Bioinformatics analysis of proteins in the complement and coagulation cascades in colon cancer: Discovering the potential biomarker SERPINA1

DA LIAN[1], DAIXIN HOU[1], KUNPENG LIU[1], HAO JIANG[1] and PENGYU XIE[2]

[1]Department of Clinical Medicine, School of Clinical Medicine, Jiamusi University, Jiamusi, Heilongjiang 154007, P.R. China; [2]Department of Biology, School of Basic Medicine, Jiamusi University, Jiamusi, Heilongjiang 154007, P.R. China

Received January 21, 2026; Accepted May 22, 2026

DOI: 10.3892/ol.2026.15707

Abstract. The complement system, a core component of innate immunity, plays a key role in the tumorigenesis and progression of colon cancer. Although dysregulation in proteins associated with the complement and coagulation cascades (CCC) pathway has been identified in colon cancer, comprehensive investigations in this area are still scarce. In the present study, differential expression analysis was performed on colon adenocarcinoma (COAD) obtained from the GEPIA2 database (www.gepia2.cancer-pku.cn), to compare gene expression profiles between tumor tissues and paired adjacent normal tissues. A threshold of $|\log_2$ fold change$|>1$ and a q-value cutoff of 0.01 were applied. A total of 88 genes from the CCC signaling pathway were cross-referenced with the differentially expressed genes (DEGs) identified in the GEPIA2 analysis. These extracted DEGs were further investigated through hub gene analysis, survival analysis and druggability assessment. The expression of one selected DEG, serpin family A member 1 (*SERPINA1*), was subsequently validated using colon tissue microarrays. Among the 88 proteins in the CCC pathway, differential expression analysis identified 19 downregulated [*SERPING1*, factor VIII (*F8*) and complement C3 (*C3*)] and 13 upregulated [including *SERPINA1* and coagulation factor XII] candidates in COAD. Survival analysis demonstrated that four of these genes (*C3*, *F8*, *SERPINA1* and *SERPING1*), were notably associated with patient survival rates. Immunohistochemical validation confirmed the upregulation of *SERPINA1*, and this elevated expression was associated with shorter survival in patients with colon cancer. The present study demonstrates widespread dysregulation of proteins in the CCC pathway in colon cancer. *SERPINA1* emerges as a promising diagnostic biomarker and potential therapeutic target. Investigating *SERPINA1* therefore offers valuable insights into colon cancer pathogenesis and guides the pursuit of novel therapeutic approaches.

Introduction

Colon cancer represents the second leading cause of global cancer-related mortality, imposing a notable economic and healthcare burden (1). In China, it ranks among the top five most prevalent types of cancer, alongside lung, liver, thyroid and gastric cancers (2,3). Worldwide, the incidence rate increased from 3.96 (95% UI: 3.69-4.21) per 100,000 individuals in 1990 to 5.37 (95% UI: 4.91-5.86) per 100,000 in 2021, while the mortality rate decreased from 2.19 (95% UI: 2.01-2.36) per 100,000 in 1990 to 2.01 (95% UI: 1.84-2.19) per 100,000 in 2021 (4), affecting not only the elderly but also adolescents and young adults (5,6). The absence of early symptoms and specific biomarkers often leads to diagnosis at advanced stages (7,8). Despite advances in treatment modalities such as surgery and targeted therapies, the prognosis for colorectal cancer (CRC) remains unfavorable, with a 5-year mortality rate of ~12.8% (9). Metastatic disease confers a worse prognosis, with 5-year survival rates falling <50% even with surgical intervention (10). The pathogenesis of colon cancer is multifactorial, involving genetic predisposition, environmental factors and chronic inflammatory conditions (11). Emerging evidence further indicates that immune dysregulation plays a key role in driving carcinogenesis (12,13).

Correspondence to: Dr Pengyu Xie, Department of Biology, School of Basic Medicine, Jiamusi University, 258 Xuefu Street, Jiamusi, Heilongjiang 154007, P.R. China
E-mail: xiepengyu@jmsu.edu.cn

Abbreviations: CCC pathway, complement and coagulation cascades pathway; DEGs, differentially expressed genes; COAD, colon cancer; KEGG, Kyoto Encyclopedia of Genes and Genomes; FDR, false discovery rate; CI, confidence interval; MCODE, Molecular Complex Detection; FC, fold change; PPI, protein-protein interaction; SERPINA1, serpin family A member 1; SERPING1, serpin family G member 1; F8, factor VIII; C3, complement C3; IHC, immunohistochemistry; CRC, colorectal cancer; CTRP, Cancer Therapeutics Response Portal; ACC, adrenocortical carcinoma; LGG, brain lower grade glioma; SKCM, skin cutaneous melanoma; KIRC, kidney renal clear cell carcinoma; BRCA, breast invasive carcinoma; AAT, anti-α-1 antitrypsin; AJCC, American Joint Committee on Cancer

Key words: colon cancer, complement and coagulation cascades, bioinformatics, serpin family A member 1

Dysregulation of the complement and coagulation cascade (CCC) pathway has been implicated in various types of cancer (14-17). Activation of this pathway promotes a procoagulant state, which can compromise vascular integrity and enhance leukocyte infiltration (18). Furthermore, specific CCC genes contribute to complement-mediated endothelial damage (19), a process involved in tumorigenesis. Despite its established relevance in pre-cancers, the role of the CCC pathway in colon cancer remains insufficiently explored (17).

To systematically investigate the role of CCC pathway proteins in colon cancer, a set of 88 genes associated with this pathway was first retrieved from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Their expression patterns were subsequently analyzed using the GEPIA2 cancer database through the 'Expression DIY' and 'Survival Analysis' modules. This approach led to the identification of four key genes: Complement C3 (*C3*), factor VIII (*F8*), serpin family A member 1 (*SERPINA1*) and *SERPING1*, which are implicated in colon cancer progression. Among these, *SERPINA1* was chosen for subsequent validation. The expression and clinical relevance of *SERPINA1* were further assessed through a meta-analysis and immunohistochemistry (IHC) performed on tissue microarrays. Additionally, the relationship between *SERPINA1* expression levels and overall patient survival was evaluated using available microarray datasets.

Materials and methods

Ethics approval. The present study received approval from Shanghai Biotechnology Co., Ltd. (approval no. SHYJS-CP-1707004) and adhered to the ethical standards outlined in the 2013 Declaration of Helsinki. Written informed consent was obtained from every participant included in the present study. From July 2006 to March 2007, tumor samples, along with matched normal colon tissue from adjacent sites, were collected from residual clinical material remaining after routine pathological assessment. Relevant clinical and pathological parameters, including demographic details, were retrospectively gathered from the institution's electronic health record system.

Screening of differentially expressed genes (DEGs) within the CCC pathway. To identify DEGs related to the CCC pathway in colon cancer, a two-step screening approach was employed. First, the gene list defining the CCC pathway was obtained from the KEGG database (http://www.genome.jp/kegg/mapper.html). Subsequently, the gene expression data for colon adenocarcinoma (COAD) were acquired from the GEPIA2 portal (http://gepia2.cancer-pku.cn/) to analyze gene expression differences. DEGs were identified using thresholds of absolute $\log_2$ fold change ($|\log_2 FC|>1$ and an adjusted q-value<0.01. Finally, the overlap between these COAD-associated DEGs and the 88 CCC pathway genes was determined using R software (version 4.4.2; Posit Software, PBC).

