

Construction and validation of a novel angiogenesis pattern to predict prognosis and immunotherapy efficacy in colorectal cancer

Zhiyong Li¹, Yang Liu², Peng Guo^{1,4,5}, Yunwei Wei^{2,3}

¹Department of Emergency Surgery, Peking University People's Hospital, Xicheng, Beijing 100044, China

²Department of Pancreatic and Gastrointestinal Surgery Division, Ningbo Second Hospital, Ningbo, Zhejiang 315010, China

³Ningbo Key Laboratory of Intestinal Microecology and Human Major Diseases, Ningbo, Zhejiang 315010, China

⁴Department of Emergency Medicine, Peking University People's Hospital, Xicheng, Beijing 100044, China

⁵Laboratory of Surgery Oncology, Peking University People's Hospital, Xicheng, Beijing 100044, China

Correspondence to: Peng Guo, Yunwei Wei; **email:** rmyyguopeng@163.com, <https://orcid.org/0000-0001-6749-9430>; hydwyw11@hotmail.com, <https://orcid.org/0000-0001-6482-4680>

Keywords: colorectal cancer, angiogenesis, prognostic, immunotherapy response, *Fusobacterium nucleatum*, biomarker

Received: April 12, 2023

Accepted: October 2, 2023

Published: November 7, 2023

Copyright: © 2023 Li et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Background: Evidence suggests that the tumor microenvironment (TME) affects the tumor active response to immunotherapy. Tumor angiogenesis is closely related to the TME. Nonetheless, the effects of angiogenesis on the TME of colorectal cancer (CRC) remain unknown.

Methods: We comprehensively assessed the angiogenesis patterns in CRC based on 36 angiogenesis-related genes (ARGs). Subsequently, we evaluated the prognostic values and therapeutic sensitivities of angiogenesis patterns using multiple methods. We then performed the machine learning algorithm and functional experiments to identify the prognostic key ARGs. Ultimately, the regulation of gut microbiota on the expression of ARGs was further investigated by using whole genome sequencing.

Results: Two angiogenesis clusters were identified and angiogenesis cluster B was characterized by increased stromal and immunity activation with unfavorable odds of survival. Further, an ARG_score including 9 ARGs to predict recurrence-free survival (RFS) was established and its predominant predictive ability was confirmed. The low ARG_score patients were characterized by a high mutation burden, high microsatellite instability, and immune activation with better prognosis. Moreover, patients with high KLK10 expression were associated with a hot tumor immune microenvironment, poorer immune checkpoint blocking treatment, and shorter survival. The *in vitro* experiments also indicated that *Fusobacterium nucleatum* (*F.n*) infection significantly induced KLK10 expression in CRC.

Conclusions: The quantification of angiogenesis patterns could contribute to predict TME characteristics, prognosis, and individualized immunotherapy strategies. Furthermore, our findings suggest that *F.n* may influence CRC progression through ARGs, which could serve as a clinical biomarker and therapeutic target for *F.n*-infected CRC patients.

INTRODUCTION

CRC is one of the most common malignancies of the digestive system, and is the second leading cause of cancer-related death worldwide. [1, 2]. The morbidity

rate of CRC continues to increase, with nearly 1.8 million individuals diagnosed and over 900,000 deaths each year [3]. Recently, the incidence and mortality of CRC in some European and Northern American countries have decreased. However, CRC incidence and

mortality continue to increase in China [4]. Chemotherapeutic regimens for the treatment of advanced CRC (e.g., oxaliplatin and 5-fluorouracil (5-FU)) have made substantial progress [5, 6]. However, the overall results of various targeted drugs currently used for the treatment of advanced CRC (such as the anti-EGFR agent cetuximab and the anti-angiogenesis agent bevacizumab) remain a challenge [7, 8]. Immunotherapy offers additional options and hope for the treatment of CRC patients. In fact, not all types of patients can benefit from immunotherapy [5]. There is therefore an urgent need to construct a novel valuable biomarker that can categorize different features of patients into distinct groups and predict the efficacy of immunotherapy in CRC.

A sufficient supply of oxygen and nutrients from the blood facilitates the survival and rapid growth of malignant tumors, which need ample vascularization to enter the circulatory system [9]. Thus, it is important to note that the initiation of tumor angiogenesis is a crucial factor in tumor development [10]. Anti-angiogenic therapy (including gastric cancer, non-small-cell lung cancer, renal cell carcinoma, and colorectal cancer) has been approved for the treatment of multiple cancers [11–14]. However, studies have revealed that anti-angiogenic therapy provides only a short-term remission and tumor growth inhibition before drug resistance is developed [9]. Immunotherapy, such as immune checkpoint inhibitors (ICIs) for PD-1, PD-L1, and CTLA-4, is a promising treatment modality for diverse tumors, whose safety and effectiveness have been proven by a growing body of clinical studies [15, 16]. Accumulative evidence shows that the TME is responsible for tumor aggressive behaviors and influences immunotherapy efficacy [17]. The TME is mainly composed of tumor cells, blood vessels, infiltrating immune, and stromal cells [18]. The formation of neovascularization characterized by continuous and disordered is a characteristic of the TME. Interestingly, cross-talk between the tumor cells and angiogenesis mediates an immunosuppression microenvironment to promote immune escape of tumor cells, which seriously interferes with anti-tumor immunity and is an important reason for promoting tumor progression [19]. Hence, a comprehensive analysis of the association between angiogenesis and the TME can be used to recognize different tumor immunophenotypes and improve the predictive power of immunotherapy.

At present, several ARGs have been found to be involved in CRC development. The large proteoglycan versican (VCAN) is one of the main components of the extracellular matrix, which is involved in cell adhesion, proliferation, migration, and angiogenesis [20, 21]. Recent studies have demonstrated a positive correlation between VCAN and VEGF expression and microvessel

density, with a worsened outcome in CRC patients with peritoneal metastasis [22]. Moreover, periosteum protein (POSTN), also known as osteoblast-specific factor 2, is a secreted extracellular matrix protein originally found in mesenchymal lineage cells (e.g., osteoblasts), and plays a key role in development and tissue regeneration [23]. There is a significant association between POSTN and the complete deletion of p53 in CRC tissues, and high POSTN expression is related to peritoneal and distant organ metastasis [24]. SERPINB5 (Serpin family B member 5) is a non-inhibitory member of the Serpin protease inhibitor superfamily with a variety of biological activities, including the regulation of cell adhesion, migration, apoptosis, oxidative stress response, and angiogenesis [25, 26]. SERPINB5 expression is significantly upregulated in CRC tissues, and is negatively correlated with progression-free survival of CRC patients [27, 28]. Dishearteningly, most current studies have focused on discovering the role of single ARG in CRC progression and prognosis. Moreover, the combination of the effects of multiple ARGs on TME infiltration characteristics of CRC have yet to be researched. Thus, a global analysis of the relationship between ARGs and TME, especially the tumor immune microenvironment, may provide the possibility of combining targeted therapy and immunotherapy to boost the predictive power of immunotherapy. On the other hand, reshaping the tumor immune microenvironment through a comprehensive understanding of the cross-talk between tumor angiogenesis and immune cells in the TME will contribute to the development of a strategy for a long-lasting anti-tumor immunity response.

Here, a comprehensive bioinformatic analysis of ARGs in CRC was performed by using TCGA and GEO databases. First, 1109 CRC patients were stratified into two discrete subtypes based on the ARGs expression. According to the differentially expressed genes (DEGs) identified in the two angiogenesis subtypes, we then identified three distinct gene subtypes and evaluated their molecular genetic features and prognostic value, as well as the abundance of infiltrating immune cells. We further established an ARG_score that precisely predicted the RFS of CRC patients and immunotherapeutic response. Finally, we observed evidence of a potential relationship between KLK10, *F.n.*, and CRC based on whole genome microarray analysis. To further verify this association, we cultured two CRC cell lines *in vitro* with or without *F.n.* The results confirmed that the mRNA and protein levels of KLK10 were significantly upregulated by *F.n.* infection at different time intervals. These findings not only extend the research community's current understanding of ARGs in the CRC immunotherapy field but also provide a promising clinical biomarker and potential therapeutic target for *F.n.*-infected CRC patients.

