Cancer Cell International

RESEARCH Open Access

A novel risk stratification approach and molecular subgroup characterization based on coagulation related genes in colon adenocarcinoma

Xiangxin Wu[1†], Lichong Zhu[2†], Xizhe Sun[3†], Mingyu Xia[4], Shihui Zhao[4], Bomiao Zhang[4] and Tianyi Xia[4*]

Abstract
Colon adenocarcinoma (COAD) represents a significant health concern within the population. Advancing our understanding of COAD is imperative for early detection, enabling personalized treatment interventions, and facilitating the development of effective preventive measures. The coagulation system plays a role in tumor-related pathological processes; however, its specific involvement in COAD and potential contributors remain unclear. This study aimed to establish a novel risk stratification approach by analyzing coagulation related genes (CRGs) associated with COAD. Through a comprehensive bioinformatics analysis of data from public databases, we screened COAD associated CRGs and characterized the associated molecular subtypes. After a comprehensive analysis of the characteristics of each subtype, we applied differentially expressed genes in CRG subtypes to establish a new risk stratification method. Clinical subgroup analysis, immunoinfiltration analysis, therapeutic reactivity prediction and other analytical methods suggest the potential clinical value of the established risk stratification method. As one of the selected targets, the effect of MS4A4A on the proliferation and invasion of COAD was confirmed by in vitro experiments, which partially verified the reliability of bioinformatics results. Our findings delineate CRGs potentially implicated in COAD pathogenesis and offer fresh insights into the influence of the coagulation process on tumorigenesis and progression.

Keywords Coagulation related genes, Colon adenocarcinoma, Immune infiltration, Risk stratification, MS4A4A

Introduction

Colon adenocarcinoma (COAD) is the most common histological subtype of colon cancer and poses a significant health threat to the population [1]. Treatment options for COAD vary depending on the stage of the disease and may involve surgery, chemotherapy, radiotherapy and targeted therapy [2, 3]. Although surgical treatment is good, most patients are in the middle or late stages of cancer at initial diagnosis, making surgery a lost opportunity [4]. In addition, resistance to platinum-based chemotherapy regimen poses a significant challenge for drug therapy for COAD patients [5]. Increasing

†Xiangxin Wu, Lichong Zhu and Xizhe Sun contributed equally to this work and share first authorship.

*Correspondence:
Tianyi Xia
xiatianyi@hrbmu.edu.cn
[1]Department of Abdominal Surgery, Ganzhou Cancer Hospital, Ganzhou, China
[2]Department of Neurology and Neuroscience Center, the First Hospital of Jilin University, Changchun, China
[3]Research Center for Drug Safety Evaluation of Hainan, Hainan Medical University, Haikou, China
[4]Department of Colorectal Surgery, Harbin Medical University Cancer Hospital, Harbin Medical University, Harbin, China

the comprehensive understanding of COAD is essential to enable early detection with a view to personalized treatment interventions and the development of effective prevention measures.

The coagulation system is a complex biological process that regulates hemostasis. Blood clots are formed at the site of injury by a complex enzyme system in response to vascular damage [6, 7]. In addition to platelets, there are a variety of tissue factors and coagulation factors/anticoagulation factors involved in the coagulation process [8–11]. As a highly regulated biological process, the coagulation system ensures effective hemostasis and maintenance of blood flow through complex interactions between various coagulation factors and regulatory molecules, while avoiding excessive bleeding [12, 13].

One of the characteristics of malignant tumors is the disorder of the hemostatic system, which makes cancer patients prone to thrombosis and bleeding [14]. During tumorigenesis, immune cells can release various pro-inflammatory molecules, including cytokines, chemokines, and growth factors. These factors can stimulate the production of tissue factor (TF) [15, 16]. TFs expressed by immune cells and tumor cells participated in the formation of hypercoagulability in tumor patients [9, 17, 18]. Activated clotting factors may exacerbate inflammation by recruiting immune cells to injury or tumor sites [17, 19]. Inflammation during the anti-tumor process can induce endothelial dysfunction, which may further lead to coagulation disorders. In addition, tumor immune processes may disrupt the balance between pro-coagulant and anticoagulant factors, leading to an increased tendency to form blood clots [20]. Although there is a lot of evidence showing the correlation between the coagulation process and carcinogenesis, the role of coagulation in COAD and its possible participants remain unclear.

In this study, we established a new risk stratification method through the analysis of coagulation related genes (CRGs), identified potential intervention targets and verified their effects on COAD proliferation and invasion ability through in vitro experiments. Our results screened the CRGs that may be involved in the pathogenesis of COAD, and explored the influence of the coagulation process on the tumorigenesis and development from a new perspective.

Materials and methods

The gene transcriptome data downloading and standardization processing

In this study, we obtained gene expression data for COAD samples from two independent and publicly available databases (TCGA and GEO). Using the TCGA database, we downloaded transcriptome data, somatic mutation burden data, clinical baseline information, and copy number variation data for COAD samples. Within the Perl language environment, we utilized Perl scripts to extract transcriptome data, TMB files, and CNV data for COAD samples. The GSE39582 dataset was downloaded from the GEO database (Platform: GPL570) and annotated with gene symbols according to the platform annotation file. Within the R language framework, we employed the "SVA" script to standardize the gene expression data for COAD samples from the two independent databases, thereby eliminating batch effects between samples. After excluding samples without clinical survival baseline data, we extracted 487 samples with complete clinical baseline information from the TCGA database (normal: 41, COAD: 446), and 562 COAD samples from the GSE39582 dataset for subsequent exploration (Supplementary Table 1).

Identification of molecular subtypes of coagulation-related genes

Using "coagulation" as a keyword, we identified coagulation-related gene signatures from the Molecular Signatures Database (MSigDB) (Supplementary Table 2) [21]. Within the environment of the "limma" script, we extracted and analyzed the differential expression of CRG between normal samples and COAD samples (p.adjust$<$0.05). We conducted visual analysis of mutation frequency and CNV frequency of CRG signatures using "maftools" and "limma" scripts. We predicted the interactions between CRG signatures using the STRING database and visualized the network diagram. Based on the gene expression characteristics of differentially expressed CRGs, we used the "ConsensusClusterPlus" script to classify the molecular subtypes of CRGs in COAD samples with k=2–9. By analyzing model parameters such as consensus clustering cumulative distribution function (CDF) curves and delta area quantification curves between k=2–9, we identified the molecular subtypes of CRGs in COAD. We predicted the clinical survival outcomes of COAD samples in CRG molecular subgroups using the "survival" script and evaluated the differences between CRG molecular subtypes using PCA pattern plots with the "ggplot2" script. According to the KEGG pathway reference file "c2.cp.kegg.v7.2.symbols.gmt," we used the "GSVA" script to investigate the differential regulation of KEGG signaling pathways between CRG molecular subtypes.

Comprehensive analysis of gene subtypes associated with CRG molecular subtypes

Under the criteria of |FC| > 2 and p.adjust$<$0.05, we utilized the "limma" R script to compute and analyze differential expression gene signatures between CRG molecular subtypes. Subsequently, using the "ggplot2" script, we visualized volcano plots depicting the differential genes. Leveraging the "clusterProfiler" script, we

conducted GO and KEGG enrichment analyses on the differential genes associated with CRG subtypes to evaluate potential molecular functional mechanisms. Based on the expression profile characteristics of differential genes, we conducted comprehensive analysis of gene subtypes in COAD using the "ConsensusClusterPlus" script. We further classified COAD samples into different gene subtypes subgroups based on the optimal classification parameters of k=2–9 models.

Constructing a CRG scoring system and evaluating the stability of the model

Based on the expression profiles of differential genes and clinical baseline data, we comprehensively assessed the potential associations of each gene signature with the clinical prognosis of COAD. Utilizing the LASSO-univariate Cox analysis algorithm, we calculated the risk value (HR) and P-value for each differential gene. Subsequently, we further employed the multivariate Cox analysis algorithm to calculate the independent prognostic significance of the prognosis signature. Based on the risk parameters of each independent prognosis signature and expression profiles, we evaluated the CRG score for each COAD sample. The formula for constructing the CRG scoring system is as follows: CRG score = (MS4A4A × 0.911) + (SCG2×0.491) + (IGFBP6×0.847) + (C11orf96×0.741) + (CCL11 × -1.109) + (CXCL10 × -0.907). Using the "caret" script, we divided COAD samples into training and validation sets in a 7:3 ratio based on the expression of independent prognosis signatures and calculated the CRG scores for samples in both independent datasets. We analyzed the survival status of COAD in different classification groups using the "survival" R script and plotted time-dependent ROC curves using the "survivalROC" script to evaluate the accuracy of the CRG scoring system in predicting COAD clinical prognosis. Additionally, we used the "ggalluvial" R script to create Sankey diagrams, illustrating the potential associations among different molecular subtypes, CRG score subgroups, and clinical survival prognosis.

Assessment of immune microenvironment infiltration characteristics and prediction of immune therapy response

We used multiple immune infiltration assessment algorithms to evaluate the immune microenvironment infiltration status of COAD samples. Employing the "ESTIMATE" algorithm, we assessed immune infiltration status among different subgroups of COAD samples and obtained four indicators immune score, stromal score, ESTIMATE score, and tumor purity to reflect immune infiltration levels. Quantitative analysis of immune cell proportions in COAD samples was conducted based on established immune cell markers. We utilized the "GSVA" algorithm to evaluate the immune function scores among COAD subgroups. Immunotherapy data for COAD samples were downloaded from The Cancer Immunome Atlas (TCIA) database to assess the potential response of COAD samples to PD-1/CTLA4 immune therapy.