Bioinformatics analysis for functional enrichment and hub gene identification. To investigate the biological relevance of the identified CCC-related DEGs, a protein-protein interaction (PPI) network was constructed using the Search Tool for the Retrieval of Interacting Genes and proteins (STRING) database (version 12.0; https://www.string-db.org/). Interaction confidence was set at a minimum score of 0.7 and statistical significance was defined by a false discovery rate <0.05. The resulting network was visualized and analyzed in Cytoscape (v3.10.0) (20). Key functional submodules within the PPI network were detected using the Molecular Complex Detection (MCODE) plugin (21), with parameters set as follows: Degree cutoff=2, node score cutoff=0.2, k-core=2, maximum depth=100 and a minimum module size of 4 genes. Hub genes were subsequently identified via the cytoHubba plugin (22), applying the maximal clique centrality algorithm to rank nodes by their topological importance within the network.

Analysis of the relationship between gene expression and patient survival. The association between gene expression and patient survival was analyzed using the GEPIA2 online database (http://gepia2.cancer-pku.cn/#survival). Overall survival data for all 32 DEGs in COAD were assessed using both median (50% cutoff) or quartile (75% cutoff-High) options, with a 95% confidence interval (CI) and the hazard ratio (HR) applied. Following plot generation, the log-rank P-value and HR (high) were extracted from the image. After comparing the median and quartile group cutoffs, the plots showing genes with lower log-rank P-values were selected for further study. Additionally, the relationship between four specific genes (*SERPINA1*, *C3*, *F8* and *SERPING1*) and patient survival was examined across 31 other tumor types available in the GEPIA2 database.

Potential drug target analyses. The therapeutic potential of *SERPINA1*, *C3*, *F8* and *SERPING1* was evaluated for druggability using the Open Targets Platform (23), which provided data on associated diseases and known targeting compounds. Further investigation into *SERPINA1* was conducted via the Cancer Therapeutics Response Portal (CTRP) (https://portals.broadinstitute.org), where candidate compounds were prioritized based on the correlation between their sensitivity profiles and *SERPINA1* expression levels, applying an interquartile multiplier cutoff.

Meta-analysis. *SERPINA1* expression data were obtained from articles indexed in PubMed (https://pubmed.ncbi.nlm.nih.gov/) using the keywords '*SERPINA1*' and 'cancer'. The inclusion criteria were as follows: i) Research conducted on adult human samples; ii) comparison between cancer patients and healthy controls; and iii) primary data clearly presenting differential groups, with DEGs that could be extracted or recalculated. Studies focusing solely on the effects of specific treatments in cancer patients or only including samples from treated patients were excluded. Additionally, preprints, reanalyses of data from databases, and reviews were excluded. Retrieved articles were managed using Endnote (version 21). Two reviewers (DL and DXH) independently extracted the data following PRISMA guidelines (24), and discrepancies were resolved through discussion. For the meta-analysis of *SERPINA1* expression data, a random-effects model was implemented in Stata 14.0 (StataCorp LP). The key metrics extracted per study were the $\log_2 FC$, reflecting the cancer vs. control expression difference and the negative log P-value (-logP) as a measure of statistical

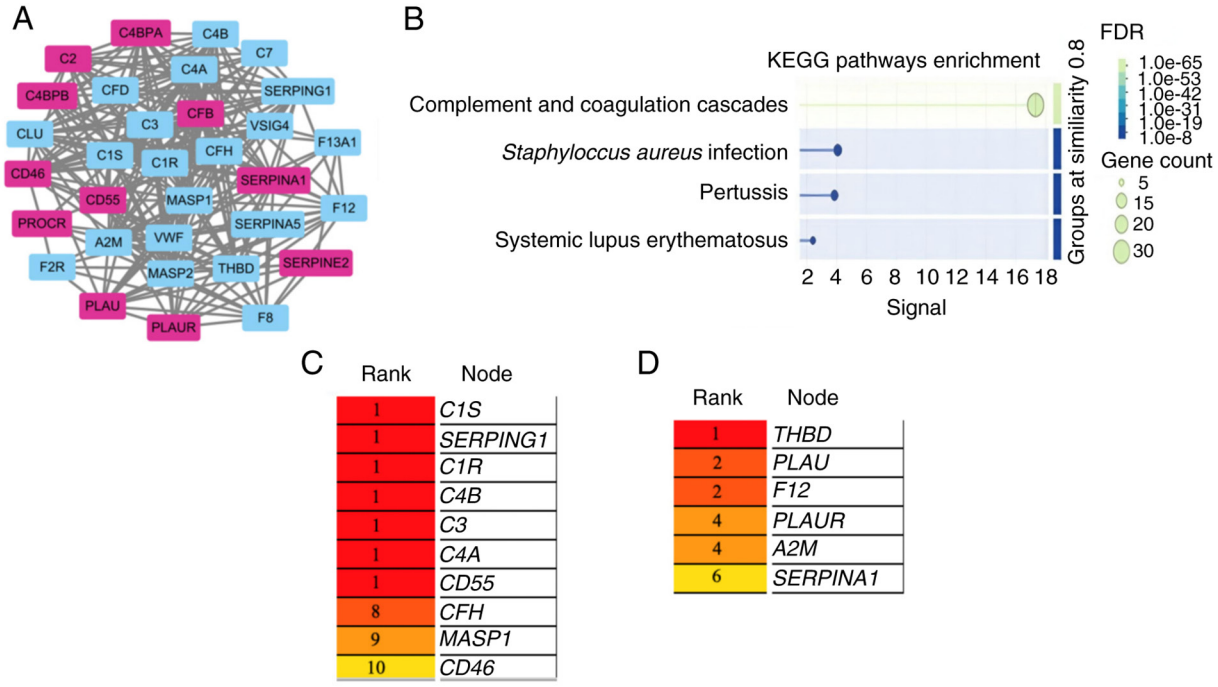

Figure 1. PPI and functional analysis of differentially expressed genes in the complement and coagulation cascades pathway. (A) PPI network visualized using Cytoscape, with up- and down-regulated genes indicated in red and blue, respectively. (B) KEGG pathway enrichment results from the STRING database. Top hub genes identified within (C) Module 1 and (D) Module 2, respectively, by applying the cytoHubba algorithm to the MCODE-derived network clusters. KEGG, Kyoto Encyclopedia of Genes and Genomes; FDR, false discovery rate; PPI, protein-protein interaction.

strength. The overall effect size was calculated as a weighted average of the $\log_2$FC values, with the weight of each study determined by the-logP-value to emphasize more reliable findings. The random-effects model was used for weight assessment. The 95% CIs for the pooled estimate were derived from this weighting scheme. Results were visualized as forest plots and the between-study heterogeneity was evaluated using the I^2 statistic. Significance was established if the 95% CI of the combined estimate did not include zero.