We hope that our results will be helpful to the development of effective CRC immunotherapies.

MATERIALS AND METHODS

Data sources preparation and preprocessing

The microarray datasets that investigated the gene expression of CRC tissues were downloaded from the Gene Expression Omnibus (GEO) database, including GSE39582, GSE17536, and GSE38832. The mRNA expression profiles of 568 CRC tissues and 44 normal tissues were obtained from The Cancer Genome Atlas (TCGA) database. In addition, the relevant clinical data of the CRC patients were acquired using TCGA on November 8, 2021. The four datasets were preprocessed via the elimination of batch effects using the “ComBat” algorithm of the “SVA” package [29]. CRC patients with missing survival values were removed from further analysis. A total of 36 ARGs were gained from the MSigDB Team (Hallmark Gene set) (Supplementary Table 1).

Unsupervised clustering analysis

Based on the expression of 36 ARGs, patients were identified as different molecular subtypes by k-means using the “ConsensusClusterPlus” package in R [30, 31]. In order to ensure the stability of the classification, the experiment was repeated 1000 times. Principal component analysis (PCA) was performed for the different clustering subtypes. The “GSVA” package in R was then used to carry out GSVA enrichment analysis [32].

Association of molecular patterns with clinicopathologic factors and prognosis in CRC

To identification the clinical significance of the two subgroups produced by cluster analysis, we compared the relationships between molecular patterns and clinicopathological features. Furthermore, we evaluated the significance of RFS in distinct subtypes by using the “survival” and “survminer” packages in R [33].

Evaluation of TME infiltration and immune checkpoints between different molecular patterns

The TME scores were assessed using the ESTIMATE algorithm, and TME cell infiltration was calculated using the CIBERSORT algorithm [34]. We also assessed the infiltrating immune cell scores and immune-related pathway activity by using the ssGSEA algorithm [35]. Furthermore, the association between the two subtypes on 47 immune checkpoints expression was analyzed [36–39].

Identification of DEGs and functional enrichment analysis

The differentially expressed genes (DEGs) were identified by using the “limma” R package between the different angiogenesis subtypes ($|\log \text{Foldchange (FC)}| > 1$ and false discovery rate (FDR) < 0.05). Further, the functions of these DEGs were performed by Gene Ontology (GO) analyses with the “clusterProfiler” package in R, which included biological process (BP), cellular component (CC), and molecular function (MF). Similarly, Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were also conducted to identify the significant enrichment pathways of these DEGs by the “clusterProfiler” package in R [40].

Construction and validation of the angiogenesis-associated prognostic ARG_score

The ARG_score was constructed to evaluate the angiogenesis patterns of CRC patients. The DEGs expression data for different subtype groups (angiogenesis cluster A and angiogenesis cluster B) were standardized in CRC specimens and the intersect genes were selected. We first performed univariate Cox regression analysis to screen those DEGs associated with CRC RFS. Then, a deeper analysis was performed using unsupervised clustering method based on the expression of prognostic DEGs to classify patients into angiogenesis gene subtype A, angiogenesis gene subtype B, and angiogenesis gene subtype C. Subsequently, the CRC patients were randomly clarified into training ($n = 555$) and testing ($n = 554$) sets at a ratio of 1:1, and the former was selected to construct the angiogenesis-related prognostic ARG_score. In short, we performed the Lasso Cox regression algorithm to minimize the risk of overfitting by using the “glmnet” package in R based on prognostic genes associated with angiogenesis. We established a predictive model using LASSO Cox regression analysis: $\text{ARG_score} = \sum \text{Exp}_i \times \beta_i$, where Exp_i and β_i represented each genes expression level and each genes coefficient index, respectively.

RNA extraction, real-time PCR (RT-qPCR), and oligonucleotide transfection

The Trizol reagent (Invitrogen, Carlsbad, CA) was used to extract the total RNA from CRC cells (HCT116 and HT29) and tissues (six pairs CRC and nearby non-tumor tissues). Then, the PrimeScript RT reagent Kit (Takara, Japan) was performed to generate cDNA from 1 μg of total RNA. Finally, mRNA transcripts were quantified with a CFX-96 instrument RT-qPCR System (Bio-Rad Laboratories) and SYBR Green assays (Takara). SiRNA was purchased from GenePharma (Shanghai, China),

and Lipofectamine 3000 (Invitrogen, USA) was used to perform transfection of siRNA. Moreover, nonspecific siRNA was used as negative controls, according to the manufacturer's protocol. The $2^{-\Delta\Delta CT}$ method was used to evaluate the relative target genes expression levels, normalizing with β -actin. The sequences of the primers used are shown in Supplementary Table 2.

Clinical significance and stratification analysis of the prognostic ARG_score

The clinical characteristics of the CRC patients were isolated from the GEO and TCGA datasets. Then, a stratified analysis was performed to assess the predictive ability of the ARG_score in different subgroups. Further, univariate and multivariable Cox regression analyses were used to confirm the independent prognostic value of the ARG_score in training, testing, and entire sets. In addition, we examined the correlations between ARG_score and the 22 immune cells infiltrating levels, 47 immune checkpoints, tumor mutation burden (TMB), microsatellite instability (MSI), and cancer stem cells (CSC) index.

Development of a nomogram for prediction

To explore independent prognostic value of the ARG_score, we established a nomogram prediction model for predicting the 1-year, 5-year, and 10-year RFS [41]. Moreover, the calibration maps described the precision of the nomogram in prognosis prediction (There are 1000 duplicate bootstrap methods).

Mutation and drug sensitivity analysis

The TCGA-COAD/READ database was used to draw the 36 ARGs mutation frequency and oncoplot waterfall plot by using the "maftools" package in R [42]. Further, the location of CNV alteration of 36 ARGs on 23 chromosomes was produced by using the "Rcircos" package in R [43]. The semi-inhibitory concentration (IC50) values of chemotherapeutic drugs, which commonly used for CRC treatment were estimated by using the "pRRophetic" package in R [44]. Moreover, we used anti-angiogenic therapy and immunotherapy clinical trials (the gene expression profile and clinical manifestations of publicly available datasets) to explore the prediction ability of ARG_score for the therapeutic responses of tumor patients (GSE53127: a phase III study of metastatic CRC patients treated with bevacizumab [45], GSE103668: a phase II study of bevacizumab vs. platinum in patients with triple negative breast cancer [46], GSE78220: a phase III study of nivolumab vs. pembrolizumab in patients with previously treated metastatic melanoma [47], GSE126044: a phase II study of nivolumab vs.

pembrolizumab in patients with non-small cell lung cancer [48], E-MTAB-3218 (ArrayExpress database, <https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-3218?accession=E-MTAB-3218#>): a phase III study of nivolumab in patients with previously treated metastatic renal cell carcinoma [49]). The transcriptional information was adjusted and normalized by using the "edgeR" package in R [50], and the data were converted by using "voom" in the "limma" package in R [51]. Meanwhile, we also collated the prognostic and immunotherapy information of patients.

Bacteria strains and cell lines

The *Fusobacterium nucleatum* strain ATCC 25586, purchased from American Type Culture Collection (ATCC, Manassas, VA, USA), was cultured overnight in brain heart infusion (BHI) broth under anaerobic conditions at 37°C. Human colorectal cancer cell lines HCT116 and HT29 were purchased from ATCC and cultured in McCoy's 5A (Gibco, USA) supplemented with 10% fetal bovine serum (FBS) at 37°C in a humidified 5% CO₂ atmosphere. Cells were seeded in 6-well plates and then incubated with or without bacteria at a multiplicity of infection (MOI) 100:1 [52, 53].

