biomarkers in colorectal cancer. Sci Rep. 2024. https://doi.org/10.1038/s41598-024-75438-6.
21. Dongmei A, Ming-hao W, Qingchuan Z, Longwei C, Yishu W, Xiuqin L, Li C. Xia. Regularized survival learning and cross-database analysis enabled identification of colorectal cancer prognosis-related immune genes. Frontiers in Genetics, (2023).;14 https://doi.org/10.3389/fgene.2023.1148470
22. Almutairi M, Alrubie TM, Alshareeda AT, Albarakati N, Almotiri A, Alamri AM, et al. Differential expression and regulation of ADAD1, DMRTC2, PRSS54, SYCE1, SYCP1, TEX101, TEX48, and TMPRSS12 gene profiles in colon cancer tissues and their in vitro response to epigenetic drugs. PLoS One. 2024;19:0307724. https://doi.org/10.1371/journal.pone.0307724.
23. Wang Z, Wang X, Zhang N, Zhang H, Dai Z, Zhang M, et al. Pentraxin 3 promotes glioblastoma progression by negative regulating cells autophagy. Front Cell Dev Biol. 2020;8:795. https://doi.org/10.3389/fcell.2020.00795.
24. Cai AQ, Landman KA, Hughes BD. Multi-scale modeling of a wound-healing cell migration assay. J Theor Biol. 2007;245(3):576–94. https://doi.org/10.1016/j.jtbi.2006.10.024.
25. Ding L, Yin Y, Han L, Li Y, Zhao J, Zhang W. TSC1-mTOR signaling determines the differentiation of islet cells. J Endocrinol. 2017;232(1):59–70. https://doi.org/10.1530/JOE-16-0276.
26. Luo Z, Rong Z, Zhang J, Zhu Z, Yu Z, Li T, et al. Circular RNA circCCDC9 acts as a miR-6792-3p sponge to suppress the progression of gastric cancer through regulating CAV1 expression. Mol Cancer. 2020;19(1):86. https://doi.org/10.1186/s12943-020-01203-8.
27. Zhang W, Zhao L, Zheng T, Fan L, Wang K, Li G. Comprehensive multi-omics integration uncovers mitochondrial gene signatures for prognosis and personalized therapy in lung adenocarcinoma. J Transl Med. 2024;22(1):952. https://doi.org/10.1186/s12967-024-05754-y.
28. Varpe R, Bhujbal, Kumar R. Enhancing Predictive Power in Personalized Breast Cancer Treatment through the Integration of Multi-Omics Data and Clinical Information with Deep Learning Models, 2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon), Pune, India, 2024, pp. 1–7. https://doi.org/10.1109/MITADTSoCiCon60330.2024.10575890
29. Mamand DR, Bazaz S, Mohammad DK, Saher O, Wiklander OPB, Sadeghi B, et al. Tumor cell derived osteopontin and prostaglandin E2 synergistically promote the expansion of myeloid derived suppressor cells during the tumor immune escape phase. Int Immunopharmacol. 2024;129:111584. https://doi.org/10.1016/j.intimp.2024.111584.
30. Czajka-Francuz P, Prendes MJ, Mankan A, Quintana Á, Pabla S, Ramkissoon S, et al. Mechanisms of immune modulation in the tumor microenvironment and implications for targeted therapy. Front Oncol. 2023;13:1200646. https://doi.org/10.3389/fonc.2023.1200646.
31. Zheng M. Integrated profile of tumor stage and mutational burden predicts disparate clinical responses to immune checkpoint inhibitors: a risk-benefit study. Semin Oncol. 2024;51(5–6):142–8. https://doi.org/10.1053/j.seminoncol.2024.08.003.
32. Rached L, Laparra A, Sakkal M, Danlos FX, Barlesi F, Carbonnel F, et al. Toxicity of immunotherapy combinations with chemotherapy across tumor indications: Current knowledge and practical recommendations. Cancer Treat Rev. 2024;127:102751. https://doi.org/10.1016/j.ctrv.2024.102751.
33. Feng L, Li G, Li D, Duan G, Liu J. Cuproptosis-related gene SERPINE1 is a prognostic biomarker and correlated with immune infiltrates in gastric cancer. J Cancer Res Clin Oncol. 2023;149(12):10851–65. https://doi.org/10.1007/s00432-023-04900-1.
34. Xu X, Zhang L, Qian Y, Fang Q, Xiao Y, Chen G, et al. A SERPINE1-Based immune gene signature predicts prognosis and immunotherapy response in gastric cancer. Pharmaceuticals (Basel). 2022;15(11):1401. https://doi.org/10.3390/ph15111401.
35. Mundhara N, Sadhukhan P. Cracking the codes behind cancer cells' immune evasion. Int J Mol Sci. 2024;25(16):8899. https://doi.org/10.3390/ijms25168899.
36. Li L, Li F, Xu Z, Li L, Hu H, Li Y, et al. Identification and validation of SERPINE1 as a prognostic and immunological biomarker in pan-cancer and in ccRCC. Front Pharmacol. 2023;14:1213891. https://doi.org/10.3389/fphar.2023.1213891.