IHC validation of SERPINA1 expression using tissue microarray. IHC analysis was performed using a colon cancer tissue microarray (cat. no. HCo1A180Su16; Shanghai Outdo Biotech Co., Ltd.), which included 180 samples [tumor (n=104) and adjacent normal tissues (n=76)] from 104 patients: 6 stage I-II, 1 stage I-III, 51 stage II, 33 stage II-III and 13 stage III (according to the 7th edition of the AJCC TNM staging system (25). The patient cohort consisted of 59 men and 45 women, with a mean age of 68.2±10.8 years (range, 24-90 years). Samples had been collected between July 2006 and May 2007. By July 2015, 41 patients remained alive, while 63 were deceased. Tissue sections were incubated with an undiluted rabbit anti-α-1 antitrypsin (AAT; *SERPINA1*) primary antibody (cat. no. ZA-0007; Beijing Zhongshan Jinqiao Biotechnology Co., Ltd.) followed by goat anti-rabbit IgG H&L (HRP) diluted 1:1,000 (cat. no. ab6721; Abcam). The sections were then stained by diaminobenzidine (DAB). Staining and slide imaging were automated using the BOND RX Research Stainer (Leica Biosystems) and the TissueFAXS 200 system (TissueFAXS Viewer software v6.0.6245.146; TissueGnostics GmbH), respectively. *SERPINA1* expression was quantified via optical density measurement with HistoQuest software (v6.0.1.114; TissueGnostics GmbH). Prior to quantification, the analysis pipeline was optimized by refining three key parameters: Total tissue area segmentation, positive cell detection and background signal correction. The final expression level was calculated as the percentage of positive cells within a total analyzed tissue volume of 3.16 mm^3.

Statistical analysis. Statistical analyses and graphical presentations were performed with GraphPad Prism (v9.0.0; Dotmatics) or R Studio (v4.4.2). Values are expressed as the mean ± standard deviation. Group comparisons were made by Student's unpaired t-test. $P<0.05$ was considered to indicate a statistically significant difference.

Results

Identification and network analysis of complement and coagulation cascade (CCC)-related DEGs in COAD. A total of 32 genes from the 88-gene CCC pathway were identified as DEGs ($|\log_2 FC|>1$ and q-value cutoff of 0.01) in COAD based on the GEPIA2 database. Among these, 13 genes were upregulated ($P<0.05$) and 19 were downregulated ($P<0.05$) (Fig. 1A and Table I), with broad representation across the intrinsic, classical, lectin and alternative complement pathways (Fig. S1). Enrichment analysis performed in the STRING database indicated that the 'CCC pathway' was the most significantly overrepresented KEGG term, followed by pathways linked to *Staphylococcus aureus* infection, pertussis and systemic lupus erythematosus (Fig. 1B). PPI interaction analysis revealed a densely interconnected network. Further clustering of this PPI network via the MCODE plugin in Cytoscape delineated two functional modules. Module 1, comprising 17 nodes and

Table I. Differentially expressed genes in colon adenocarcinoma from GEPIA2 database.

Genes	Median (tumor)	Median (normal)	log_2FC	P-value
C7	1.67	111.39	-5.40	$6.80x10^{-112}$
MASP1	0.30	23.52	-4.24	$7.16x10^{-73}$
CFD	9.65	162.01	-3.94	$6.45x10^{-97}$
CLU	56.28	744.12	-3.70	$3.50x10^{-65}$
F13A1	1.65	17.67	-2.82	$2.72x10^{-69}$
VSIG4	4.15	22.13	-2.17	$2.24x10^{-41}$
SERPING1	67.62	261.89	-1.94	$2.52x10^{-36}$
C1S	70.60	241.45	-1.76	$5.43x10^{-35}$
C3	68.83	234.82	-1.76	$4.95x10^{-21}$
C4A	18.13	59.54	-1.66	$2.46x10^{-26}$
C4B	18.15	59.09	-1.65	$1.28x10^{-25}$
CFH	6.62	19.90	-1.46	$1.91x10^{-35}$
SERPINA5	0.34	2.64	-1.44	$8.93x10^{-60}$
MASP2	0.24	1.92	-1.24	$3.64x10^{-97}$
A2M	58.60	139.21	-1.23	$1.56x10^{-28}$
C1R	77.29	176.88	-1.18	$5.10x10^{-23}$
THBD	3.53	9.01	-1.14	$1.64x10^{-27}$
F8	1.72	4.88	-1.11	$3.70x10^{-25}$
VWF	16.77	36.48	-1.08	$2.30x10^{-23}$
C4BPA	1.78	0.18	1.24	$5.55x10^{-27}$
CD46	174.82	71.39	1.28	$2.42x10^{-78}$
F2R	12.27	4.27	1.33	$1.45x10^{-47}$
PROCR	35.05	12.79	1.39	$2.04x10^{-35}$
CD55	65.26	22.58	1.49	$9.23x10^{-42}$
CFB	69.48	21.30	1.66	$3.87x10^{-45}$
C4BPB	7.37	1.57	1.70	$7.87x10^{-27}$
PLAU	29.05	6.56	1.99	$5.20x10^{-68}$
SERPINE2	47.67	10.86	2.04	$5.04x10^{-41}$
PLAUR	59.00	11.74	2.24	$1.58x10^{-66}$
C2	57.80	9.82	2.44	$1.14x10^{-86}$
SERPINA1	120.77	17.31	2.73	$6.53x10^{-33}$
F12	18.78	1.94	2.75	$3.72x10^{-96}$

log_2FC, log2 fold change.

121 edges, included complement C1s, *SERPING1* and *C3*. Module 2, consisting of 6 nodes and 12 edges, encompassed the hub genes thrombomodulin and *SERPINA1* (Fig. 1C and D).

Four genes related to the survival of patients with pan-cancers. Overall survival analysis using a median group cutoff demonstrated an association between four genes from the CCC pathway (*SERPINA1*, *C3*, *F8* and *SERPING1*) and patient survival in COAD (Fig. 2). This association remained significant for *SERPINA1* alone when a more stringent quartile cutoff with Cutoff-high of 75% was applied. Extending the analysis to other types of cancer revealed broader prognostic relevance. *SERPINA1* expression was associated with patient survival in brain lower grade glioma (LGG), skin cutaneous melanoma (SKCM) and breast invasive carcinoma. The other three genes also showed significant associations across multiple types of cancer, with *C3* linked to adrenocortical carcinoma, COAD, LGG and SKCM; *F8* to COAD and kidney renal clear cell carcinoma; and *SERPING1* to COAD, LGG and SKCM (Table II).

Evaluating the druggability profile of SERPINA1, C3, F8 and SERPING1. To investigate the potential druggability of the four identified genes (*SERPINA1*, *C3*, *F8* and *SERPING1*), established disease associations were analyzed using the Open Targets Platform (Table II). All genes displayed associations with various types of cancer, with strength of evidence varying. Gene-cancer association scores were obtained from the GeneCards Suite, which integrates multi-omics and clinical evidence to generate a normalized ALIscore (0-1) for each gene-disease pair, with higher scores indicating stronger association. For instance, *SERPINA1* showed strong associations with gastric and breast cancer, whereas the link between *SERPING1* and types of cancer, such as breast cancer and prostate cancer, was comparatively weak. Subsequent review of