Western blot analysis

Firstly, the cells were washed twice with ice-cold PBS and collected cell extracts with RIPA lysis buffer supplemented with Halt™ Protease Inhibitor Single-Use Cocktail (Thermo Fisher Scientific). Then, protein concentrations were quantified using BCA Protein Assay Kit (Thermo Fisher Scientific). Further, an equal amount of protein (10 μ g) was separated by 10% SDS-PAGE and transferred to PVDF membranes (Invitrogen). The membranes were blocked in 1× TBS, 0.1% Tween-20 detergent (TBST) (Thermo Fisher Scientific) with 5% non-fat dry milk at room temperature for 2 h and then incubated with primary antibodies (E-cadherin (3195; Cell Signaling Technology), N-cadherin (13116; Cell Signaling Technology), Vimentin (5741; Cell Signaling Technology), β -actin (8457; Cell Signaling Technology)) overnight at 4°C. Next, after washing with TBST, the membranes were incubated with secondary antibodies for another 1 h at room temperature. Finally, the enhanced chemiluminescent (ECL) assay kit (Thermo Scientific, PA, USA) was applied for film visualization.

Cell proliferation assay and wound healing assay

Firstly, approximately 1×10^4 cells in 100 μ l of medium were seeded in 96-well plates. Then, 10 μ l of Cell Counting Kit-8 (CCK8, Yeasen, China) solution was

added to each well at 24, 48, and 72 h. After an additional incubation period at 37°C for 3 h, the OD value was measured at 450 nm using a microplate reader. The wound healing assay was performed as previously described [54]. Briefly, the two cell lines (1×10^5) were cultured in 6-well plates. The complete medium was replaced with a low concentration of serum fresh medium (2%) after 16 h. A 200- μ l pipette tip was used to scratch across each well and wash gently with PBS two times. Meanwhile, multiple location markers were performed at the inoculated cells surface to compare the same wound area in future. The scratch area was photographed with an inverted microscope at 0, 24 and 48 h, respectively. The experiment was conducted in triplicate.

Statistical analysis

R (v4.3.1), GraphPad Prism (version 10.0), and SPSS (27.0) were used to perform all statistical analyses. A *p*-value less than 0.05 was considered statistically significant.

Availability of data and materials

In this study, publicly available datasets were used to perform analyses, as described in the materials and methods. The datasets are available from TCGA (<https://portal.gdc.cancer.gov/>) and Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>), including GSE39582, GSE17536, and GSE38832.

RESULTS

Overview of the genetic and transcriptional alterations of ARGs in CRC

Firstly, the somatic mutations pattern of 36 ARGs in CRC cohort (COAD and READ samples) was investigated. The Figure 1A demonstrated that 157 (39.35%) had mutations of these ARGs in 399 COAD cohort. Notably, VCAN was the gene with the highest mutation frequency (10%), however, there were no mutations in two ARGs (CCND2 and TIMP1). In addition, the Figure 1B showed that 34 (24.82%) had mutations of these ARGs in 137 READ cohort, and VCAN was also the gene with the highest mutation rate (9%). Further, the protein–protein interaction (PPI) analysis was performed to construct, and indicated that VEGFA, SPP1, POSTN, VTN, COL3A1, ITGAV, and TIMP1 were the hub genes (Figure 1C). We then performed to compare the mRNA expression levels between normal and tumor tissues. A total of 26 DEGs were determined, most of which were rich in the tumor tissues (Figure 1D). Moreover, the somatic copy number alterations of ARGs were investigated, and the copy number variation (CNV) alterations on their

respective chromosomes were demonstrated in Figure 1E. As shown in Figure 1F, 36 ARGs demonstrated evident CNV alterations. Notably, there was a positive association between the levels of ARGs and CNV alteration; for example, the expression levels of PRG2, SERPINA5, and THBD were low in tumor tissues, while VEGFA, PF4, and PTK2 were expressed at high levels. These results indicate that CNV may modulate the expression of ARGs in CRC. According to clinical characteristics, the CRC patients were divided into early-stage groups (I + II) and advanced-stage groups (III + IV). As shown in Figure 1G, there were significant differences between the two groups in 36 ARGs. The expression levels of 12 ARGs (VEGFA, VCAN, POSTN, STC1, COL5A2, ITGAV, COL3A1, SPP1, OLR1, PTK2, TIMP1, and JAG2) were higher in both CRC tissues and the advanced-stage groups, which indicated that their potential function as carcinogenic genes in CRC.

Identification of angiogenesis patterns in CRC

Supplementary Figure 1 displays the article framework and workflow. Our study integrated 1109 patients from four eligible CRC datasets to identify potential angiogenesis patterns of CRC. Detailed clinical information of 1109 patients are presented in Supplementary Table 3. We extracted the expression levels of 36 ARGs and then revealed their prognostic value using uniCox and Kaplan–Meier analysis [55] (Supplementary Table 4). Next, we established the angiogenesis network to visualize the landscape of 36 ARGs, including their interactions, connections, and prognostic values (Figure 2A, Supplementary Table 5). Based on these ARGs, we identified two different angiogenesis-related patterns by consensus clustering, including 612 cases in Cluster A and 497 cases in Cluster B (Figure 2B). Further, PCA analysis confirmed that the CRC patients were well-differentiated when $k = 2$, indicating the robust and reliable clustering of the samples (Figure 2C). The Kaplan–Meier curves demonstrated that the survival outcome of Cluster A was better than that of Cluster B ($P < 0.001$, Figure 2D). Furthermore, the heatmap was used to visualize the relationship between 36 ARGs expression and clinical features. Compared to Cluster B, we found that Cluster A was markedly associated with early pathological stage ($P < 0.001$), left-sided CRC ($P < 0.01$), without BRAF mutations ($P < 0.001$), and low risk of recurrence ($P < 0.001$, Figure 2E).

Variation in characteristics of TME infiltration between two angiogenesis subtypes

GSVA enrichment analysis showed that cluster B was significantly enriched in cancer-associated pathways

(e.g., renal cell carcinoma, glioma, and small cell lung cancer), metastasis-associated pathways (e.g., cell adhesion molecules cams, ECM receptor interaction,

and focal adhesion), and immune fully-activated pathways (e.g., B cell receptor signaling pathway, cytokine receptor interaction, chemokine signaling