The independence evaluation of clinical prognosis predicted by the CRG scoring system and analysis of drug sensitivity

We integrated the clinical baseline characteristics of COAD samples and employed the "ggplot2" script to comprehensively analyze the distribution of CRG scores across different clinical features. Utilizing the "survival" script as a foundation, we conducted univariate and multivariate Cox analyses to calculate the hazard ratios (HR) and p-values of each clinical baseline feature and CRG score, evaluating the independent prognostic value of each variable. Subsequently, we applied the "rms" script to construct a novel nomogram model based on clinical baseline features and CRG scores to predict the survival probabilities of COAD samples at three-time intervals. Additionally, we utilized the "ggDCA," "regplot," and "rms" scripts to plot Decision Curve Analysis (DCA) curves, Concordance Index, and calibration curves to assess the accuracy and reliability of CRG scores and clinical baseline features in predicting COAD survival probabilities. Leveraging the transcriptomic features of samples, we used the Genomics of Drug Sensitivity in Cancer (GDSC) database to predict potential responses to targeted drugs (https://www.cancerrxgene.org/).

Cell culture

Human cell lines Caco2 and HCT116 were purchased from the American Type Culture Collection (ATCC). The above-mentioned cells were cultured in 1640 medium containing 10% fetal bovine serum and 1% penicillin-streptomycin. All cells were maintained in a sterile humidified incubator at 37 °C and 5% CO_2. Regular mycoplasma testing was conducted to verify that all cells were free of mycoplasma contamination.

Western blot analysis

Cells were harvested and lysed on ice using RIPA buffer (Beyotime, P0013B) supplemented with 1% PMSF. Equal amounts of proteins were separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis and subsequently transferred onto polyvinylidene difluoride membranes. Following blocking with 5% non-fat dry milk in Tris-buffered saline (TBS) for 1 h at room temperature (RT), the membranes were incubated with primary antibodies against MS4A4A (1:1000; Abcam, ab67134) and ATCB (1:5000; ABclonal, AC026) overnight at 4 °C. Then they were incubated with horseradish-peroxidase (HRP)-conjugated secondary antibody at RT for 1 h. Bands were visualized using the SuperPico ECL Chemiluminescence

Kit (Vazyme, E422-01), and band intensities were quantified using Image J software.

Reverse transcription, real-time qualitative PCR (RT-PCR)

Total RNA was extracted and reverse transcribed using RNA Purification Kit (Fastagen, 220011) and RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher, K1622) respectively, following the manufacturer's instructions. RT-PCR was performed using Hieff® qPCR SYBR Green Master Mix (TEASEN, 11201ES03). The MS4A4A primer sequences used were as follows: forward: ACCATGCAA GGAATGGAACAG; reverse, TTCCCATGCTAAGGCT CATCA.

CCK-8 assay

Cells were seeded in 96-well plates at a density of 1000 cells per well. Subsequently, 10 μL of CCK-8 solution dissolved in serum-free RPMI 1640 was added to each well and incubated for 1 h at 37 °C. The absorbance was then measured at a wavelength of 450 nm. The optical density (OD) was recorded daily at the same time.

Colony formation assays

Cells were seeded at a density of 1000 cells per well in a six-well plate and incubated at 37 °C with 5% CO2 for 1–2 weeks until visible colonies formed. Following this, the culture medium was removed, and the cells were fixed with methanol for 30 min. Subsequently, the cells were stained with a 0.1% crystal violet solution for 30 min. After staining, the excess staining solution was gently washed off with water. Using a microscope, the number of clones containing more than 50 cells was counted and recorded via photography. The clone formation rate was determined by averaging the number of clones formed in three replicate wells.

Transwell assay

Cells were harvested, centrifuged, and resuspended in serum-free medium for cell counting. A volume of 100 μL containing 100,000 cells was added to the upper chamber of a 24-well plate, while 600 μL of complete culture medium was added to the lower chamber. The plate was then placed in a cell culture incubator and incubated for an additional 48 h. Subsequently, the cells were fixed with methanol for 30 min and stained with a 0.5% crystal violet solution for 30 min. After rinsing the upper chamber with PBS, excess dye on the upper surface was gently removed with a cotton swab. Finally, the cells were observed, photographed, and counted under an inverted microscope.

Data analysis

Using the R and Perl programming languages, we conducted preprocessing and visualization analyses of the raw data from COAD samples. Statistical analysis between two groups was performed using the Wilcoxon rank-sum test, while statistical analysis among multiple groups utilized the One-way ANOVA test. Data from cell-based experiments were analyzed in triplicate. In this study, all statistical differences were subjected to multiple testing correction, with significance defined as $p<0.05$. Significance symbols were denoted as follows: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$.

Results

Analysis of gene expression patterns and variation landscape features of coagulation-related genes (CRG) in COAD

To elucidate the potential regulatory role of coagulation-related genes (CRG) signatures in COAD, we identified 40 CRG gene signatures from the MSigDB database and analyzed their differential expression and landscape features of genetic variation in COAD. Based on the results of the differential expression analysis using the "limma" package, significant differences in expression were observed for 34 CRG gene signatures between normal and COAD tissues (Fig. 1A, $p<0.05$). The waterfall plot of CRG mutation features revealed potential mutation burden in 156 out of 454 COAD samples, with mutation frequencies of 10% for VWF, 9% for F8, 6% for PLG, and 5% for F5 (Fig. 1B). Results of copy number frequency elucidated the variation characteristics of CRG gene signatures in COAD, with HNF4A, F7, F10, VWF, KNG1, and GNA12 showing significant CNV amplifications, while THBD, ENTPD1, LMAN1, MMRN1, F2RL2, and F2R exhibited noticeable CNV deletions (Fig. 1C). Based on these findings, we evaluated the expression status and genetic information features of CRG gene signatures in COAD, inferring their potential critical role in the onset of COAD. Subsequently, we further assessed the potential relationship between CRG gene signatures and the clinical prognosis of COAD. The prognosis network diagram indicated that 11 CRG gene signatures were associated with the clinical survival outcomes of COAD, comprising 8 risk factors and 3 beneficial factors (Fig. 1D). Additionally, the PPI network diagram revealed potential interactions among CRG gene signatures (Fig. 1E).

Identification of molecular subtypes of coagulation-related genes (CRGs) in COAD samples

Based on the expression characteristics of 34 differentially expressed CRG genes, we identified molecular subtype features related to CRG in COAD. Employing unsupervised consensus clustering algorithms, we divided COAD samples into two significantly distinct CRG molecular subtypes based on the optimal classification ratio, with 375 COAD samples in subtype A and 633 COAD samples in subtype B. Additionally, we

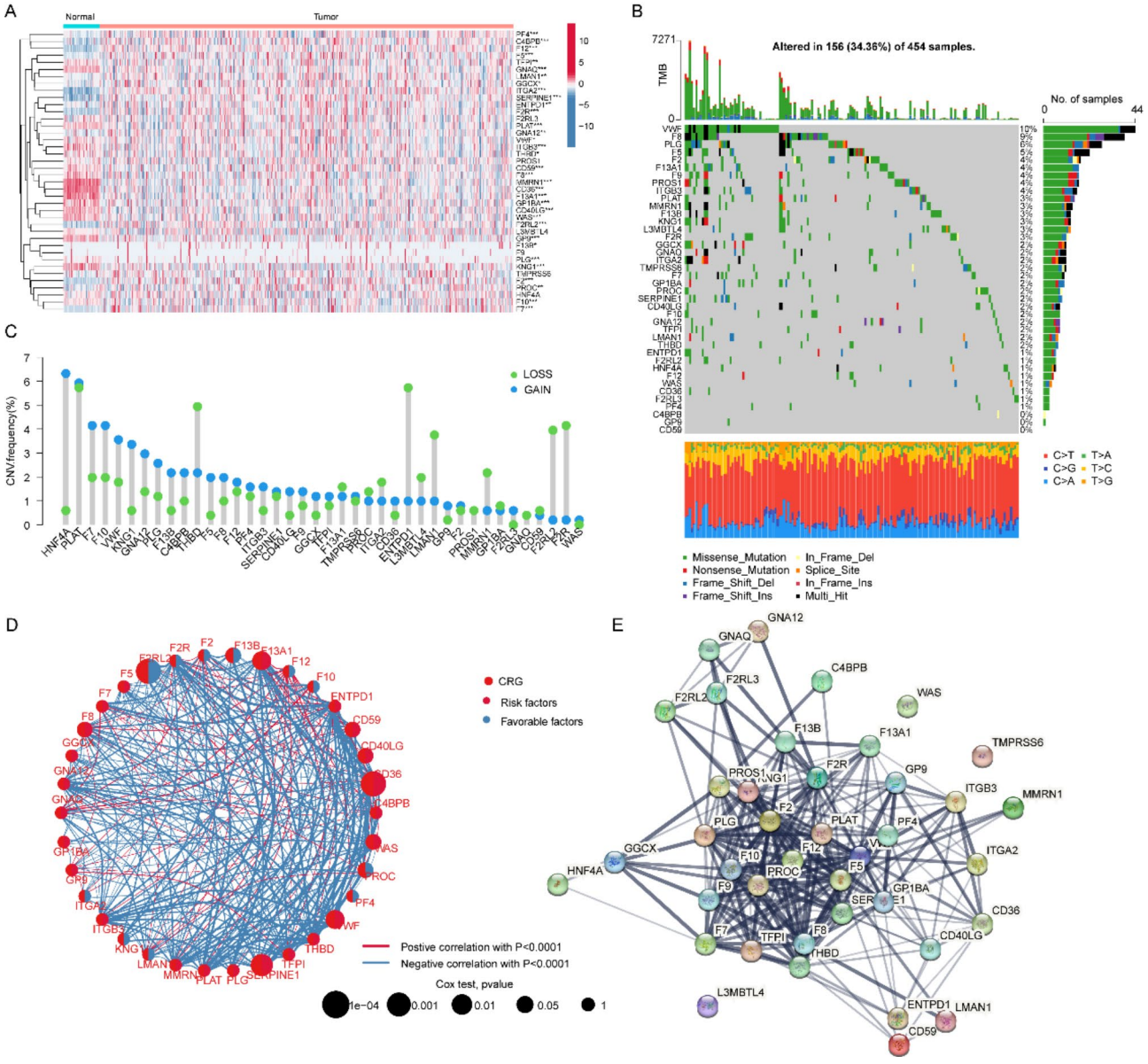

Fig. 1 Analysis of differential expression and genetic variation features of coagulation-related genes (CRG) in COAD. (**A**) Differential expression analysis of CRG in normal and COAD samples. (**B**) Assessment of mutation features of CRG gene signatures. (**C**) Quantitative assessment of copy number variation frequency in COAD. (**D**) Analysis of prognosis of CRG gene signatures in COAD. (**E**) PPI network diagram revealing potential interaction relationships among CRG gene signatures