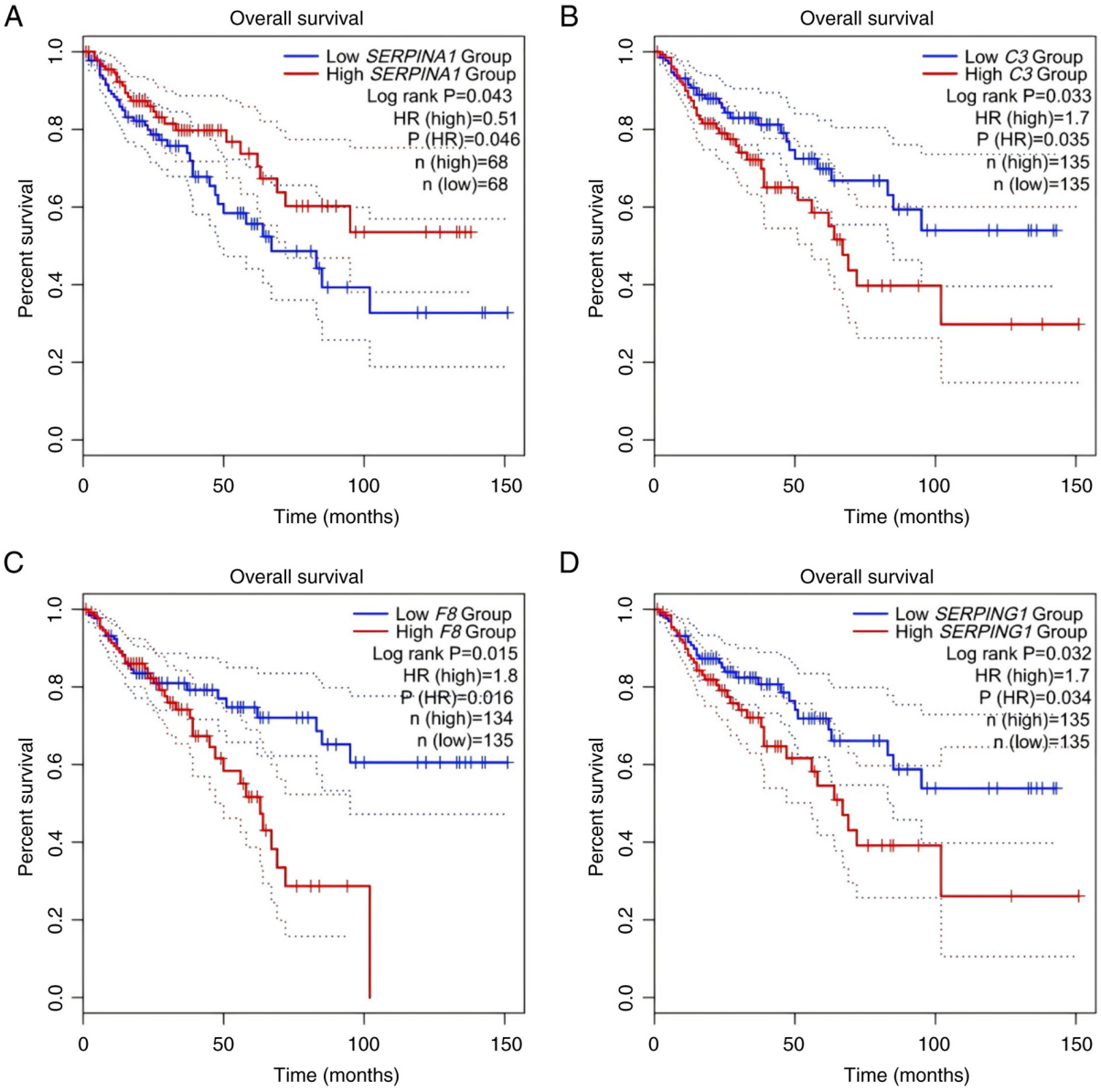

Figure 2. Association of four genes with overall survival in colon cancer. (A) Survival analysis for *SERPINA1* using quartile cutoff. Survival analyses for (B) *C3*, (C) *F8* and (D) *SERPING1*, respectively, using median cutoff. *SERPINA*1, serpin family A member 1; *C3*, complement C3; *F8*, factor VIII; HR, hazard ratio.

approved targeted therapies revealed that *F8*, *C3* and *SERPING1* are already targeted by existing drugs (Table III) (26-28). Notably, however, no clinically approved drugs currently target *SERPINA1*, suggesting its potential as a novel therapeutic candidate (https://www.genecards.org/card/SERPINA1).

In the absence of approved drugs targeting *SERPINA1*, the CTRP database was used to identify potential compounds interacting with this gene. As presented in Fig. 3, five compounds showed a correlation with *SERPINA1* expression, listed in order of decreasing correlation score: Omacetaxine mepesuccinate (0.2330), ML030 (0.2140), vincristine (0.2090), SB-743921 (0.2070) and KX2-391 (0.2020). These agents may represent potential therapeutic candidates for colon cancer via modulation of *SERPINA1* activity.

Meta-analysis of SERPINA1 expression. Based on multiple criteria, including prognostic relevance, hub gene status and the absence of approved targeted therapies, *SERPINA1* was prioritized for further investigation. Following a literature review in PubMed (https://pubmed.ncbi.nlm.nih.gov/) using the keywords 'cancer' and '*SERPINA1*', four eligible studies (29-32) were included (Table SI). As shown in Fig. 4, the meta-analysis demonstrated that *SERPINA1* was consistently upregulated in cancer tissues compared with healthy controls, with a pooled $\log_2$FC of 1.58 (95% CI: 0.56, 2.60). However, high heterogeneity was observed across the included studies (I^2=85.6%; P<0.001). This may be attributed to the limited reference data (derived from only four articles) and inherent data heterogeneity. In the present study, the data came from two types of cancer: Colon cancer (n=3) and pancreatic ductal adenocarcinoma (n=1). In addition, different sample types were involved, including tissue (n=2), plasma exosomes (n=1) and serum (n=1).

SERPINA1 validation by IHC analysis. To validate *SERPINA1* expression at the protein level, IHC analysis was performed. Although previous data from the GEPIA2

Table II. Survival analysis of four genes in pan-cancer panel from GEPIA2 database.

	Type of cancer													
Genes	COAD	LGG	SKCM	BRCA	ACC	CHOD	KICH	KRIC	MESO	PCPG	CESC	THYM	LUSC	SARC
SERPINA1														
Log rank P-value	0.043	$2.7x10^{-6}$	0.00026	$9.3x10^{-12}$	-	-	-	-	-	-	-	-	-	-
HR (high)	0.51	2.4	0.61	0.61	-	-	-	-	-	-	-	-	-	-
C3														
Log rank P-value	0.033	0.0031	0.017	-	$9.5x10^{-5}$	0.033	0.032	$4.1x10^{-5}$	0.0025	0.013	-	-	-	-
HR (high)	1.7	1.7	0.72	-	0.19	1.7	0.14	1.9	0.47	$1.6x10^{-9}$	-	-	-	-
F8														
Log rank P-value	0.015	-	-	-	0.019	-	-	0.0003	-	-	0.0026	0.036	-	-
HR (high)	1.8	-	-	-	0.39	-	-	0.57	-	-	0.49	6.9	-	-
SERPING1														
Log rank P-value	0.032	$3.3x10^{-7}$	0.00052	-	-	-	-	-	0.0003	-	-	-	0.037	$3.9x10^{-5}$
HR (high)	1.7	2.6	0.62	-	-	-	-	-	0.42	-	-	-	1.3	0.43

COAD, colon adenocarcinoma; LGG, brain lower grade glioma; SKCM, skin cutaneous melanoma; BRCA, breast invasive carcinoma; ACC, adrenocortical carcinoma; CHOL, cholangiocarcinoma; KICH, kidney chromophobe; KIRC, kidney renal clear cell carcinoma; MESO, mesothelioma; PCPG, pheochromocytoma and paraganglioma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; THYM, thymoma; LUSC, lung squamous cell carcinoma; SARC, sarcoma; HR, hazard ratio; *SERPINA1*, serpin family A member 1; *C3*, complement C3; *F8*, factor VIII.