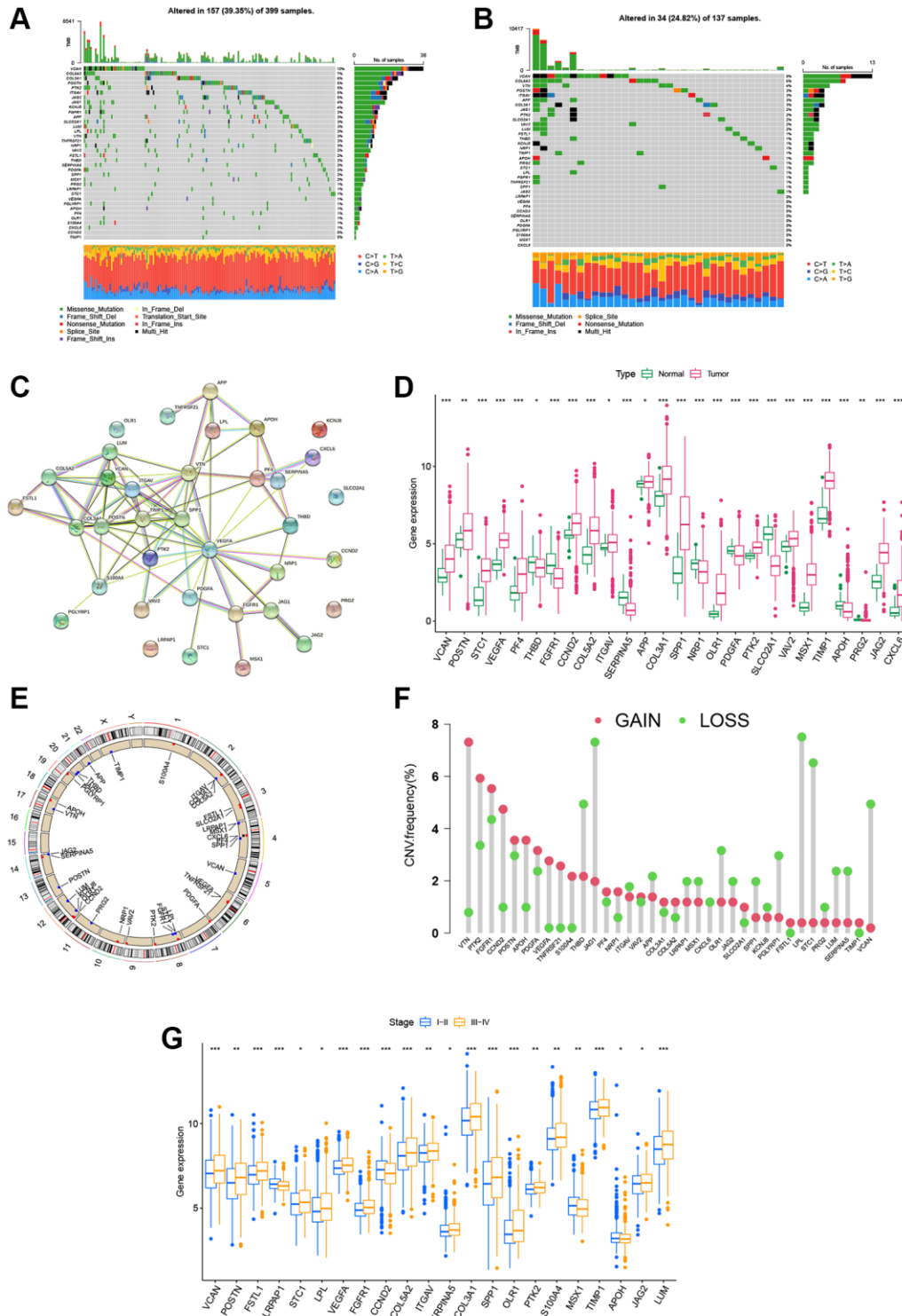


Figure 1. Landscape of the ARG genetic alterations in CRC. (A, B) Mutation frequencies of 36 ARGs in 399 and 137 patients with COAD and READ based on the TCGA cohort, respectively. (C) PPI network showing the interactions of ARGs (the minimum required interaction score was 0.4). (D) Expression distributions of DEGs between tumor and normal tissues. (E, F) Locations of CNV alterations in ARGs on 23 chromosomes and the frequencies of CNV gain, loss, and non-CNV among ARGs, respectively. (G) Expression distributions of DEGs between the high- and low-stage groups. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

pathway activation, and the NOD-like and Toll-like receptor signaling pathways) (Figure 3A). To further investigate the correlation between ARGs and the TME

of CRC, we analyzed the infiltrating levels of 23 human immune cell subsets in distinct subtypes with the CIBERSORT algorithm. We observed that, except for

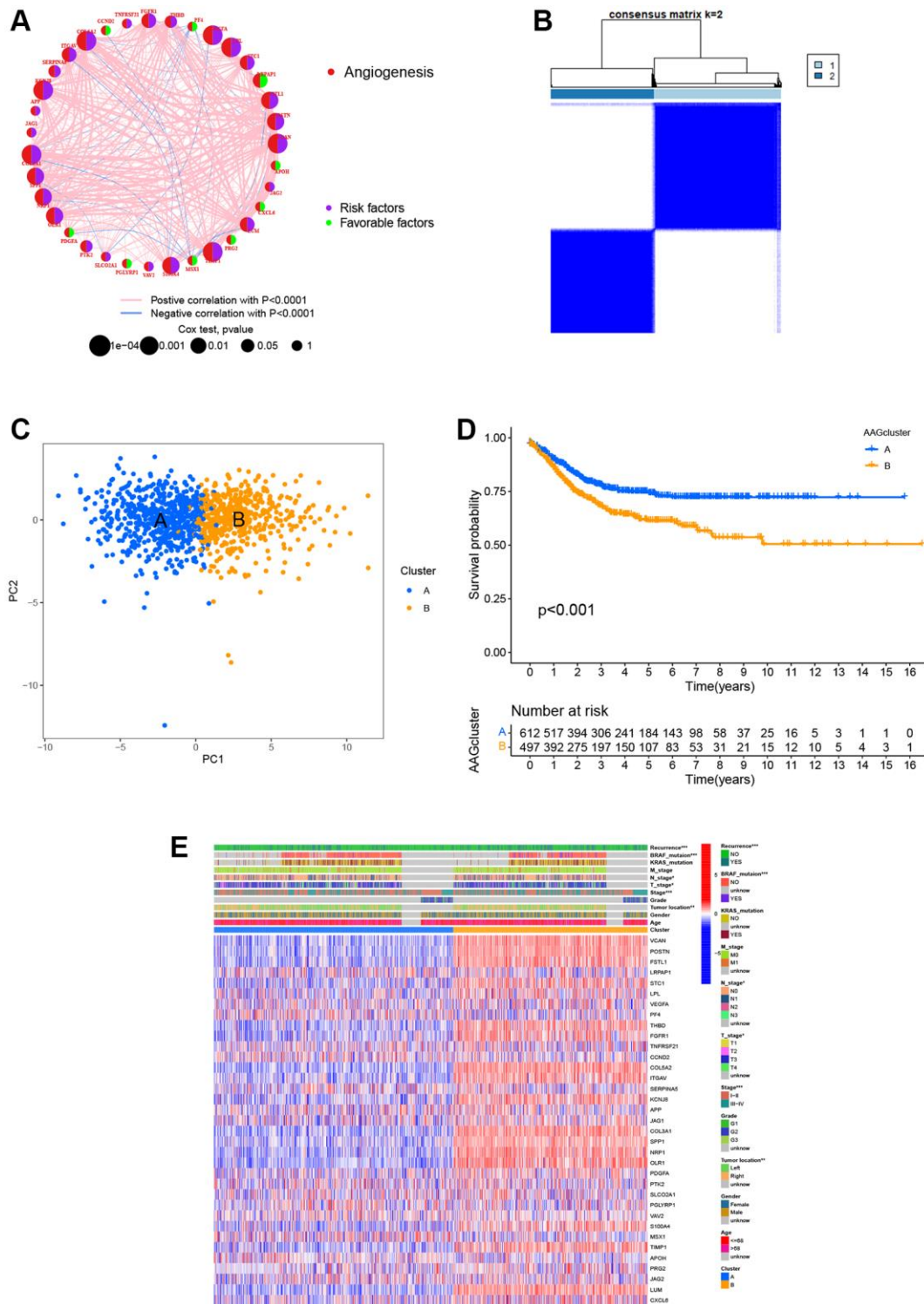


Figure 2. ARG-related subtypes and clinicopathological and biological characteristics of two distinct subgroups of samples divided by consistent clustering. (A) The correlation network of ARGs in CRC (red line: positive correlation; blue line: negative correlation). (B) Consensus matrix heatmap defining two clusters ($k = 2$) and their correlation area. (C) PCA showing a remarkable difference in transcriptomes between the two subtypes. (D) Univariate analysis showing 36 ARGs related to the RFS time. (E) Differences in clinicopathological and biological characteristics of the two distinct subtypes. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

the monocyte cells, the other 22 types of immune cells were all poorly activated in Cluster A (Figure 3B). Further, we analyzed the differences in terms of the immune score, stromal score, and estimate score between Cluster A and Cluster B using the ESTIMATE algorithm. Our results revealed that Cluster B had higher TME scores than Cluster A, which indicates that Cluster B had a higher levels of angiogenesis-dependent extracellular matrix components ($P < 0.001$, Figure 3C). Moreover, we analyzed the expression of 47 important immune-oncology targets between Cluster A and Cluster B. As shown in Figure 3D, significant differences were observed in 36 immune checkpoints between the two subtypes, including PD-L1, CTLA-4, and LAG3 ($P < 0.001$). Altogether, our findings suggest that ARGs were involved in the formation of TME infiltration and represent different prognostic features in patients of CRC.