further elucidated the potential relationship between differentially expressed CRG gene signatures and clinical features as well as molecular subtypes in COAD. Unsupervised PCA pattern plots demonstrated a significant separation pattern between subtypes A and B, indicating significant independence between the two CRG molecular subtypes (Fig. 2A). Survival prognosis results of CRG molecular subtypes suggested that the overall survival rate of COAD was significantly shorter in subtype A than in subtype B ($p<0.001$), indicating a potential association of high CRG gene expression subgroup with adverse prognosis in COAD (Fig. 2B). Based on GSVA analysis, we initially revealed potential regulatory KEGG signaling pathways between CRG molecular subtypes. The results indicated significant differences in tumor- and immune-related signaling pathways between CRG molecular subgroups, which may be important mechanisms mediating differences in clinical prognosis among CRG molecular subtypes (Fig. 2C). We further evaluated the mutation characteristic landscape among different CRG molecular subgroups, and the results showed that the mutation frequency of TP53, TTN, SYNE1 was higher in CRG subtype A, while the gene mutation frequency of APC and KRAS was higher in CRG subtype B (Fig. 2D).

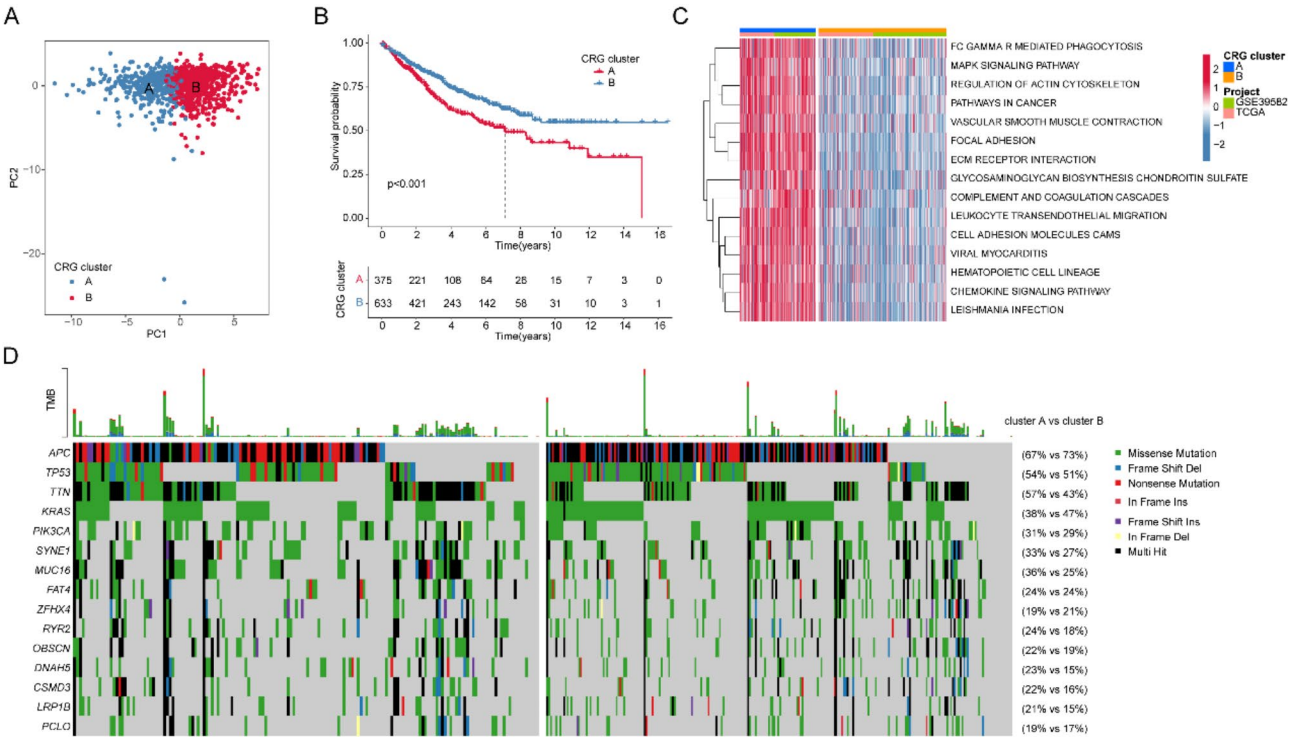

Fig. 2 Molecular subtype identification, mutation characterization and prognosis analysis based on CRG gene signatures in COAD. (**A**) PCA pattern recognition plot of CRG molecular subtypes. (**B**) Evaluation of clinical overall survival (OS) rates in CRG molecular subgroups. (**C**) Prediction analysis of KEGG signaling pathways between CRG molecular subtypes. (**D**) Tumor mutation characterization of CRG subtypes

Analysis of immune infiltration characteristics and prediction of immune therapy response in CRG molecular subtypes

Using the ESTIMATE assessment method, we quantitatively evaluated the immune infiltration status of CRG molecular subtypes. As depicted in Fig. 3A-D, we observed a significant decrease in immune score, ESTIMATE score, and stromal score, along with a significant increase in tumor purity in CRG molecular subtype B. This suggests a lower immune infiltration status in CRG molecular subtype B, which may contribute to better clinical prognosis in COAD. Results from the ssGSEA assessment algorithm indicated that COAD samples in CRG subtype A exhibited higher proportions of immune cells compared to CRG subtype B, including Activated B cells, activated CD4 T cells, activated CD8 T cells, and activated dendritic cells, suggesting a more pronounced immune-suppressive state in COAD samples in CRG subtype B. Meanwhile, quantitative analysis of immune functions revealed significant decreases in APC co inhibition, APC co stimulation, CCR, and cytolytic activity in CRG subtype B (Fig. 3E). Based on these results, we speculate that the immune infiltration characteristics and immune functional phenotypes of COAD samples in CRG subtype B are poorer compared to those in CRG subtype A, which may be crucial factors contributing to adverse prognosis. In the subsequent study, we utilized the TCIA database to predict and assess the response of COAD samples in CRG subtypes to CTLA4/PD-1 immune therapy. Results of the IPS phenotype indicated a potentially better response rate to CTLA4 immune therapy in COAD samples in CRG subtype B (Fig. 3F). Based on these findings, we preliminarily conclude that there are significant differences in immune infiltration status and immune functional phenotypes among COAD samples in CRG molecular subtypes, which may be important factors contributing to differences in clinical prognosis.

Comprehensive analysis of gene subtypes associated with CRG subtypes

To better explore the potential molecular mechanisms between CRG molecular subtypes, we conducted a comprehensive analysis of differential gene expression between the subtypes (threshold: $|FC|>2$, $p.adjust<0.05$). The volcano plot results elucidated that the majority of differentially expressed genes were significantly downregulated in CRG subtype B (Fig. 4A). GO analysis of differentially expressed genes in CRG subtypes suggested their involvement in regulating biological functions such as leukocyte migration, extracellular matrix organization, collagen-containing extracellular matrix, and extracellular matrix structural constituent (Fig. 4B). KEGG analysis indicated that differentially expressed genes in