Table III. Summary of disease associations and approved or investigational drugs targeting the four genes.

Genes	Type of cancer (score[a])	Known drugs	Main treatment	Earliest approved	(Refs.)
SERPINA1	Gastric cancer (0.82); Pancancer (0.81) Breast cancer (0.75); Colorectal cancer (0.59)	No	/	/	N/A
C3	Ovarian cancer (0.83); Pancancer (0.81); Gastric cancer (0.72); Colorectal cancer (0.36)	Pegcetacoplan	Paroxysmal nocturnal hemoglobinuria and immune system disease	2021	(26)
F8	Pancancer (0.67); Esophageal cancer (0.39); Breast cancer (0.20); Lung cancer (0.14)	Moroctocog alfa	Hemophilia a	1999	(27)
SERPING1	Breast cancer (0.24); Pancancer (0.18); Prostate cancer (0.15); Pituitary cancer (0.13)	Conestat alfa	Hereditary angioedema	2010	(28)

[a]Association scores. A score of 1.0 corresponds to the highest level of relevance. N/A, not applicable; *SERPINA1*, serpin family A member 1; *F8*, factor VIII; *C3*, complement C3.

database and meta-analysis had indicated upregulation of *SERPINA1* in colon cancer, direct protein validation using established methods such as IHC, western blot or ELISA remained necessary. To address this, IHC was conducted on a tissue microarray containing 180 spots of colon tumor and matched adjacent normal tissues. Anti-AAT staining demonstrated markedly stronger AAT accumulation in tumor tissues (Fig. 5A) compared with adjacent normal tissues (Fig. 5B). Quantitative optical density analysis further demonstrated that AAT was upregulated by 1.84-fold in tumors ($P<0.0001$; Fig. 5C). When comparing AAT expression between surviving and deceased patient groups, higher AAT levels ($P<0.05$) were observed in the deceased group (Fig. 5D). Consistently, survival analysis revealed significantly shorter overall survival in patients with higher *SERPINA1* expression compared with those with lower expression (Fig. 5E).

Discussion

CRC is a highly prevalent malignancy of the digestive system, whose development is influenced by a combination of genetic predisposition, environmental factors, chronic inflammation and gut microbiota dysbiosis. Prior research has highlighted the involvement of the complement system, a central element of innate immunity, in CRC pathogenesis (33,34), as it contributes to host defense against pathogens and modulates intestinal inflammatory responses during cancer progression. Furthermore, complement signaling has been implicated across multiple stages of tumorigenesis, including initiation, proliferation, metastasis and response to therapy (34-36). Dysregulation of the CCC pathway has been documented in several malignancies, such as metastatic urothelial carcinoma (14), acute lymphoblastic leukemia (37), lower-grade glioma (15) and bladder cancer (38). Nevertheless, the role of the CCC pathway in CRC remains largely unexplored. In the present study, the expression profiles of 88 CCC genes in COAD were evaluated using the GEPIA2 database and 32 differentially expressed CCC genes (13 upregulated and 19 downregulated) were identified, suggesting that widespread dysregulation may be implicated in clinical outcomes for patients with CRC.

Following PPI and survival analyses, four hub genes associated with patient survival were identified: *SERPINA1* (upregulated), *C3* (downregulated), *F8* (downregulated) and

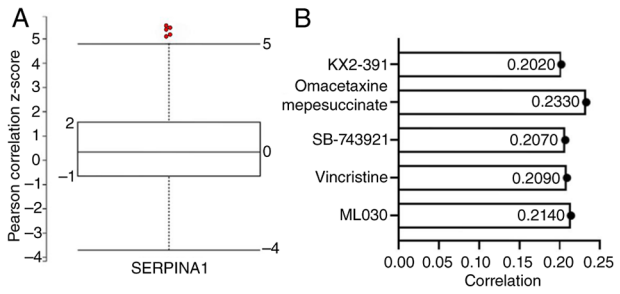

Figure 3. Correlation analysis of small molecules with *SERPINA1* expression using Cancer Therapeutics Response Portal data. (A) Distribution of Pearson correlation coefficients presented as a box plot (interquartile multiplier method). (B) Bar chart representation of the correlation scores visualized using GraphPad Prism 9 software. *SERPINA1*, serpin family A member 1.

SERPING1 (downregulated). *SERPINA1* is a serine protease inhibitor primarily targeting elastase, and is capable of irreversibly inhibiting trypsin, chymotrypsin and plasminogen activator. The aberrant isoform of *SERPINA1* inhibits insulin-induced nitric oxide synthesis in platelets, shortens coagulation time and exhibits proteolytic activity against insulin and plasmin (39). *SERPINA1* has been implicated in colon cancer (29-32) and in combination with fibrinogen demonstrates superior diagnostic efficacy compared with conventional markers such as carcinoembryonic antigen and carbohydrate antigen 19-9 (30). In CRC, *SERPINA1* is highly expressed, is associated with unfavorable clinical outcomes and promotes CRC cell proliferation and migration via activation of the STAT3 pathway (31). *C3* serves as a precursor for non-enzymatic constituents of the classical, alternative, lectin and granzyme K complement pathways. These pathways comprise a proteolytic cascade that drives pathogen phagocytosis and degradation while enhancing adaptive immune signaling. *C3* deficiency exacerbates inflammatory responses in the colon (40). Although downregulated in colon cancer, elevated *C3* expression is associated with worse overall survival in gastric cancer (34). *F8*, in the presence of calcium and phospholipids, functions as a cofactor for factor IXa during the conversion of factor X to its active form, factor Xa. Deficiency in coagulation factor VIII underlies hemophilia A, for which recombinant or plasma-derived factor VIII remains first-line therapy (41). *SERPING1*, another serine protease inhibitor, regulates the classical complement pathway. *SERPING1* is downregulated in prostate cancer and reduced *SERPING1* expression is associated with higher Gleason scores, advanced pathological grade and more progressive tumor stages (42). In summary, *SERPINA1*, *C3*, *F8* and *SERPING1* are four cancer-related genes likely to play notable roles in colon cancer, despite the limited number of studies specifically focused on their functions in this malignancy.

The druggability of *SERPINA1*, *C3*, *F8* and *SERPING1* was further evaluated based on reports from the Open Targets Platform (23). Among these, only *SERPINA1* has not yet been established as a known drug target, to the best of our knowledge. However, based on predictive screening using the CTRP database, omacetaxine mepesuccinate was identified as a potential compound targeting *SERPINA1*. This agent is an approved anticancer drug currently used in the treatment of chronic myeloid leukemia (43,44). Taken together, these findings suggest that *SERPINA1*, *C3*, *F8* and *SERPING1* may represent promising candidate targets for therapeutic intervention.