Generation of gene subtypes based on two angiogenesis clusters

To confirm the biological behavioral differences between these two angiogenesis subtypes, 927 angiogenesis clusters related DEGs were obtained using the “limma” package in R to obtain insights into their biological function. These angiogenesis subtypes related DEGs were mainly enriched in biological metastasis processes (Figure 4A). KEGG analysis showed enrichment of immune-, cancer- and metastasis-related pathways (Figure 4B), indicating that angiogenesis plays a vital role in the immune regulation of the TME and the modulation of tumor metastasis. Next, uniCox analysis was performed to explore the prognostic value of these DEGs and 620 genes related to RFS time were extracted ($P < 0.05$, Supplementary Table 6). To further understand specific regulation mechanism, a consensus

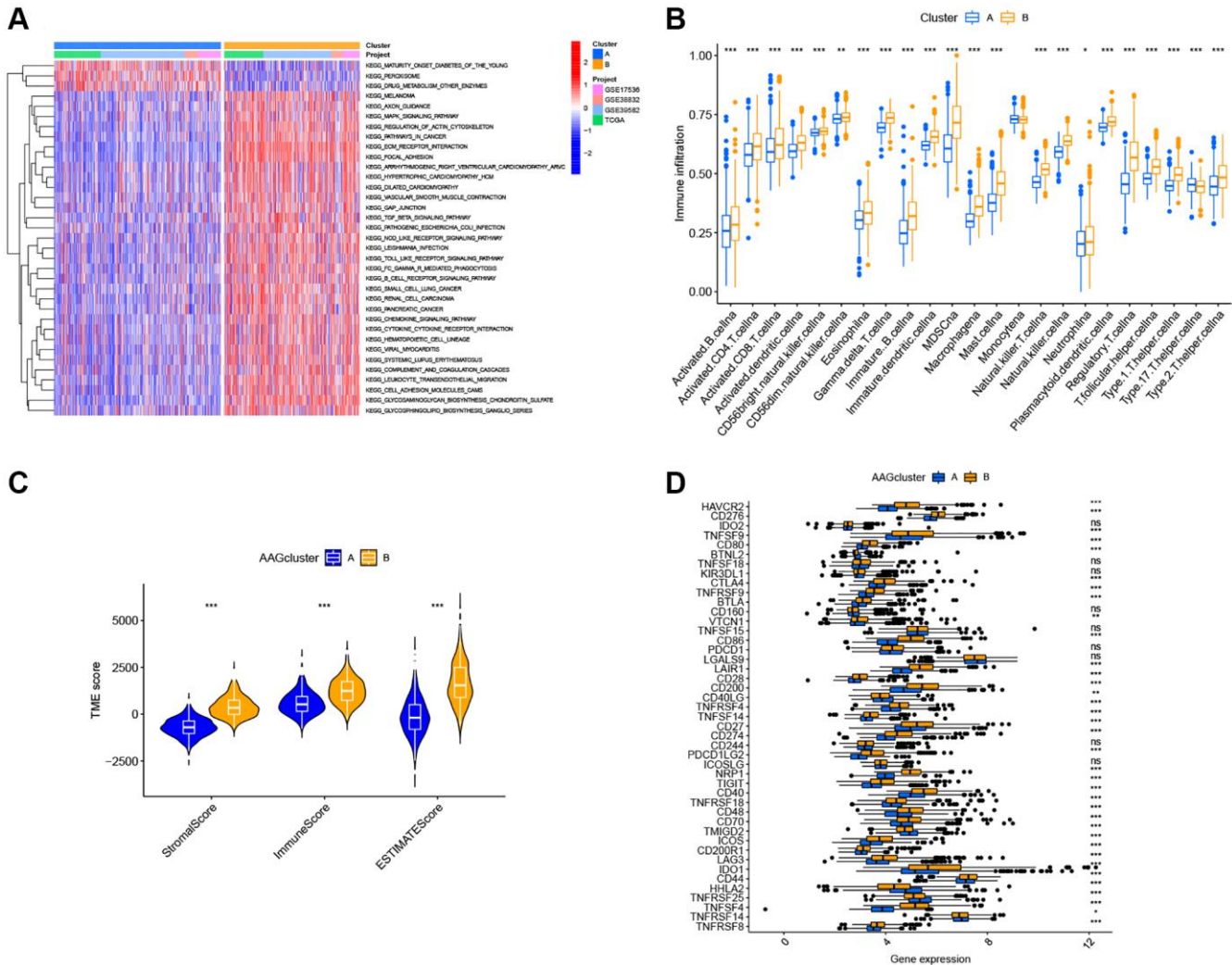


Figure 3. Two ARG subtypes showed diverse tumor immune cell microenvironments. (A) Biological processes analyzed by GSEA showing the active biological pathways in distinct subtypes. (B) The abundance of 23 TME infiltrating cells between the two subtypes of ARGs. (C) Correlations between the two ARG subgroups and the TME score. (D) Expression levels of 47 immune checkpoints in the two subtypes. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

angiogenesis-related gene clusters demonstrated significant differences in ARG expression, which were consistent with the expected results of the angiogenesis clusters (Figure 4E).

Construction and validation of the prognostic ARG_score in CRC

The ARG_score was constructed based on the DEGs associated with angiogenesis clusters. First, we randomly assigned the patients into training ($n = 555$) and testing ($n = 554$) groups at a ratio of 1:1. LASSO and multivariate Cox analyses for 620 angiogenesis cluster-related prognostic DEGs were then used to identify an optimal prognostic signature (Supplemen-

tary Figure 3). A total of 9 RFS-associated genes were identified based on the minimum partial likelihood deviance, and the ARG_score for each patient in the training, testing, and entire datasets were calculated based on the risk formula, as follows: risk score = (expression of SLC2A3 \times 0.2758) + (expression of SCG2 \times 0.1617) + (expression of KDR \times 0.3094) + (expression of MMP11 \times 0.1725) + (expression of CXCL10 \times -0.2112) + (expression of CXCL13 \times -0.0965) + (expression of SPINK1 \times -0.1023) + (expression of KLK10 \times 0.0877) + (expression of MMP3 \times -0.1427). A Sankey diagram was constructed to visualize the patients' distribution in the two angiogenesis clusters, three gene clusters, and two ARG_score groups (Figure 5A). In addition, significant