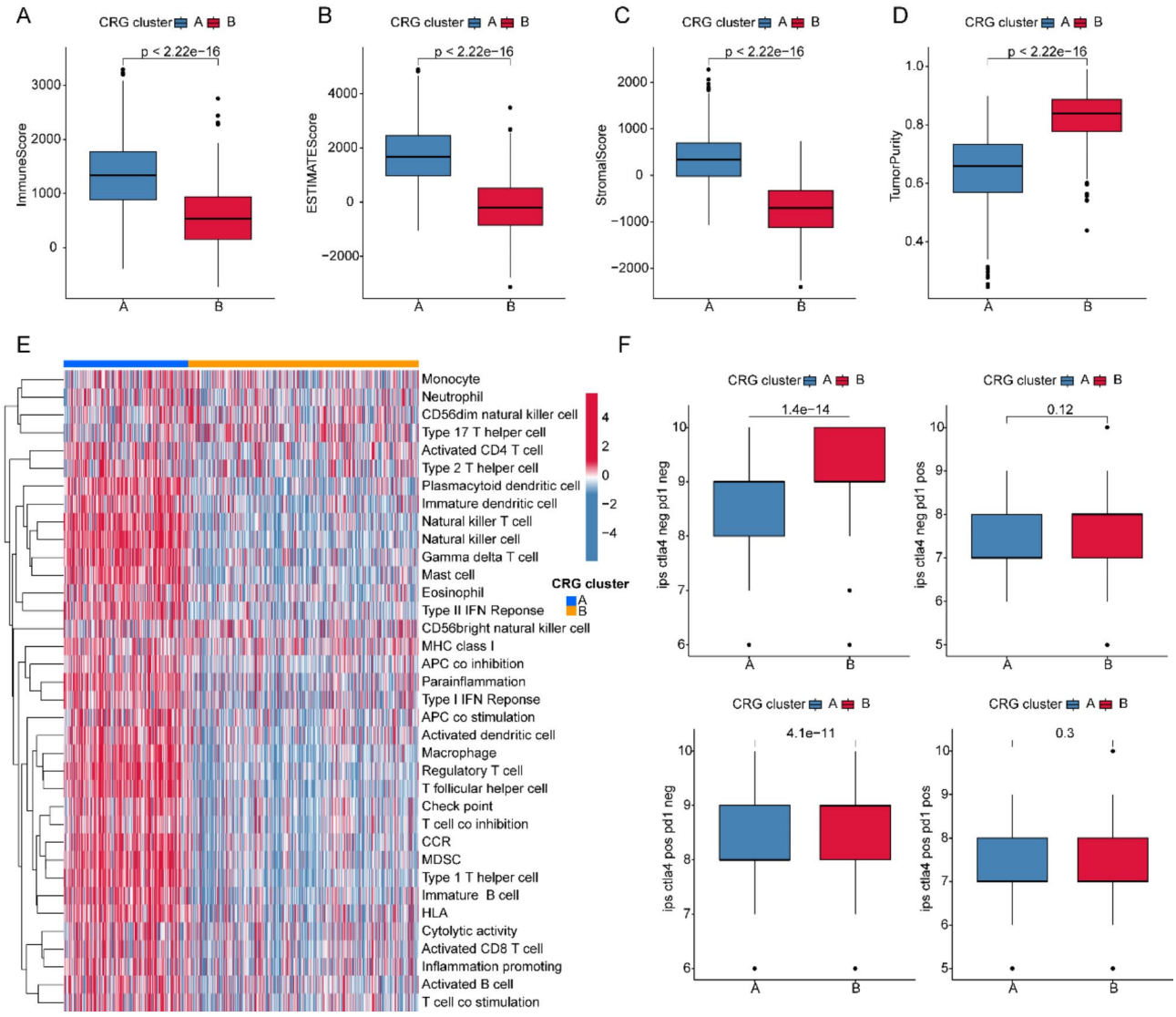

Fig. 3 Evaluation of immune infiltration characteristics and immune functional phenotypes in CRG molecular subtypes. (**A**-**D**) Assessment of immune status in CRG molecular subtypes. (**E**) Quantitative analysis of relative proportions of immune cells and immune functional phenotypes of CRG molecular subtypes. (**F**) Phenotypic assessment of response to CTLA-4/PD-1 immune therapy in CRG molecular subtypes

CRG subtypes were involved in the regulation of signaling pathways such as Phagosome, Cytokine-cytokine receptor interaction, Focal adhesion, and Staphylococcus aureus infection (Fig. 4C). Based on the expression profiles of differentially expressed genes in CRG subtypes, we further analyzed the gene subtype patterns of COAD samples. Unsupervised consensus clustering analysis revealed that COAD samples could be accurately classified into two distinct gene subtype patterns based on differentially expressed genes in CRG subtypes, with 381 samples in gene subtype A and 627 samples in gene subtype B (Fig. 4D). The PCA plot results indicated significant differentiation between the two gene subtypes, explaining the differences between them (Supplementary Fig. 1A). Clinical prognosis analysis curves demonstrated that the clinical overall survival (OS) rate of gene subtype B was significantly better than that of gene subtype A, suggesting that COAD samples in gene subtype B had better clinical survival benefits (Supplementary Fig. 1B). Visualization analysis of heatmap results revealed the expression of differentially expressed genes in CRG subtypes across different subtypes and pathological features, indicating a significant decrease in expression of differentially expressed genes in the subgroup with better clinical prognosis (Supplementary Fig. 1C). Meanwhile, among the gene subtypes, we found significant differences in the expression characteristics of CRG signatures as well (Fig. 4E).

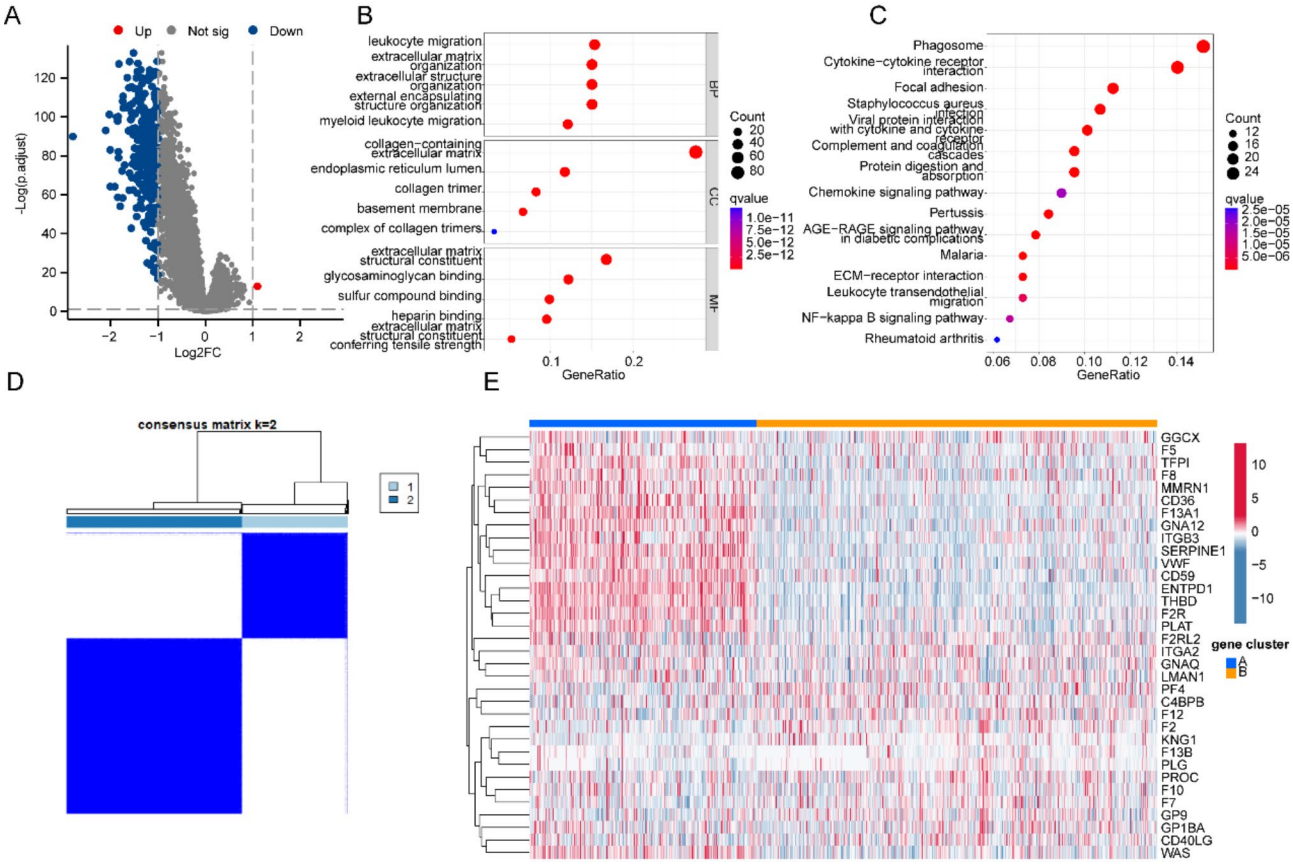

Fig. 4 Gene subtype identification based on CRG subtype-related genes. (**A**) Analysis of differential gene expression between CRG subtypes, with filtering threshold: |FC|>2, *p*.adjust < 0.05. (**B**, **C**) GO and KEGG enrichment analysis predictions of differentially expressed genes in CRG subtypes. (**D**) Gene subtype identification analysis based on the expression profile of differentially expressed genes in CRG subtypes. (**E**) Differential analysis of CRG gene signatures in gene subtypes

Comprehensive analysis of the potential value of the CRG scoring system in predicting COAD prognosis

Using the expression profiles of differentially expressed genes in CRG subtypes and clinical survival data from two independent databases (TCGA and GSE39582), we conducted a comprehensive analysis of the potential association between CRG subtype differentially expressed genes and clinical survival in COAD, and constructed a novel CRG scoring system to optimize risk stratification of COAD samples. Employing the LASSO-univariate Cox analysis algorithm, we identified variables from the CRG subtype differentially expressed gene signature associated with COAD clinical prognosis (Fig. 5A, B). Subsequently, using multivariate Cox analysis, we further identified variables with independent prognostic value and calculated the risk coefficients for each variable. Based on these risk coefficients and expression profiles, we computed the CRG scores for COAD samples and divided them into three independent validation cohorts in a 7:3 ratio. In the complete CRG score cohort, COAD samples in the low CRG score subgroup exhibited significantly better clinical prognosis than those in the high CRG score subgroup, indicating a significant association between high CRG scores and adverse COAD prognosis ($p<0.001$, Fig. 5C). Time-dependent ROC curve results suggested AUC values of 0.689, 0.669, and 0.644 at 1, 3, and 5 years, respectively (Fig. 5D). Furthermore, in the training and validation cohorts of the CRG scoring system, we observed that COAD samples in the low CRG score subgroup exhibited significantly better clinical prognosis than those in the high CRG score subgroup. The AUC values of the time-dependent ROC curves were 0.689, 0.673, and 0.685 in the training cohort, and 0.690, 0.664, and 0.567 in the validation cohort at 1, 3, and 5 years, respectively (Fig. 5E-H). Based on these results, we conclude that constructing a CRG scoring system for assessing and predicting clinical survival prognosis in COAD is stable and reliable. Utilizing a Sankey model diagram, we clearly demonstrated the associations between CRG molecular subtypes, gene subtypes, CRG scoring system, and COAD clinical prognosis, indicating a preference for high CRG scores in COAD samples with better prognosis (Fig. 5I). Across different molecular subtypes, we found that CRG subtype A and gene subtype A