SERPINA1 was selected for meta-analysis and IHC validation using tissue microarrays. While prior studies have examined the expression and function of *SERPINA1* in colon cancer (30-32), the present approach distinguishes itself from these earlier works by employing different specimens and different validation methods. For instance, Li *et al* (30) and Peltier *et al* (32) employed plasma samples from patients with CRC and validated *SERPINA1* via ELISA, whereas Ma *et al* (31) utilized a mouse CRC model. The present study, based on tissue microarray analysis, thus serves as a methodological complement to existing research. Furthermore, the expression pattern of *SERPINA1* in colon cancer remains contentious. The cProSite database, which incorporates proteomic and phosphoproteomic data from the National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium and International Cancer Proteogenome Consortium (45), indicates that *SERPINA1* is downregulated in colon cancer (Fig. S2). By contrast, the GEPIA2 database, which performs gene expression analysis based on tumor and normal samples from the TCGA and GTEx databases (46), reports its upregulation. In the present study, *SERPINA1* expression was elevated in tumor tissues compared with adjacent normal colon tissues, and was higher in deceased patients compared with surviving patients. Furthermore, in the GEPIA2 database cohort, patients with higher *SERPINA1* expression exhibited longer survival. However, in the present study, increased *SERPINA1* expression was associated with shortened survival in patients with colon cancer, a finding consistent with observations in pancreatic ductal adenocarcinoma (30). This discrepancy in survival outcomes between the GEPIA2 database and the present study may be attributed to the difference between RNA-sequencing data (from GEPIA2) and protein expression data (from the present study). The post-transcriptional and post-translational modifications of SERPINA1 may lead to inconsistencies between its mRNA and protein expression levels, which also highlights the necessity of protein-level experimental validation for bioinformatics and omics findings. Additionally, a review of published literature indicated that *SERPINA1* was commonly upregulated across multiple types of cancer (30-32). Thus, the present study provides experimental evidence that helps to clarify the discrepant findings between the cProSite and GEPIA2 database cohorts.

The present study has several limitations that need to be addressed. First, the protein expression levels of the other three candidate genes (*C3*, *F8* and *SERPING1*) were not experimentally validated. Since mRNA expression levels do not always correlate with protein abundance due to post-transcriptional regulation, future studies should therefore employ immunohistochemistry or western blot analysis to confirm their protein expression in colon cancer tissues. Second, the meta-analysis was performed solely for *SERPINA1* expression. Consequently, the diagnostic or prognostic value of the other three genes (C3, F8, SERPING1) across different cohorts remains elusive. Further meta-analyses integrating multiple independent datasets are needed to evaluate their clinical relevance. Third, no experiments were conducted to validate potential compounds targeting *SERPINA1*. Thus, the druggability of SERPINA1 suggested by the Open Targets Platform is purely computational and lacks experimental

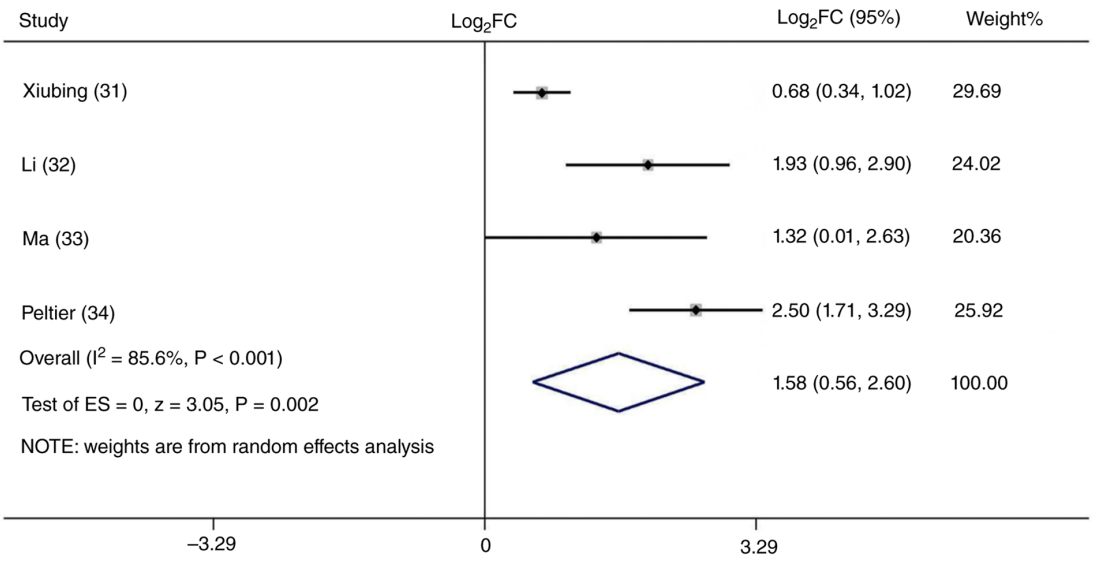

Figure 4. Forest plot of *SERPINA1* based on the Random Effects Model by meta-analysis. Individual study estimates are shown as black dots (log_2FC) with horizontal error bars representing 95% ICs. The pooled effect size across studies is indicated by a diamond marker. A dashed vertical line at log_2FC=0 corresponds to no differential expression. The plot also reports key summary statistics, including the ES (test of ES=0; z=3.05; P=0.002), the heterogeneity index (I^2=85.6%) and the P-value for heterogeneity (P<0.001), supporting the robustness of this finding while indicating high heterogeneity. log_2FC, log_2 fold change; ES, effect size; CI, confidence interval; *SERPINA1*, serpin family A member 1.