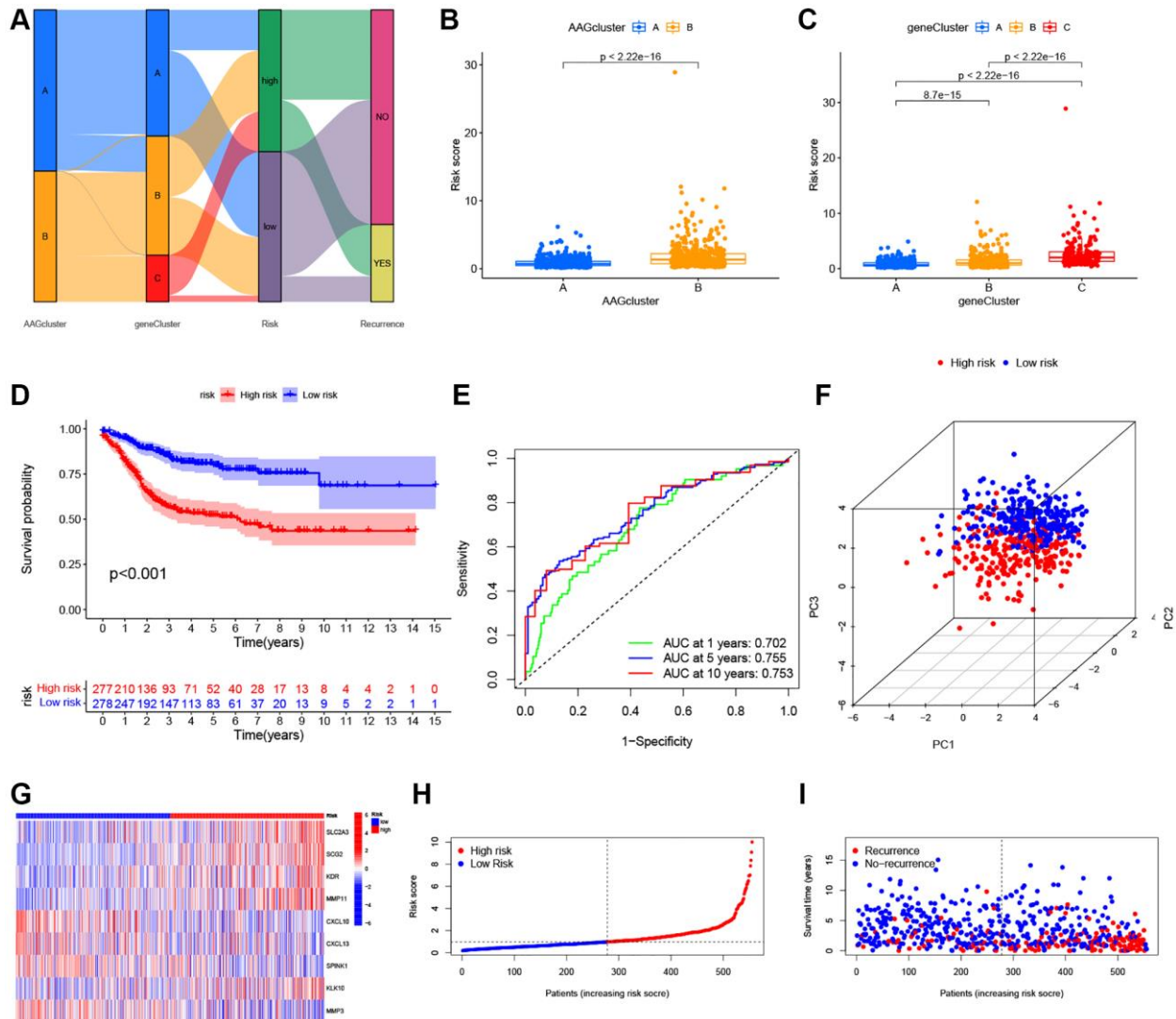


Figure 5. Construction of the ARG_score in the training set. (A) Alluvial diagram of subtype distributions in groups with a different ARG_score and prognosis. (B) Differences in ARG_score between the two ARGs clusters. (C) Differences in ARG_score among the three gene subtypes. (D) Kaplan–Meier analysis of the RFS between the low- and high-ARG_score groups. (E) ROC curves to predict the sensitivity and specificity of 1-, 5-, and 10-year survival according to the ARG_score. (F) PCA analysis based on the prognostic model. (G–I) The distribution of the risk score, survival time, and survival status, and the heatmap of the 9 selected prognostic genes between the high- and low-ARG_score groups, respectively.

differences were observed in the ARG_score of the angiogenesis clusters and gene clusters (Figure 5B, 5C). The highest and lowest ARG_score was observed in gene_Cluster C and gene_Cluster A, respectively, suggesting a low ARG_score may be closely associated with immune activation-associated features. According to the results of survival analysis, we found that the higher ARG_score of both classifications were associated with a worse RFS. Moreover, Kaplan-Meier analysis implied that low ARG_score patients had a better RFS over patients with a high ARG_score in the training cohort ($P < 0.001$, Figure 5D), and the AUCs of the 1-, 5-, and 10-year RFS were 0.702, 0.755, and 0.753 in the training cohort, respectively (Figure 5E). PCA analysis demonstrated a clear distribution between the low- and high-ARG_score groups (Figure 5F). The relationships between the 9 selected prognostic signatures and the ARG_score can be witnessed in the heatmap (Figure 5G). Meanwhile, the distribution of the ARG_score and RFS time of patients is shown in Figure 5H, 5I. To assess the predictive robustness of the risk model, the ARG_score of the testing cohort and the entire cohort were obtained (Supplementary Figures 4, 5). Based on the median score of the training cohort, we also separated patients into different risk subgroups. Also, survival analysis revealed a superior RFS of low score patients compared to high score patients ($P < 0.001$). Further, the classification efficiency analysis of 1-, 5- and 10-years survival prediction showed that ARG_score had a high AUC values, suggesting that the model had the ability to evaluate the survival of CRC patients.

Validation of the 9 prognostic ARGs in CRC tissues

To more effectively verify our findings, the expression levels of 9 prognostic signatures were measured in 6 pairs of tumor tissues and normal adjacent tissue samples by RT-qPCR. As shown in Supplementary Figure 6I–6O, the expression levels of SLC2A3, MMP11, CXCL10, KLK10, and MMP3 were upregulated, while those of SCG2 and CXCL13 were downregulated in CRC tissues compared to normal tissues. No statistically significant difference was observed in the mRNA levels of KDR and SPINK1 between the normal and tumor tissues (data not shown). These experimental results were consistent with those predicted by the bioinformatics methods (Supplementary Figure 6A) and the GEPIA database (Supplementary Figure 6B–6H).

Stratified analysis and independent prognostic value of ARG_score

To determine the influence of the ARG_score on clinical features, the association between ARG_score

and clinical factors was investigated in CRC patients, including age, gender, TNM stage, tumor location, KRAS, and BRAF mutation status. We found a higher ARG_score in patients in the stage III–IV subgroup than those in the stage I–II subgroup ($P < 0.001$, Supplementary Figure 7A). To determine whether the ARG_score could be used as an independent prognostic factor, we carried out univariate and multivariate Cox regression analyses by using ARG_score and multiple clinical characteristics. As shown in Supplementary Figure 7B–7D, ARG_score was an independent predictor of CRC outcome in the training cohort, with consistent results observed across the testing cohort and the entire cohort ($P < 0.001$). To further investigate the prognostic significance of the ARG_score in CRC, we performed a stratified analysis in different clinical parameter subgroups. As shown in Supplementary Figure 8, CRC patients in the high ARG_score group tended to have a shorter RFS than that in the low ARG_score group ($P < 0.001$), as demonstrated by the results for age, sex, tumor location, TNM stage, and KRAS and BRAF mutations.

Establishment of a clinical nomogram for the prediction model in CRC

Given the high correlation between the ARG_score and patients' prognosis in CRC, we further performed to establish a nomogram by using multivariate Cox regression to predict the 1-, 5-, and 10-year RFS (Figure 6A). Calibration plots showed that 1-, 5-, and 10-year RFS could be predicted relatively well (Figure 6B). Further, the AUC values of the nomogram with that of the TNM stage were estimated for predicting the 1-, 5-, and 10-year RFS, respectively. Compared with TMN staging features, the prediction ability of ARG_score in the nomogram was superior (Figure 6C–6E).

Relationship between ARG_score and TME infiltration in CRC

There is increasing evidence that TME cell infiltration is critical to tumor development and therapeutic response. Thus, we used the CIBERSORT algorithm to investigate the abundance of different immune cell types between the low- and high-ARG_score groups. As shown in Figure 7A, the ARG_score was positively correlated with the infiltration of M0 and M2 macrophages, neutrophils, regulatory T cells (Tregs), and activated mast cells, while being negatively correlated with the infiltration of M1 macrophages, naïve B cells, follicular helper T cells, CD8 + T cells, CD4 + memory-activated T cells, plasma cells, activated NK cells, and resting NK cells. In addition, a low ARG_score was also closely related to a high immune score, whereas a high ARG_score was linked

to a high stromal score (Figure 7B). We then evaluated the potential relationship between the 9 selected genes in the prognostic signature and the abundance of immune cells, and found that most immune cells were significantly associated with these 9 genes (Figure 7C). Furthermore, we evaluated the correlations between 47 immune checkpoints and this prognostic signature. As shown in Figure 7D, 30 immune checkpoints were discrepantly represented in the two groups, including PD-L1, CTLA-4, and LAG3.

The ARG_score had observable correlations with the TMB, MSI, and CSC indices

Several studies have reported that patients with high TMB or MSI-H status may be more sensitive to immunotherapy [5, 56, 57]. As shown in Figure 8A, a higher TMB was observed in the low ARG_score groups compared to the high ARG_score groups ($P < 0.05$), indicating that low ARG_score patients may benefit more from immunotherapy drugs. In addition, our findings indicate that the ARG_score was negatively correlated with TMB based on Spearman correlation analysis ($R = -0.14$, $P < 0.05$, Figure 8B).