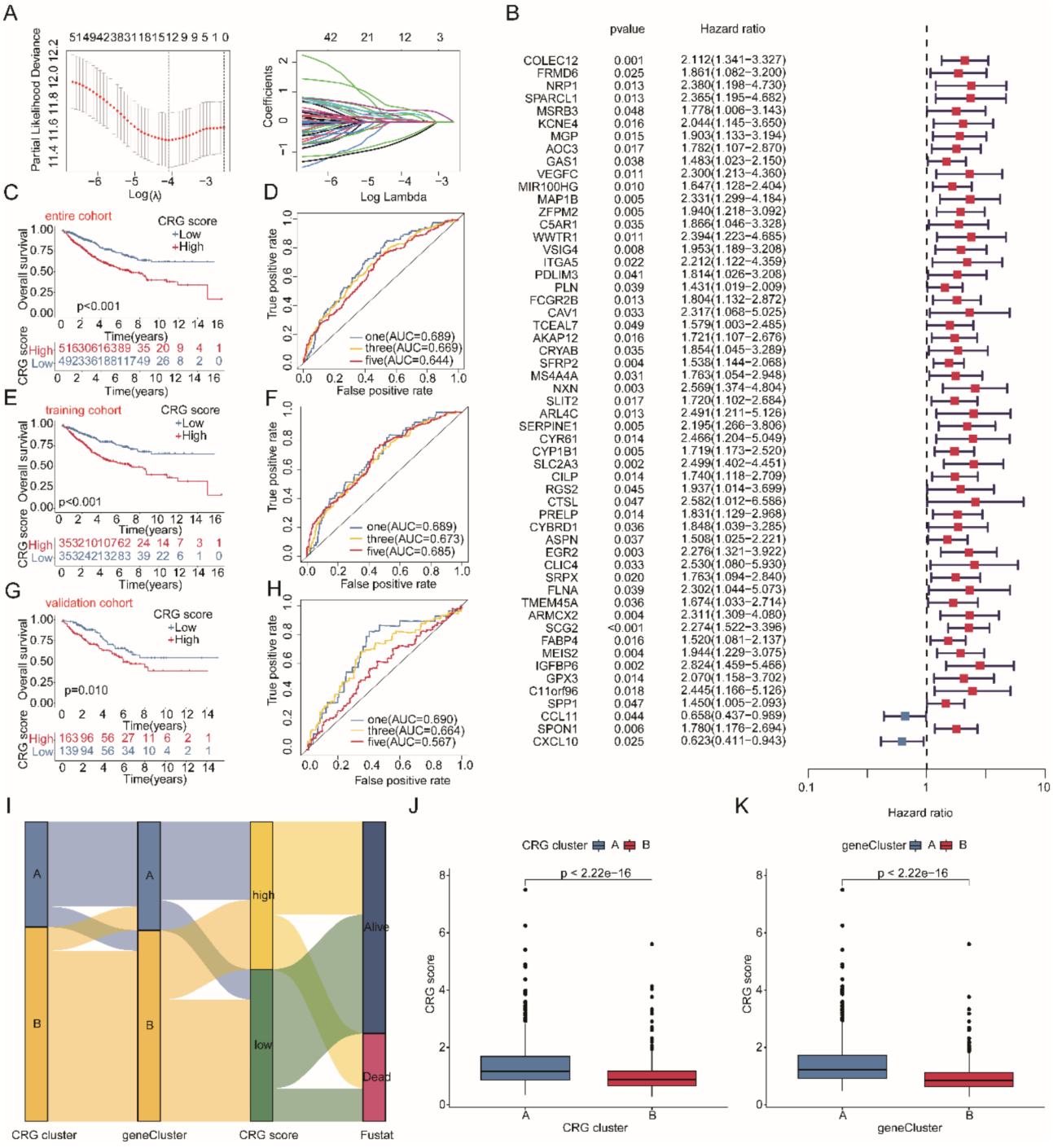

Fig. 5 Construction of the CRG scoring system and stability evaluation for predicting clinical prognosis in COAD. (**A**, **B**) Identification of gene signatures associated with COAD prognosis based on LASSO-univariate Cox analysis. (**C**, **D**) Model construction and time-dependent ROC curve analysis of the CRG scoring system in the complete cohort. (**E**-**H**) Establishment of the CRG scoring system model and time-dependent ROC curve analysis in the training and validation cohorts. (**I**) Sankey diagram analysis of the relationships among molecular subtypes, CRG scoring system, and clinical prognosis. (**J**, **K**) Differential analysis of CRG scores in CRG subtypes and gene subtypes

had significantly higher CRG scores than the other subtype, elucidating the close relationship between the CRG scoring system and molecular subtypes (Fig. 5J, K).

Independent prognostic analysis of the CRG scoring system and establishment of the Nomogram model

We further analyzed the potential association between different clinical features of COAD samples and the CRG scoring system. As depicted in Fig. 6A-E, we observed no significant differences in CRG scoring expression based on age and gender pathology features. However, significant differences were noted in CRG scoring expression based on disease-related N staging, T staging, and Stage, indicating a correlation between higher CRG scoring and advanced clinical staging (Fig. 6A-E). Based on these clinical-pathological feature variables and CRG scoring, we evaluated the association of each variable with the clinical prognosis of COAD and assessed the independent prognostic value of these variables. Univariate Cox analysis results indicated that stage (HR=2.084(1.741–2.494), $p<0.001$), T (HR=2.043(1.594–2.618), $p<0.001$), N (HR=1.484(1.270–1.733), $P<0.001$), and CRG scoring (HR=1.559(1.368-1,776), $p<0.001$) were correlated with poor prognosis of COAD. Multivariate Cox analysis results showed that CRG scoring (HR=1.440(1.244-1,667), $p<0.001$) was an independent prognostic factor for predicting COAD prognosis (Fig. 6F, G). Based on the clinical-pathological variables and CRG scoring, we established a nomogram model to assess the survival probability of COAD samples at 1 year, 3 years, and 5 years (Fig. 6H). Results from decision curve analysis (DCA), Concordance index, and calibration curve indicated that the nomogram established using clinical-pathological variables and CRG scoring reliably predicted the survival probability of COAD samples at 1 year, 3 years, and 5 years, with the predictive value of CRG scoring variables significantly superior to clinical variables (Fig. 6I-K). Based on these findings, we conclude that the CRG scoring system is a key independent prognostic factor for predicting the clinical prognosis of COAD samples, with significantly better predictive ability than clinical-pathological features.

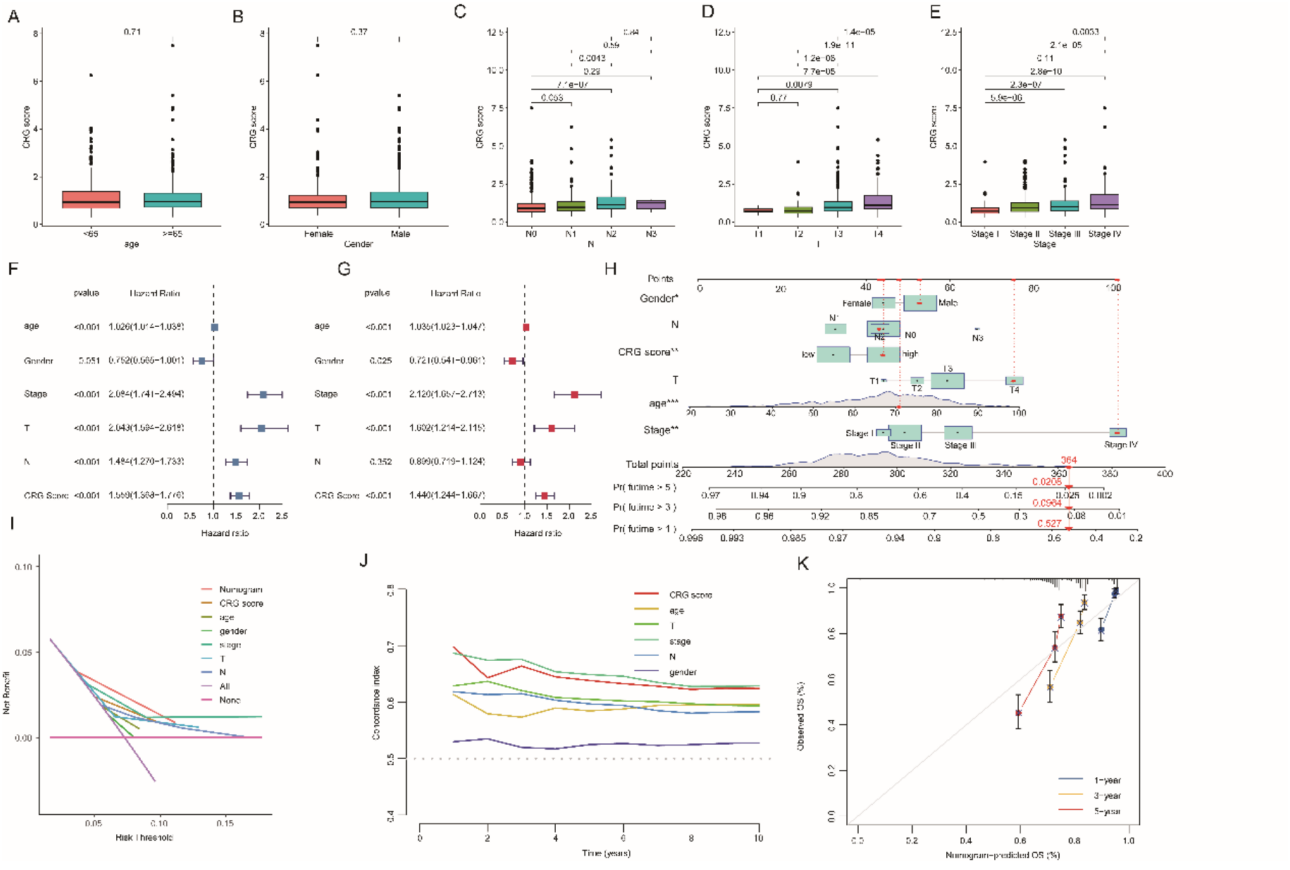

Fig. 6 Comprehensive Evaluation of the Independent Prognostic Value of the CRG Scoring System and Establishment of the Nomogram Model. (**A**-**E**) Distribution of CRG scoring across different clinical-pathological features of COAD. (**F**, **G**) Univariate and multivariate Cox analyses of clinical-pathological variables and the CRG scoring system. (**H**) Establishment of the nomogram model based on clinical variables and CRG scoring. (**I**) Decision curve analysis (DCA) curve computation. (**J**) Concordance index evaluation of different variable features at different survival times. (**K**) Analysis of consistency index between predicted OS using the nomogram model and actual OS over time