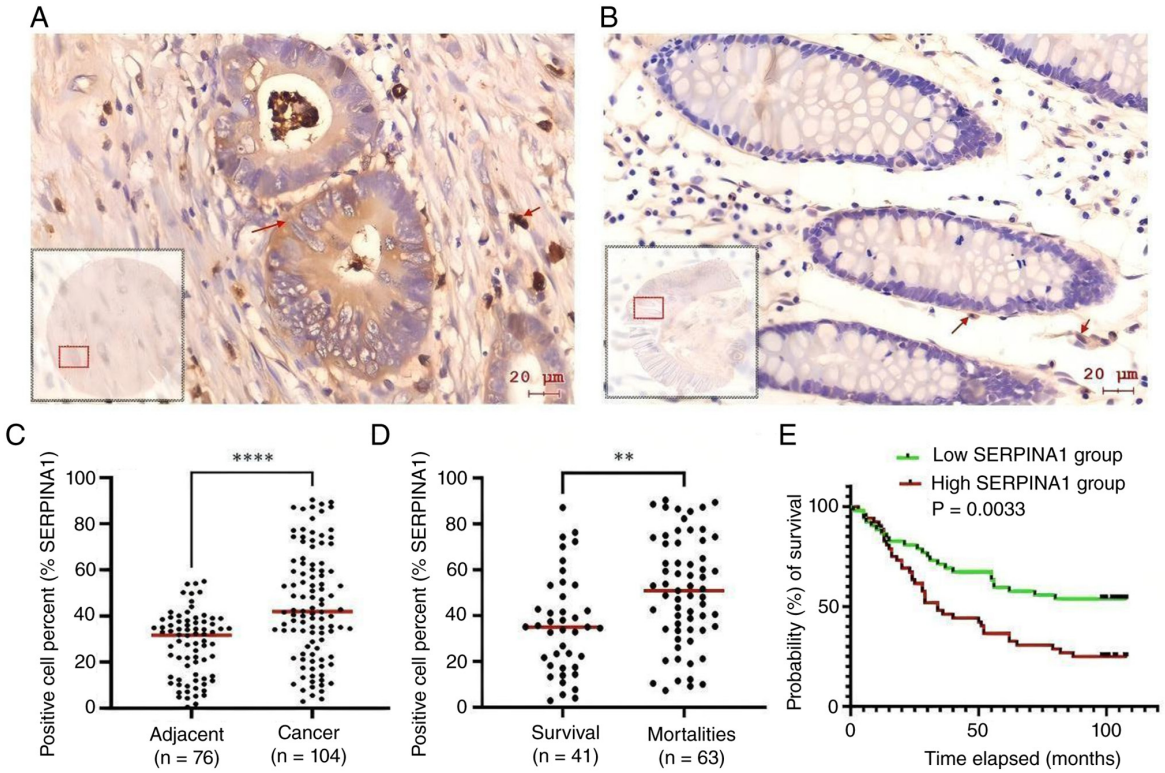

Figure 5. IHC detection of *SERPINA1* in colon tissues. (A) Representative IHC staining of *SERPINA1* in colon tumor tissue. Lower left image depicts the whole tissue, while the boxed area is enlarged and displayed on the right. Red arrows depict positively stained cells. Scale bar, 20 μm. (B) Representative IHC staining of *SERPINA1* in adjacent normal colon tissue. Lower left image shows the whole tissue, while the boxed area is enlarged and displayed on the right. Red arrows depict positively stained cells. Scale bar, 20 μm. (C) Quantitative comparison of *SERPINA1* expression between tumor and adjacent normal tissues using optical density analysis. (D) Comparison of *SERPINA1* expression between deceased and surviving patient groups based on optical density measurements. (E) Survival analysis comparing patients with high vs. low *SERPINA1* expression. **P<0.01 and ****P<0.0001. IHC, immunohistochemistry; *SERPINA1*, serpin family A member 1.

support. Future research should perform *in vitro* or *in vivo* assays (e.g., viability, apoptosis or targeted inhibition assays) to test candidate compounds. Finally, the precise functional role of *SERPINA1* in patients with colon cancer has yet to be experimentally elucidated. This is a critical limitation because without functional validation (e.g., via knockdown or

overexpression models), the observed expression differences cannot be causally linked to tumor progression or patient outcomes. Future investigations using gain- or loss-of-function approaches in colon cancer cell lines or animal models are required to define its mechanistic role.

In summary, the present study analyzed the expression of 88 genes in the CCC pathway and identified 32 DEGs. Among these, four hub genes were associated with the survival of patients with colon cancer. *SERPINA1*, an upregulated gene, was further validated in colon tissue microarrays via IHC analysis, showing that its upregulation was associated with worse survival outcomes. These findings suggest that *SERPINA1* may serve as a potential diagnostic biomarker for colon cancer and represents a promising candidate for therapeutic targeting.

Acknowledgements

Not applicable.

Funding

This research received financial support from University Student Innovation and Entrepreneurship Training Plan of Heilongjiang Province (grant no. S202510222084).

Availability of data and materials

The data generated in the present study may be requested from the corresponding author.

Authors' contributions

PX conceived and designed the study and revised the manuscript. DL collected, analyzed and interpreted the data, and wrote the manuscript. DH designed the differential expression analysis pipeline and interpreted the transcriptome data. KL performed the functional enrichment analysis and curated all public datasets used in this study. HJ conducted statistical analysis of bioinformatics data and generated all related figures. DH, KL and HJ critically revised the bioinformatics sections of the manuscript for important intellectual content. PX and DL confirm the authenticity of all the raw data. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The research complied with the principles outlined in the Declaration of Helsinki and received formal approval from Shanghai Biotechnology Co., Ltd. (approval no. SHYJS-CP-1707004). Written informed consent was secured from all participants prior to their involvement.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Use of artificial intelligence tools

During the preparation of this work, DeepSeek-V3.1 (https://chat.deepseek.com/) was used to improve the readability and language of the manuscript or to generate images, and subsequently, the authors revised and edited the content produced by the artificial intelligence tools as necessary, taking full responsibility for the ultimate content of the present manuscript.

References

1. Lukic M, Licaj I, Laaksonen MA, Weiderpass E, Borch KB and Rylander C: The burden of colon cancer attributable to modifiable factors-The Norwegian women and cancer study. Int J Cancer 152: 195-202, 2023.
2. Han B, Zheng R, Zeng H, Wang S, Sun K, Chen R, Li L, Wei W and He J: Cancer incidence and mortality in China, 2022. J Natl Cancer Cent 4: 47-53, 2024.
3. He S, Xia C, Li H, Cao M, Yang F, Yan X, Zhang S, Teng Y, Li Q and Chen W: Cancer profiles in China and comparisons with the USA: A comprehensive analysis in the incidence, mortality, survival, staging, and attribution to risk factors. Sci China Life Sci 67: 122-131, 2024.
4. Zhang J, Ou D, Xie A, Chen D and Li X: Global burden and cross-country health inequalities of early-onset colorectal cancer and its risk factors from 1990 to 2021 and its projection until 2036. BMC Public Health 24: 3124, 2024.
5. Bleyer A, Ries LAG, Cameron DB, Mansfield SA, Siegel SE and Barr RD: Colon, colorectal, and all cancer incidence increase in the young due to appendix reclassification. J Natl Cancer Inst 117: 1340-1349, 2025.
6. Aggarwal S, Lavingiya V, Krishna V, Chitalkar P, Ostwal V and Parikh PM: Young onset colorectal cancer. South Asian J Cancer 13: 225-228, 2024.
7. Lo Nigro C, Ricci V, Vivenza D, Granetto C, Fabozzi T, Miraglio E and Merlano MC: Prognostic and predictive biomarkers in metastatic colorectal cancer anti-EGFR therapy. World J Gastroenterol 22: 6944-6954, 2016.
8. Alaiyan B, Ilyayev N, Stojadinovic A, Izadjoo M, Roistacher M, Pavlov V, Tzivin V, Halle D, Pan H, Trink B, *et al*: Differential expression of colon cancer associated transcript1 (CCAT1) along the colonic adenoma-carcinoma sequence. BMC Cancer 13: 196, 2013.
9. Sahli H, Dahlbäck C, Lydrup ML and Buchwald P: Impact of previous diverticulitis on 5-year survival and recurrence rates in patients with colorectal cancer. Scand J Gastroenterol 58: 1280-1285, 2023.
10. Stewart CL, Warner S, Ito K, Raoof M, Wu GX, Kessler J, Kim JY and Fong Y: Cytoreduction for colorectal metastases: Liver, lung, peritoneum, lymph nodes, bone, brain. When does it palliate, prolong survival, and potentially cure? Curr Probl Surg 55: 330-379, 2018.
11. Jahanafrooz Z, Mosafer J, Akbari M, Hashemzaei M, Mokhtarzadeh A and Baradaran B: Colon cancer therapy by focusing on colon cancer stem cells and their tumor microenvironment. J Cell Physiol 235: 4153-4166, 2020.
12. Yang WW, Zhou X and He G: Patients with colorectal cancer combined with HIV had a worse overall survival after surgery: A meta-analysis. Front Oncol 15: 1440105, 2025.
13. Fidelle M, Yonekura S, Picard M, Cogdill A, Hollebecque A, Roberti MP and Zitvogel L: Resolving the paradox of colon cancer through the integration of genetics, immunology, and the microbiota. Front Immunol 11: 600886, 2020.
14. Gong Z, He Y, Mi X, Li C, Sun X, Wang G, Li L, Han Y, Xu C, Wang W, *et al*: Complement and coagulation cascades pathway-related signature as a predictor of immunotherapy in metastatic urothelial cancer. Aging (Albany NY) 15: 9479-9498, 2023.
15. Yang J, Shen L, Yang J, Qu Y, Gong C, Zhou F, Liu Y, Luo M and Zhao L: Complement and coagulation cascades are associated with prognosis and the immune microenvironment of lower-grade glioma. Transl Cancer Res 13: 112-136, 2024.
16. Conway EM: Reincarnation of ancient links between coagulation and complement. J Thromb Haemost 13 (Suppl 1): S121-S132, 2015.