37. Badodekar N, Sharma A, Patil V, Telang G, Sharma R, Patil S, et al. Angiogenesis induction in breast cancer: a paracrine paradigm. Cell Biochem Funct. 2021;39(7):860–73. https://doi.org/10.1002/cbf.3663.
38. David J. Stewart. Lung Cancer Resistance to Chemotherapy. (2010).331–93. https://doi.org/10.1007/978-1-60761-524-8_15
39. Jun X, Kai L, Feng. Yan. CASC15 mediated upregulation of SERPINE1 by sponging miR-30c- 5p correlates with poor prognosis and immune infiltration of gastric cancer. (2022). https://doi.org/10.21203/rs.3.rs-1251004/v1
40. Wang J, Wang J, Gao J, Ding M, Li H. The expression of SERPINE1 in colon cancer and its regulatory network and prognostic value. BMC Gastroenterol. 2023. https://doi.org/10.1186/s12876-022-02625-y.
41. Mazzoccoli G, Pazienza V, Panza A, Valvano MR, Benegiamo G, Vinciguerra M, et al. ARNTL2 and SERPINE1: potential biomarkers for tumor aggressiveness in colorectal cancer. J Cancer Res Clin Oncol. 2012;138(3):501–11. https://doi.org/10.1007/s00432-011-1126-6.
42. Winkler J, Abisoye-Ogunniyan A, Metcalf KJ, Werb Z. Concepts of extracellular matrix remodelling in tumour progression and metastasis. Nat Commun. 2020;11(1):5120. https://doi.org/10.1038/S41467-020-18794-X.
43. Polo-Generelo S, Rodriguez-Mateo C, Torres B, Pintor-Tortolero J, Guerrero-Martínez JA, König J, et al. Serpine1 mRNA confers mesenchymal characteristics to the cell and promotes CD8 + T cells exclusion from colon adenocarcinomas. Cell Death Discov. 2024. https://doi.org/10.1038/s41420-024-01886-8.

44. Xu DM, Chen LX, Xue T, Zhuang XY, Wei LC, Han H, et al. Decoding the impact of MMP1 + malignant subsets on tumor-immune interactions: insights from single-cell and spatial transcriptomics. Cell Death Discov. 2025;11(1):244. https://doi.org/10.1038/s41420-025-02503-y.
45. Wang Q, Feng W, Tan Y, Qiao J, Liu H, Liu Q, Wang H, Zhang Q, Tao J, Li Z, Yang B, Xu Z, Wang C, Yang S, Cui L. Antiphospholipid antibodies inhibit the migration and invasion of trophoblast cells by suppressing the JNK/C-Jun/MMP1 signaling pathway. J Transl Med. 2025;23(1):581. https://doi.org/10.1186/s12967-025-06596-y.
46. Tian S, Chen X, Li J. Zinc finger transcription factor ZNF24 inhibits colorectal cancer growth and metastasis by suppressing MMP2 transcription. Genes Dis. 2025;12(5):101529. https://doi.org/10.1016/j.gendis.2025.101529.
47. Xing P, Liu H, Xiao W, Zhang G, Zhang C, Liao Z, Li T, Yang J. The fusion gene LRP1-SNRNP25 drives invasion and migration by activating the pJNK/37LRP/MMP2 signaling pathway in osteosarcoma. Cell Death Discov. 2024;10(1):198. https://doi.org/10.1038/s41420-024-01962-z.
48. Wang Z, Sun H, Zhu S, Wang F, Li Q, Zhou J. RGS3 acts as a tumor promoter by facilitating the regulation of the TGF-β signaling pathway and promoting EMT in ovarian cancer. Cell Death Discov. 2025;11(1):262. https://doi.org/10.1038/s41420-025-02536-3.
49. Liu C, Wang S, Zhang X, Han Y, Tan M, Fan J, et al. The biomechanical signature of tumor invasion. Genes Dis. 2025;13(1):101771. https://doi.org/10.1016/j.gendis.2025.101771.
50. Hao Z, Zhang M, Du Y, Liu J, Zeng G, Li H, Peng X. Invadopodia in cancer metastasis: dynamics, regulation, and targeted therapies. J Transl Med. 2025;23(1):548. https://doi.org/10.1186/s12967-025-06526-y.
51. Li Y, Wang H, Mao D, Che X, Chen Y, Liu Y. Understanding pre-metastatic niche formation: implications for colorectal cancer liver metastasis. J Transl Med. 2025;23(1):340. https://doi.org/10.1186/s12967-025-06328-2.
52. Chen MM, Gao Q, Ning H, Chen K, Gao Y, Yu M, Liu CQ, Zhou W, Pan J, Wei L, Dou W, Zhang D, Zhu L, Zhang Q, Chen R, Zhang Z. Integrated single-cell and spatial transcriptomics uncover distinct cellular subtypes involved in neural invasion in pancreatic cancer. Cancer Cell. 2025;43(9):1656–e167610. https://doi.org/10.1016/j.ccell.2025.06.020.
53. Ito I, Hanyu A, Wayama M, Goto N, Katsuno Y, Kawasaki S, et al. Estrogen inhibits transforming growth factor beta signaling by promoting Smad2/3 degradation. J Biol Chem. 2010;285(19):14747–55. https://doi.org/10.1074/jbc.M109.093039.

Publisher's note