The correlation analysis between the CRG scoring system and the immune microenvironment infiltration and somatic mutation landscape of COAD

We further analyzed the association between the CRG scoring system and the immune infiltration microenvironment in COAD. Utilizing the ESTIMATE algorithm, we predicted the immune infiltration status of CRG scoring subgroups, revealing that in the high CRG scoring subgroup, COAD samples exhibited higher stromal and ESTIMATE scores, along with lower tumor purity (Fig. 7A-D). These findings suggest potential differences in the immune infiltration status of COAD within CRG scoring subgroups. Employing the ssGSEA evaluation algorithm, we elucidated the proportions of immune cell infiltration in COAD samples within CRG scoring subgroups. The results indicated a significant upregulation in the infiltration proportions of most immune cells in the high CRG scoring subgroup, including CD56dim natural killer cells, eosinophils, and gamma delta T cells. Additionally, predictions of immune function indicated a significant decrease in most immune function scores in the low CRG scoring subgroup compared to the high CRG scoring subgroup (Fig. 7E). Furthermore, we assessed the somatic mutation landscape of COAD samples within CRG scoring subgroups, revealing that compared to the low CRG scoring subgroup, COAD samples in the high CRG scoring subgroup may exhibit a higher somatic mutation characteristic. Gene mutation frequency results demonstrated a significantly higher mutation frequency of APC, TP53, and TTN in the low CRG scoring subgroup, while the mutation frequency of KRAS was lower (Fig. 7F).

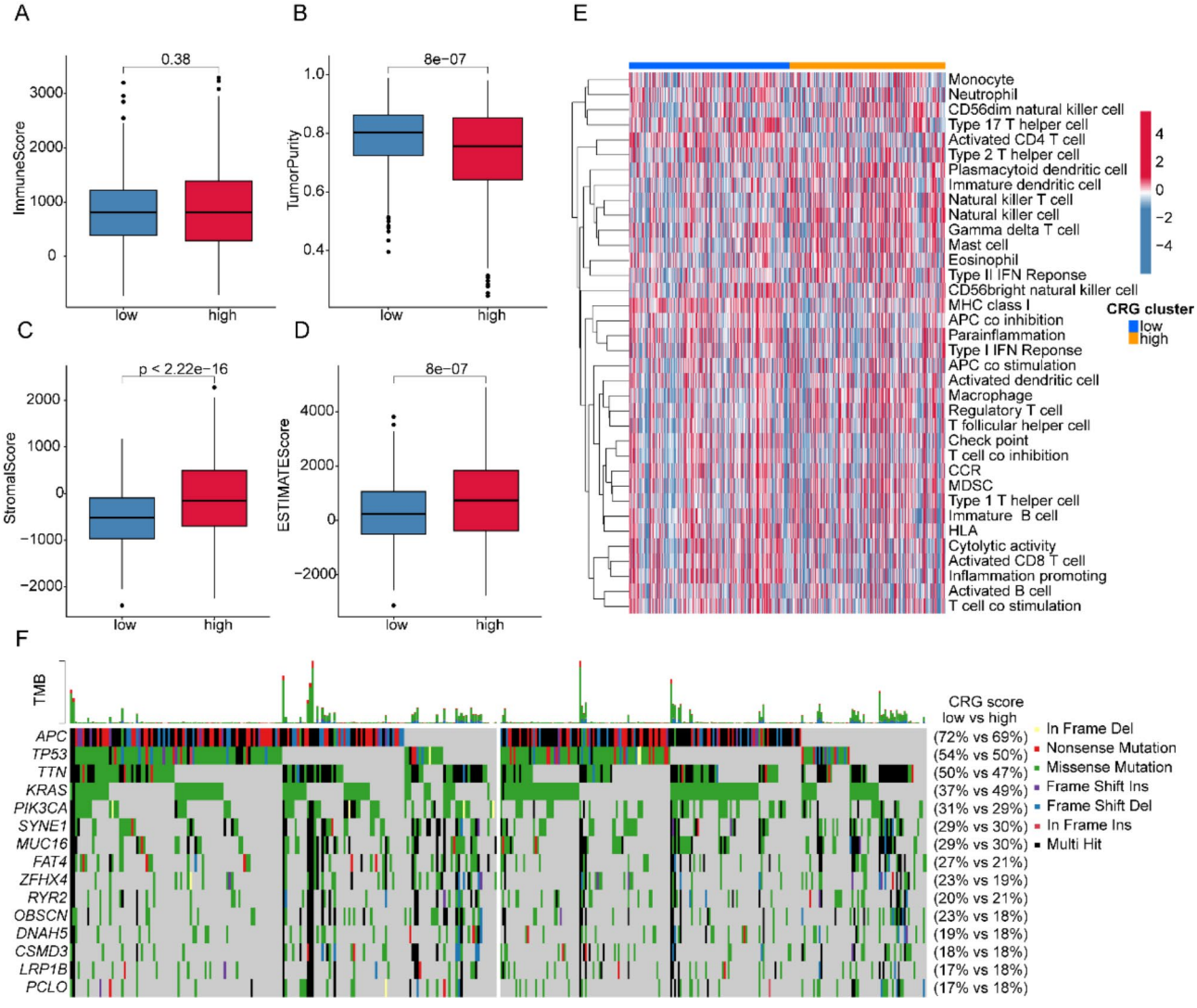

Fig. 7 Correlation analysis of the CRG scoring system with immune infiltration microenvironment and somatic mutation characteristics. (**A**-**D**) Evaluation of the immune infiltration status of the CRG scoring system. (**E**) Quantitative analysis of immune cell proportions and immune function scores in CRG scoring subgroups. (**F**) Assessment of somatic mutation characteristics in CRG scoring subgroups

The association of the CRG scoring system with immune therapy response and prediction of potential therapeutic drugs

Based on the quantitative analysis of the somatic mutation features and immune infiltration of the CRG scoring subgroups, we predicted the response of CRG scoring subgroups to PD-1 and CTLA-4 immune therapies. The IPS quantitative results elucidated that in the high CRG scoring subgroup, COAD samples had significantly lower IPS scores compared to the low CRG scoring subgroup, suggesting that COAD samples associated with high CRG scores may derive greater clinical treatment benefits from PD-1/CTLA-4 immune therapies (Fig. 8A-D). Meanwhile, based on the GDSC database, we also predicted some small molecule compounds that may have potential clinical benefits. As shown in Fig. 8E-L, the IC50 results of drug sensitivity indicated that the high CRG scoring subgroup may be more sensitive to GNF-2, Paclitaxel, Rapamycin, and VX-680, while the low CRG scoring subgroup may have better responses to CGP-082996, Dasatinib, Erlotinib, and Saracatinib. Based on these results, we predicted and identified the response degree and sensitivity of CRG scoring subgroups to immune therapy and targeted drug therapy, providing new perspectives for personalized treatment stratification of COAD in the future.

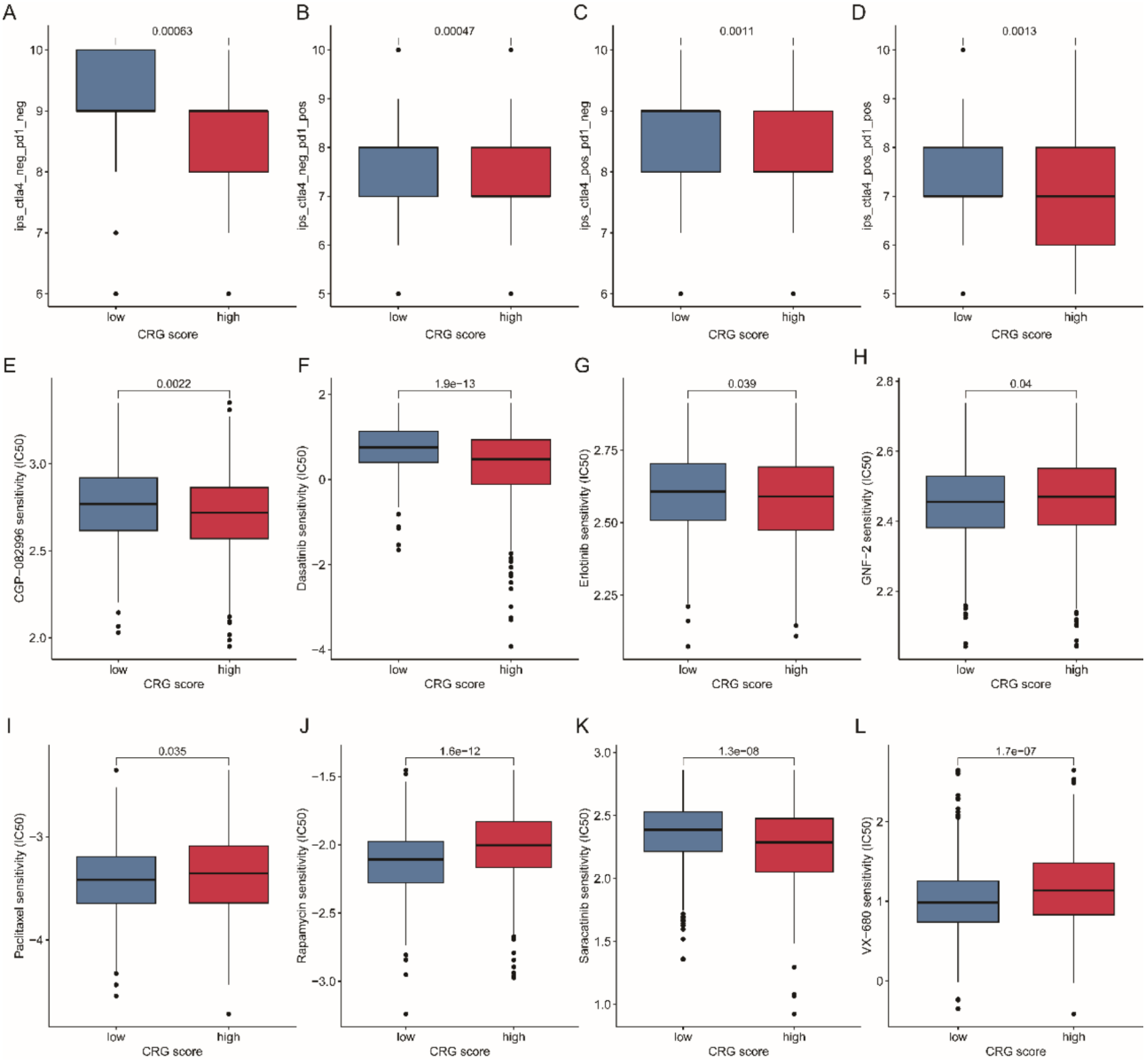

Fig. 8 Prediction of the response of CRG scoring subgroups to PD-1/CTLA-4 and targeted drugs. (**A**-**D**) Prediction of the response to immune therapy based on the TCIA database. (**E**-**L**) Assessment of potential targeted drug sensitivity associated with CRG scoring subgroups