17. Bonetto A, Aydogdu T, Kunzevitzky N, Guttridge DC, Khuri S, Koniaris LG and Zimmers TA: STAT3 activation in skeletal muscle links muscle wasting and the acute phase response in cancer cachexia. PLoS One 6: e22538, 2011.
18. Gavriilaki E, Ho VT, Schwaeble W, Dudler T, Daha M, Fujita T and Jodele S: Role of the lectin pathway of complement in hematopoietic stem cell transplantation-associated endothelial injury and thrombotic microangiopathy. Exp Hematol Oncol 10: 57, 2021.
19. Ray A, Winter KAK, Naik DSL and Okorie C: Prognostic significance of the coagulation and complement systems in critical COVID-19 infection. Prague Med Rep 124: 77-93, 2023.
20. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T: Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498-2504, 2003.
21. Bader GD and Hogue CWV: An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 4: 2, 2003.
22. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT and Lin CY: cytoHubba: Identifying hub objects and sub-networks from complex interactome. BMC Syst Biol 8 (Suppl 4): S11, 2014.
23. Ochoa D, Hercules A, Carmona M, Suveges D, Baker J, Malangone C, Lopez I, Miranda A, Cruz-Castillo C, Fumis L, *et al*: The next-generation open targets platform: Reimagined, redesigned, rebuilt. Nucleic Acids Res 51 (D1): D1353-D1359, 2023.
24. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, *et al*: The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. Rev Esp Cardiol (Engl Ed) 74: 790-799, 2021 (In English, Spanish).
25. Edge SB and Compton CC: The American joint committee on cancer: The 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol 17: 1471-1474, 2010.
26. U.S. Food and Drug Administration (FDA): FDA approves new treatment for adults with serious rare blood disease (EB/OL). (2021-05-18). FDA, Silver Spring, MD. https://www.fda.gov/drugs/news-events-human-drugs/fda-approves-new-treatment-adults-serious-rare-blood-disease
27. European Medicines Agency: ReFacto AF (moroctocog alfa) EPAR[EB/OL]. (1999-04-13). European Medicines Agency, Amsterdam, 2016. https://www.ema.europa.eu/en/medicines/human/EPAR/refacto-af
28. European Medicines Agency: Ruconest (conestat alfa) EPAR[EB/OL]. (2010 10 28). European Medicines Agency, Amsterdam, 2017. https://www.ema.europa.eu/en/medicines/human/EPAR/ruconest
29. Xiubing C, Huazhen L, Xueyan W, Jing N, Qing L, Haixing J, Shanyu Q and Jiefu L: SERPINA1 promotes the invasion, metastasis, and proliferation of pancreatic ductal adenocarcinoma via the PI3K/Akt/NF-κB pathway. Biochem Pharmacol 230: 116580, 2024.
30. Li L, Song X, Chen G, Zhang Z, Zheng B, Zhang Q, Wang S and Xie L: Plasma exosomal protein PLG and SERPINA1 in colorectal cancer diagnosis and coagulation abnormalities. J Cancer Res Clin Oncol 149: 8507-8519, 2023.
31. Ma Y, Chen Y, Zhan L, Dong Q, Wang Y, Li X, He L and Zhang J: CEBPB-mediated upregulation of SERPINA1 promotes colorectal cancer progression by enhancing STAT3 signaling. Cell Death Discov 10: 219, 2024.
32. Peltier J, Roperch JP, Audebert S, Borg JP and Camoin L: Quantitative proteomic analysis exploring progression of colorectal cancer: Modulation of the serpin family. J Proteomics 148: 139-148, 2016.
33. Ding P, Xu Y, Li L, Lv X, Li L, Chen J, Zhou D, Wang X, Wang Q, Zhang W, *et al*: Intracellular complement C5a/C5aR1 stabilizes β-catenin to promote colorectal tumorigenesis. Cell Rep 39: 110851, 2022.
34. Krieg C and Guglietta S: The complement system in intestinal inflammation and cancer. J Clin Invest 135: e188348, 2025.
35. Bao D, Zhang C, Li L, Wang H, Li Q, Ni L, Lin Y, Huang R, Yang Z, Zhang Y and Hu Y: Integrative analysis of complement system to prognosis and immune infiltrating in colon cancer and gastric cancer. Front Oncol 10: 553297, 2021.
36. Roumenina LT, Daugan MV, Petitprez F, Sautès-Fridman C and Fridman WH: Context-dependent roles of complement in cancer. Nat Rev Cancer 19: 698-715, 2019.
37. Tang Y, Chen L, Xiao Y, Ran Q, Li Z and Chen M: Clinical significance of complement and coagulation cascades genes for patients with acute lymphoblastic leukemia. Int J Lab Hematol 47: 266-275, 2025.
38. Liu Y, Xiong S, Liu S, Chen J, Yang H, Liu G and Li G: Analysis of gene expression in bladder cancer: Possible involvement of mitosis and complement and coagulation cascades signaling pathway. J Comput Biol 27: 987-998, 2020.
39. Long GL, Chandra T, Woo SL, Davie EW and Kurachi K: Complete sequence of the cDNA for human alpha 1-antitrypsin and the gene for the S variant. Biochemistry 23: 4828-4837, 1984.
40. Choi YJ, Kim JE, Lee SJ, Gong JE, Jin YJ, Lee H and Hwang DY: Promotion of the inflammatory response in mid colon of complement component 3 knockout mice. Sci Rep 12: 1700, 2022.
41. Gualtierotti R, Solimeno LP and Peyvandi F: Hemophilic arthropathy: Current knowledge and future perspectives. J Thromb Haemost 19: 2112-2121, 2021.
42. Peng S, Du T, Wu W, Chen X, Lai Y, Zhu D, Wang Q, Ma X, Lin C, Li Z, *et al*: Decreased expression of serine protease inhibitor family G1 (SERPING1) in prostate cancer can help distinguish high-risk prostate cancer and predicts malignant progression. Urol Oncol 36: 366.e1-366.e9, 2018.
43. Damlaj M, Lipton JH and Assouline SE: A safety evaluation of omacetaxine mepesuccinate for the treatment of chronic myeloid leukemia. Expert Opin Drug Saf 15: 1279-1286, 2016.
44. Rosshandler Y, Shen AQ, Cortes J and Khoury HJ: Omacetaxine mepesuccinate for chronic myeloid leukemia. Expert Rev Hematol 9: 419-424, 2016.
45. Wang D, Qian X, Du YCN, Sanchez-Solana B, Chen K, Kanigicherla M, Jenkins LM, Luo J, Eng S, Park B, *et al*: cProSite: A web based interactive platform for online proteomics, phosphoproteomics, and genomics data analysis. J Biotechnol Biomed 6: 573-578, 2023.
46. Tang Z, Kang B, Li C, Chen T and Zhang Z: GEPIA2: An enhanced web server for large-scale expression profiling and interactive analysis. Nucleic Acids Res 47 (W1): W556-W560, 2019.