We further evaluated the potential relationship between TMB status and RFS, and found significant differences in clinical outcome between patients with high and low TMB groups ($P < 0.05$, Figure 8C). Meanwhile, this prognostic benefit in the high TMB groups was more affected by the ARG_score than that of the low TMB groups ($P < 0.01$, Figure 8D). Furthermore, correlation analysis showed that the high ARG_score group was associated with a microsatellite stable (MSS) state, while the low ARG_score group was associated with a high microsatellite unstable (MSI-H) state (Figure 8E, 8F). This result also implies that the low ARG_score groups benefitted from immunotherapy, with consistent results observed in the TMB analyses. Furthermore, the relationships between ARG_score and cancer stem cell (CSC) index was analyzed. As shown in Figure 8G, the results showed that ARG_score was negatively correlated with CSC index values ($R = -0.48$, $P < 0.001$), meaning that a lower ARG_score was associated with more pronounced stem cell characteristics and lower degree of cell differentiations. Additionally, the distribution differences of the top 20 somatic mutations genes among distinct ARG_score groups were analyzed based on TCGA-COAD cohort.

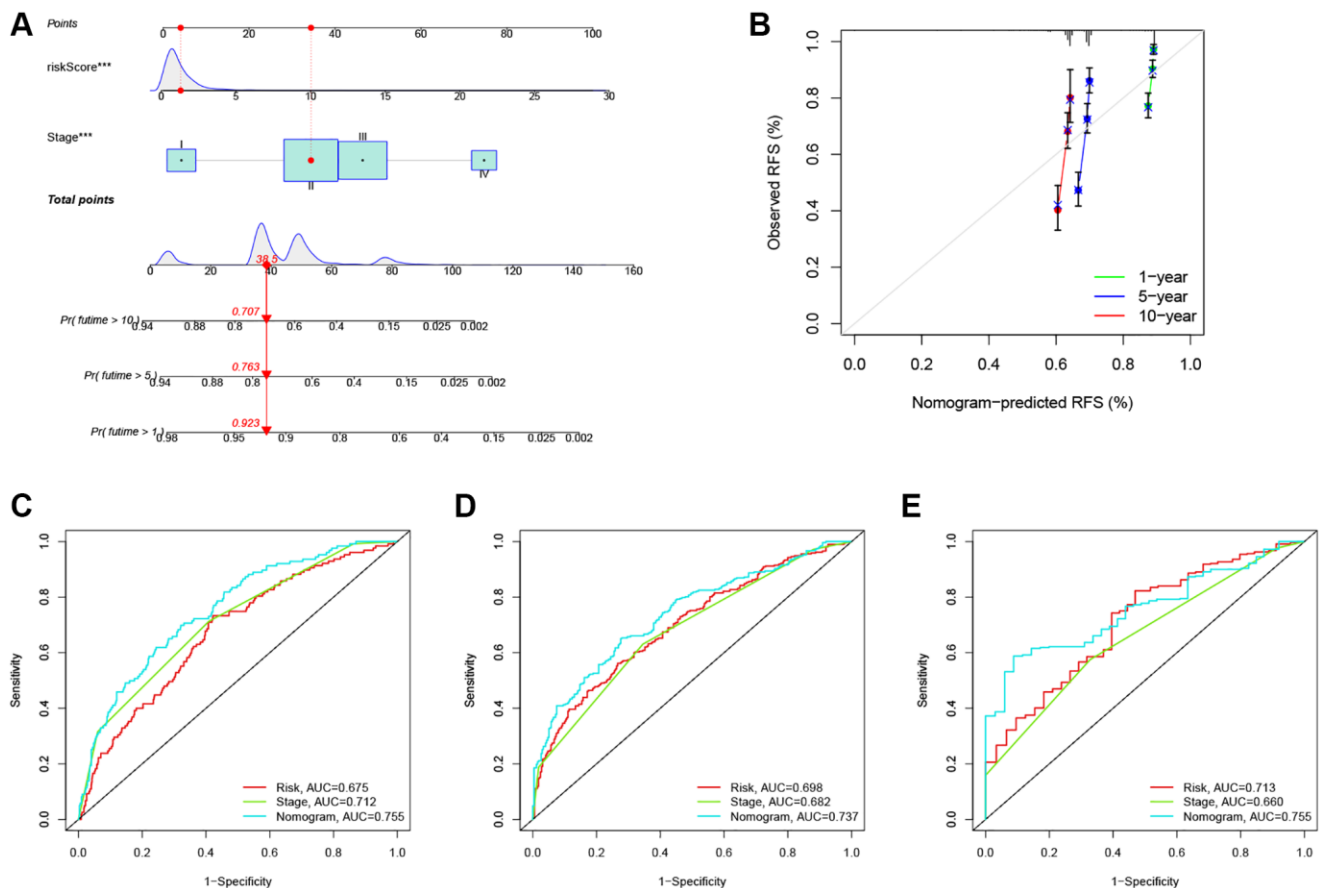


Figure 6. Construction and evaluation of a predictive nomogram. (A) The nomogram predicts the probability of the 1-, 5-, and 10-year RFS. (B) The calibration plot of the nomogram predicts the probability of the 1-, 3-, and 5-year RFS. (C–E) The time-dependent ROC curves of the nomograms compared for the 1-, 5-, and 10-year RFS in CRC, respectively. *** $P < 0.001$.

As shown in Figure 8H, the mutation incidences of APC, TP53, TTN, KRAS, MUC16, SYNE1, PIK3CA, and FAT4 were higher than or equal to 20% in the high ARG_score groups. Surprisingly, except for the USH2A, all other mutated genes were higher than or equal to 20% in the low ARG_score groups (Figure 8I). Taken together, our results indicate that these genes were more likely to be mutated in the low ARG_score groups compared to the high ARG_score groups.

Screening of the sensitive molecular drugs and prediction of the immunotherapy efficacy

Currently, CRC patients usually receive chemotherapy, targeted therapy, and immunotherapy after radical surgery, which has been shown to significantly improve clinical outcomes [58]. To explore the potential therapeutic drugs of patients in the low- and high-ARG_score groups, we analyzed the IC50 values of