MS4A4A promotes COAD tumor proliferation and migration

Considering the fact that MS4A4A demonstrates the highest HR coefficient during the screening process, we chose it as the focal point of our research to validate the reliability of the bioinformatics analysis conducted in this study. Firstly, we assessed the basic protein expression levels of MS4A4A in COAD cell lines Caco2 and HCT116 (Fig. 9A). Subsequently, we established a high-expression cell line with relatively low expression levels of HCT116, while creating a knockdown cell line with a relatively high expression level of Caco2. Western Blot and qRT-PCR analyses confirmed the high expression and low knockdown efficiency (Fig. 9B-E). CCK8 assay and clone formation assay demonstrated that knockdown of the MS4A4A gene inhibited tumor cell proliferation (Fig. 9F-H). Transwell experiments further validated that the high expression of the MS4A4A gene could promote tumor migration (Fig. 9I, J).

Discussion

Due to tumor heterogeneity, COAD clinical and histopathological features do not always accurately predict patient outcomes. However, the development of bioinformatics technology has greatly increased the reliability of predicting patient outcomes through genetic analysis [22, 23]. In this study, we established a new risk stratification method through the analysis of COAD associated CRGs, thereby providing new clues to the role of the coagulation process in the development of COAD carcinogenesis.

Our results suggest the potential value of coagulation related therapy in the treatment of COAD. Despite the lack of evidence in COAD, adjuvant therapy targeting coagulation related pathways can significantly improve patient outcomes in other tumor types [24]. One example is the discovery of the anti-cancer properties of low molecular weight heparin. Low molecular weight heparin has a positive anti-cancer effect in patients with various tumors, including pancreatic cancer and lung cancer [25–27]. Its specific mechanisms include inhibition of tumor-induced angiogenesis, regulation of tumor cell heparinase action, selectin mediated inhibition of tumor cell distant metastasis and activation of coagulation system [28–31]. In addition, the potential of natural thrombomodulin and thrombomodulin as new targets to inhibit cancer progression is being recognized [32–34]. These results, together with ours, suggest a role for the coagulation process in the development of COAD.

The membrane-spanning 4 A (MS4A) family is expressed differently and selectively in immunocompetent cells and macrophages and is associated with and regulates signal transduction activity of different types of immune receptors, including pattern recognition receptors (PRRs) and Ig receptors [35–37]. One of the members, MS4A4A protein was confirmed to be selectively expressed by macrophages in healthy tissues and tumor-associated macrophages in COAD patients [35]. Some evidence shows the role of MS4A4A in the development of tumors, including COAD. MS4A4A enhances the ability of dectin-1 mediated NK cells and macrophages to fight distant metastasis of tumors in mouse models [38]. On the contrary, macrophages lacking MS4A4A showed signal transduction deficits [36]. MS4A4A targeting tumor-associated macrophages (TAMs) restores anti-tumor immunity mediated by $CD8^+$ T cells [38]. In addition, the effect of MS4A4A on tumor-related immunity has also been confirmed by a series of articles. MS4A4A is involved in the modulation of the biological functions of immunoreceptors such as FcεRI, Dectin-1, and Vslg4 [36, 39, 40]. In addition, Syk phosphorylation and ROS, IL-6, and TNF production were reduced in MS4A4A-deficient macrophages compared to controls [35, 41]. Based on the above evidence, further investigation of the role of MS4A4A in cancer has positive significance.

Due to the limited conditions, the molecular mechanism of the selected target in COAD could not be further verified in this paper. In addition, the target of bioinformatics screening is inevitably biased in time and space. However, this paper provides a new way to study the role of coagulation in COAD and a new risk stratification method with potential clinical application value. Additionally, research on COAD is increasingly highlighting differences in the pathogenesis and treatment of different types of tumors. It should be noted that the results of this study are more focused on COAD of colon cancer rather than rectal cancer. Further analysis of the possible molecular mechanisms of colon cancer on this basis will be of positive significance.

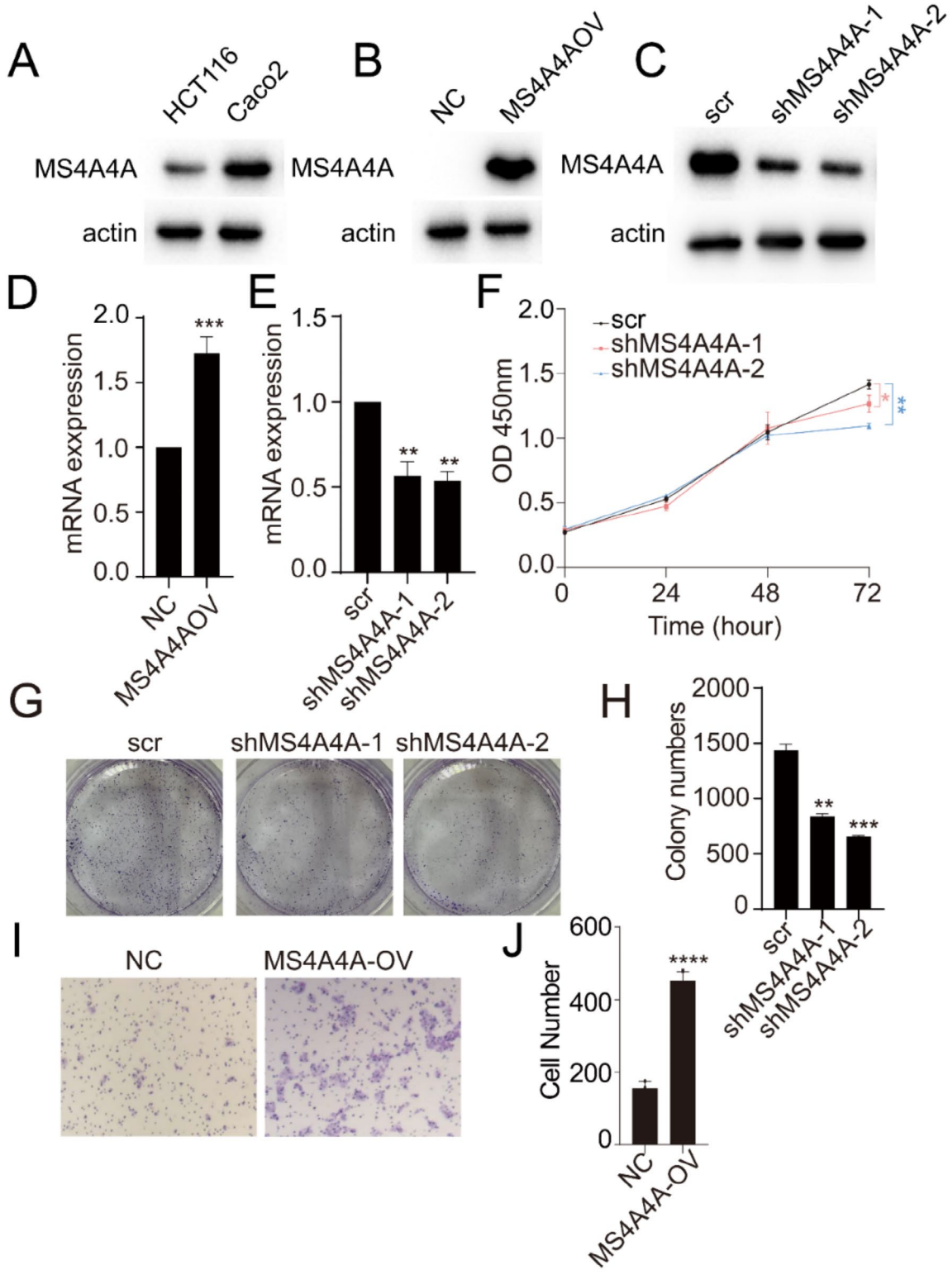

Fig. 9 Effects of MS4A4A on cell proliferation and migration in COAD cell lines. (**A**) Protein expression levels of MS4A4A across various COAD cell lines. (**B**) Representative western blot bands illustrating MS4A4A overexpression in HCT116 cells. (**C**) Representative bands depicting MS4A4A knockdown in Caco2 cells. (**D**) RT-PCR results demonstrating the efficiency of MS4A4A overexpression in HCT116 cells. (**E**) RT-PCR results showing the efficiency of MS4A4A knockdown in Caco2 cells. (**F**, **G**, **H**) CCK8 assay and clone formation assay conducted in Caco2 cells with low MS4A4A expression. (**I**, **J**) Transwell assay carried out in HCT116 cells with stable MS4A4A overexpression

Supplementary information

The online version contains supplementary material available at https://doi.org/10.1186/s12935-024-03491-2.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

Supplementary Material 7

Supplementary Material 8

Author contributions

T.X. designed the study. X.W., L.Z. and X.S. collected and analyzed the data. L.Z. and X.S. wrote the manuscript. M.X., S.Z. and B.Z. wrote and edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study is supported by Postdoctoral Research Project in Heilongjiang Province (LBH-Z22229).

Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate

The data for this study were obtained from public databases and therefore did not require ethics committee approval and review.

Consent for publication

All authors agree to be published.

Competing interests

The authors declare no competing interests.

Conflict of interest

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

Received: 28 April 2024 / Accepted: 28 August 2024
Published online: 09 September 2024

References

1. Feng RM, Zong YN, Cao SM, Xu RH. Current cancer situation in China: good or bad news from the 2018 Global Cancer statistics? Cancer Commun (Lond). 2019;39(1):22.
2. Kohne-Wompner CH, Schoffski P, Schmoll HJ. Adjuvant therapy for colon adenocarcinoma: current status of clinical investigation. Ann Oncol. 1994;5(Suppl 3):97–104.
3. Brivio F, Fumagalli L, Fattori L, Nespoli L, Denova M, Sargenti E, Nespoli A. [Clinical role of interleukin-2 in the surgical treatment of liver metastasis due to colon adenocarcinoma]. Minerva Chir. 2004;59(6):573–82.
4. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J. Cancer statistics in China, 2015. CA Cancer J Clin. 2016;66(2):115–32.
5. Brzozowa-Zasada M, Matysiak N, Piecuch A, Gawelek E, Michalski M, Kucharzewski M, Los MJ. The Prognostic significance of apoptotic protease activating factor (Apaf-1) protein expression in Colon adenocarcinoma tissue-preliminary Report. Front Biosci (Landmark Ed). 2023;28(2):29.
6. Mackman N. Role of tissue factor in hemostasis, thrombosis, and vascular development. Arterioscler Thromb Vasc Biol. 2004;24(6):1015–22.
7. Neubauer K, Zieger B. Endothelial cells and coagulation. Cell Tissue Res. 2022;387(3):391–8.
8. Adams RL, Bird RJ. Review article: coagulation cascade and therapeutics update: relevance to nephrology. Part 1: overview of coagulation, thrombophilias and history of anticoagulants. Nephrol (Carlton). 2009;14(5):462–70.
9. Ahmadi SE, Shabannezhad A, Kahrizi A, Akbar A, Safdari SM, Hoseinnezhad T, Zahedi M, Sadeghi S, Mojarrad MG, Safa M. Tissue factor (coagulation factor III): a potential double-edge molecule to be targeted and re-targeted toward cancer. Biomark Res. 2023;11(1):60.
10. Zelaya H, Rothmeier AS, Ruf W. Tissue factor at the crossroad of coagulation and cell signaling. J Thromb Haemost. 2018;16(10):1941–52.
11. Chen Z, Herzog RW, Kaufman RJ. Cellular stress and coagulation factor production: when more is not necessarily better. J Thromb Haemost. 2023;21(12):3329–41.
12. Gobel K, Eichler S, Wiendl H, Chavakis T, Kleinschnitz C, Meuth SG. The coagulation factors fibrinogen, Thrombin, and Factor XII in inflammatory Disorders-A systematic review. Front Immunol. 2018;9:1731.
13. Al-Koussa H, AlZaim I, El-Sabban ME. Pathophysiology of Coagulation and Emerging roles for Extracellular vesicles in Coagulation cascades and disorders. J Clin Med 2022, 11(16).
14. Falanga A, Marchetti M, Vignoli A. Coagulation and cancer: biological and clinical aspects. J Thromb Haemost. 2013;11(2):223–33.
15. Palacios-Acedo AL, Mege D, Crescence L, Dignat-George F, Dubois C, Panicot-Dubois L. Platelets, Thrombo-Inflammation, and Cancer: collaborating with the enemy. Front Immunol. 2019;10:1805.
16. Panova-Noeva M, Schulz A, Arnold N, Hermanns MI, Prochaska JH, Laubert-Reh D, Spronk HM, Blettner M, Beutel M, Pfeiffer N, et al. Coagulation and inflammation in long-term cancer survivors: results from the adult population. J Thromb Haemost. 2018;16(4):699–708.
17. Koizume S, Miyagi Y. Tissue factor in cancer-associated thromboembolism: possible mechanisms and clinical applications. Br J Cancer. 2022;127(12):2099–107.
18. Hisada Y, Mackman N. Tissue factor and extracellular vesicles: activation of Coagulation and Impact on Survival in Cancer. Cancers (Basel) 2021, 13(15).
19. Galmiche A, Rak J, Roumenina LT, Saidak Z. Coagulome and the tumor microenvironment: an actionable interplay. Trends Cancer. 2022;8(5):369–83.
20. Setiawan B, Budianto W, Sukarnowati TW, Rizky D, Pangarsa EA, Santosa D, Setiabudy RD, Suharti C. Correlation of inflammation and coagulation markers with the incidence of deep vein thrombosis in Cancer patients with high risk of thrombosis. Int J Gen Med. 2022;15:6215–26.
21. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;27(12):1739–40.
22. Ibrahiem AT, Eladl E, Toraih EA, Fawzy MS, Abdelwahab K, Elnaghi K, Emarah Z, Shaalan AAM, Ehab Z, Soliman NA. Prognostic Value of BRAF, programmed cell death 1 (PD1), and PD Ligand 1 (PDL1) protein expression in Colon adenocarcinoma. Diagnostics (Basel) 2023, 13(2).
23. Koshkin SA, Anatskaya OV, Vinogradov AE, Uversky VN, Dayhoff GW 2nd, Bystriakova MA, Pospelov VA, Tolkunova EN. Isolation and characterization of human Colon Adenocarcinoma Stem-Like cells based on the endogenous expression of the stem markers. Int J Mol Sci 2021, 22(9).
24. Kuderer NM, Ortel TL, Francis CW. Impact of venous thromboembolism and anticoagulation on cancer and cancer survival. J Clin Oncol. 2009;27(29):4902–11.
25. Laner-Plamberger S, Oeller M, Rohde E, Schallmoser K, Strunk D. Heparin and derivatives for Advanced Cell therapies. Int J Mol Sci 2021, 22(21).
26. Chen D. Heparin beyond anti-coagulation. Curr Res Transl Med. 2021;69(4):103300.
27. Schrag D, Uno H, Rosovsky R, Rutherford C, Sanfilippo K, Villano JL, Drescher M, Jayaram N, Holmes C, Feldman L, et al. Direct oral anticoagulants vs low-molecular-weight heparin and recurrent VTE in patients with Cancer: a Randomized Clinical Trial. JAMA. 2023;329(22):1924–33.
28. Falanga A, Vignoli A, Diani E, Marchetti M. Comparative assessment of low-molecular-weight heparins in cancer from the perspective of patient outcomes and survival. Patient Relat Outcome Meas. 2011;2:175–88.
29. Noble S. Low-molecular-weight heparin and survival in lung cancer. Thromb Res. 2012;129(Suppl 1):S114–118.
30. van Doormaal FF, Di Nisio M, Otten HM, Richel DJ, Prins M, Buller HR. Randomized trial of the effect of the low molecular weight heparin nadroparin on survival in patients with cancer. J Clin Oncol. 2011;29(15):2071–6.

31. Riess H, Pelzer U, Hilbig A, Stieler J, Opitz B, Scholten T, Kauschat-Bruning D, Bramlage P, Dorken B, Oettle H. Rationale and design of PROSPECT-CONKO 004: a prospective, randomized trial of simultaneous pancreatic cancer treatment with enoxaparin and chemotherapy). BMC Cancer. 2008;8:361.
32. Horowitz NA, Palumbo JS. Mechanisms coupling thrombomodulin to tumor dissemination. Thromb Res. 2012;129(Suppl 1):S119–121.
33. Spek CA, Arruda VR. The protein C pathway in cancer metastasis. Thromb Res. 2012;129(Suppl 1):S80–84.
34. Chen LM, Wang W, Lee JC, Chiu FH, Wu CT, Tai CJ, Wang CK, Tai CJ, Huang MT, Chang YJ. Thrombomodulin mediates the progression of epithelial ovarian cancer cells. Tumour Biol. 2013;34(6):3743–51.
35. Mattiola I, Mantovani A, Locati M. The tetraspan MS4A family in homeostasis, immunity, and disease. Trends Immunol. 2021;42(9):764–81.
36. Mattiola I, Tomay F, De Pizzol M, Silva-Gomes R, Savino B, Gulic T, Doni A, Lonardi S, Astrid Boutet M, Nerviani A, et al. The macrophage tetraspan MS4A4A enhances dectin-1-dependent NK cell-mediated resistance to metastasis. Nat Immunol. 2019;20(8):1012–22.
37. Sanyal R, Polyak MJ, Zuccolo J, Puri M, Deng L, Roberts L, Zuba A, Storek J, Luider JM, Sundberg EM, et al. MS4A4A: a novel cell surface marker for M2 macrophages and plasma cells. Immunol Cell Biol. 2017;95(7):611–9.
38. Li Y, Shen Z, Chai Z, Zhan Y, Zhang Y, Liu Z, Liu Y, Li Z, Lin M, Zhang Z, et al. Targeting MS4A4A on tumour-associated macrophages restores CD8 + T-cell-mediated antitumour immunity. Gut. 2023;72(12):2307–20.
39. Huang X, Feng Z, Jiang Y, Li J, Xiang Q, Guo S, Yang C, Fei L, Guo G, Zheng L, et al. VSIG4 mediates transcriptional inhibition of Nlrp3 and Il-1beta in macrophages. Sci Adv. 2019;5(1):eaau7426.
40. Donnadieu E, Jouvin MH, Kinet JP. A second amplifier function for the allergy-associated fc(epsilon)RI-beta subunit. Immunity. 2000;12(5):515–23.
41. Arthur GK, Ehrhardt-Humbert LC, Snider DB, Jania C, Tilley SL, Metcalfe DD, Cruse G. The FcepsilonRIbeta homologue, MS4A4A, promotes FcepsilonRI signal transduction and store-operated ca(2+) entry in human mast cells. Cell Signal. 2020;71:109617.

Publisher's note