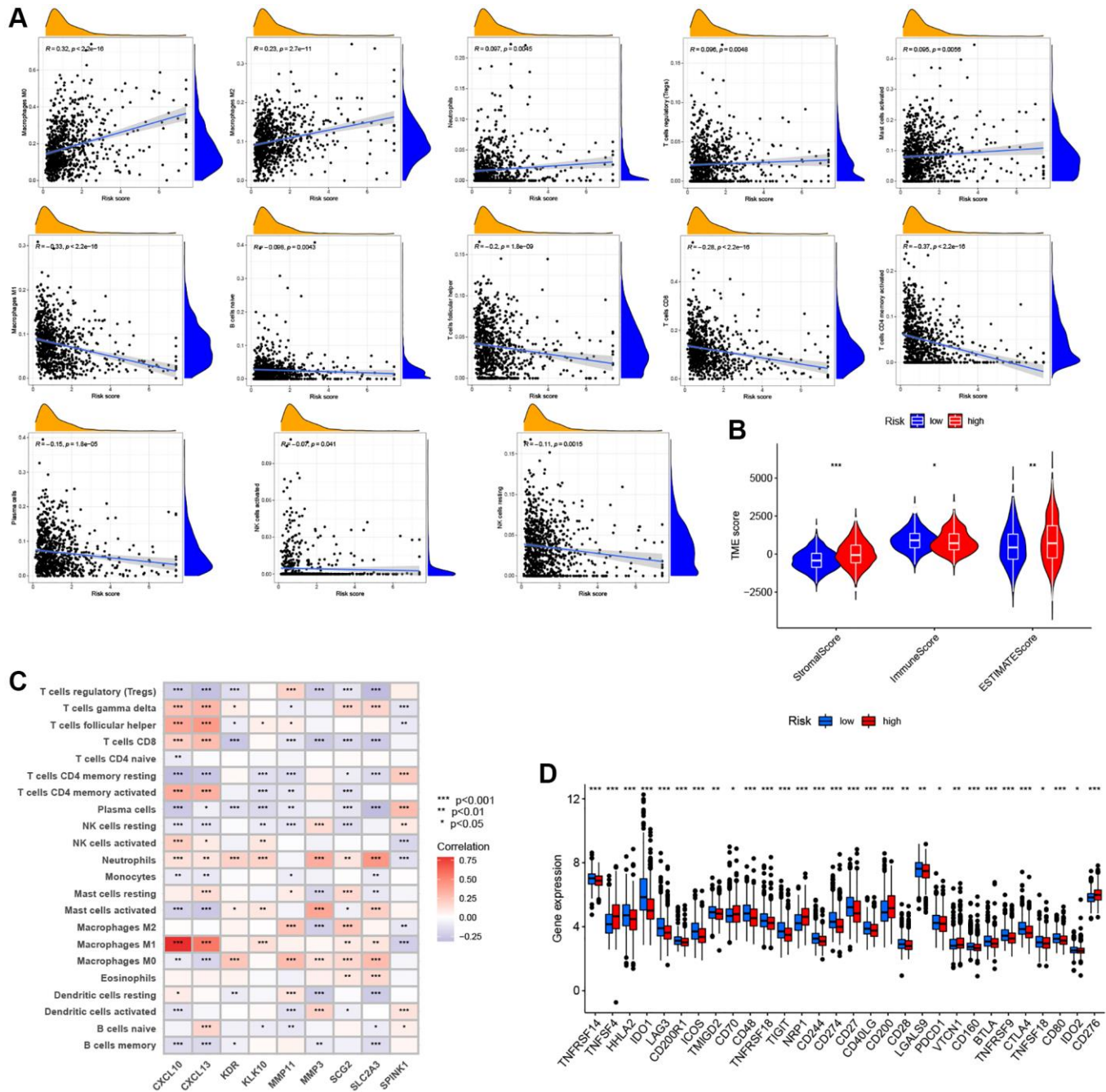


Figure 7. Evaluation of the TME and immune checkpoints between the distinct ARG_score groups. (A) Correlations between ARG_score and immune cell types. **(B)** Correlations between ARG_score and both immune and stromal scores. **(C)** Correlations between the abundance of immune cells and 9 selected genes in the proposed model. **(D)** Expression of 47 immune checkpoints in the low- and high-ARG_score groups. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

common therapeutic drugs in CRC patients. Our data showed that the low ARG_score groups may positively react to cisplatin, ATRA, gefitinib, sunitinib, gemcitabine, obatoclox, mesylate, paclitaxel, camptothecin, and bosutinib, while the high ARG_score groups may respond better to shikonin, 11asatinib, erlotinib, imatinib, lapatinib, and nilotinib (Figure 9A). The 15 associated candidate drugs were ranked by the *p*-value of differences, and top 6 most associated molecule drugs were screened for CRC patients. The 3D structures of ATRA, gemcitabine, camptothecin, shikonin, imatinib, and 11asatinib were displayed through the PubChem database (Figure 9B). The vascular endothelial growth factor (VEGF) pathway has

been shown to play a critical role in the control of CRC angiogenesis [59]. Anti-angiogenic drugs targeting the VEGF signaling pathways, including bevacizumab, has been shown to be an effective and tolerable therapy that improves survival in advanced CRC patients [60]. We further investigated the ability of the ARG_score to predict response to bevacizumab therapy in the GSE53127 cohort. As shown in Figure 9D, the high ARG_score groups had a higher objective response rate compared to the low ARG_score groups ($P < 0.001$). Also, in the GSE103668 cohort, the proportion of response patients in the high ARG_score groups were significantly higher than that in the low ARG_score groups ($P < 0.001$, Figure 9E). Furthermore,

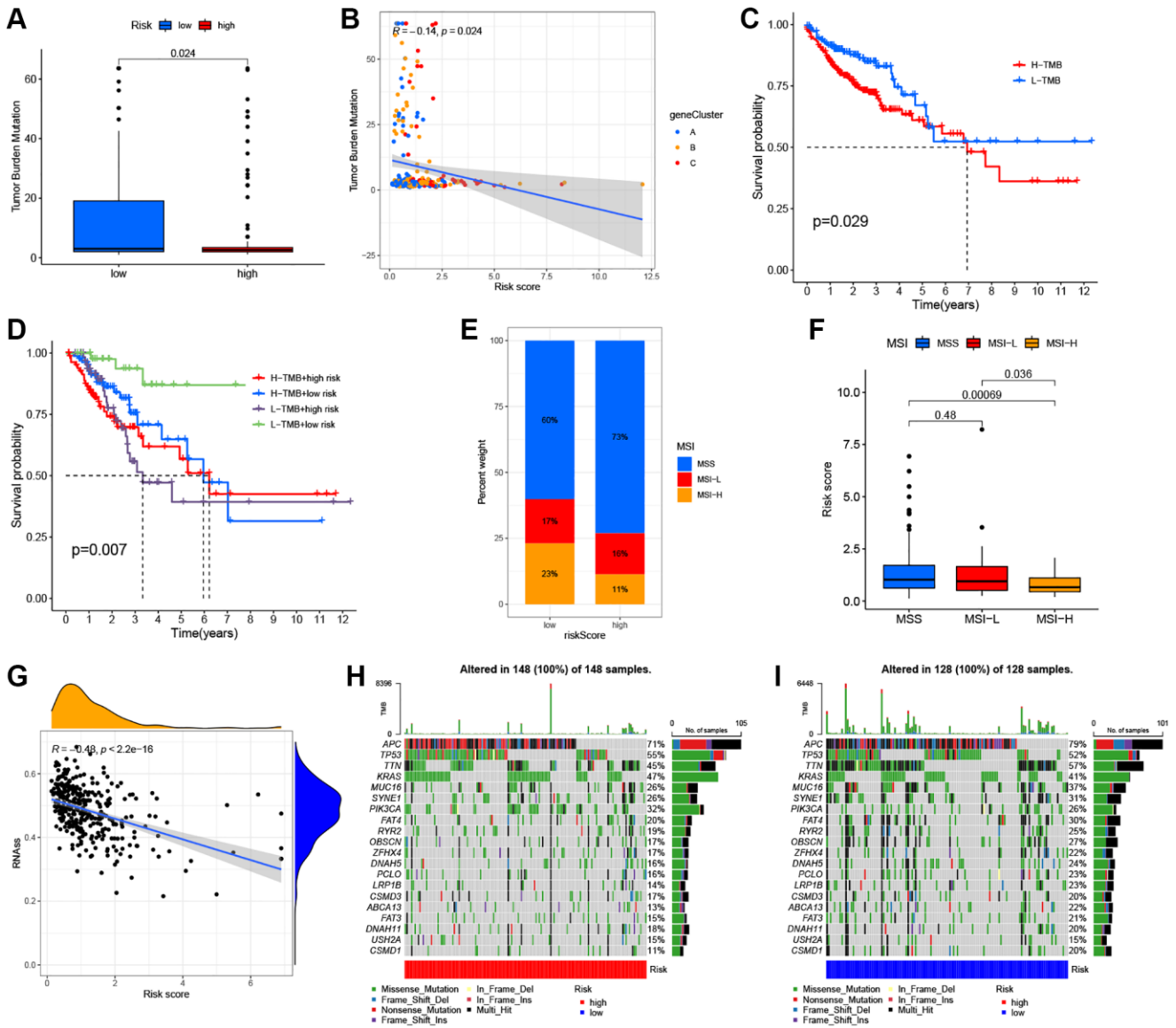


Figure 8. Comprehensive analysis of the ARG_score in CRC. (A, B) Relationships between the ARG_score and TMB. (C) Kaplan–Meier analysis of the RFS between the low- and high-TMB groups. (D) Survival analysis among four groups stratified by both TMB and ARG_score in CRC. (E, F) Relationships between ARG_score and MSI. (G) The correlation of the ARG_score with CSC index. (H, I) The waterfall plot of somatic mutation features established with the high- and low-ARG_score groups.

we observed significant negative associations between immune checkpoint levels (e.g., PD-1/L1, CTLA-4, and LAG3) and the ARG_score during the evaluation of

ARG-related TME characteristics. Thus, we used two methods to assess the predictive ability of the ARG_score in the immunotherapy response. The

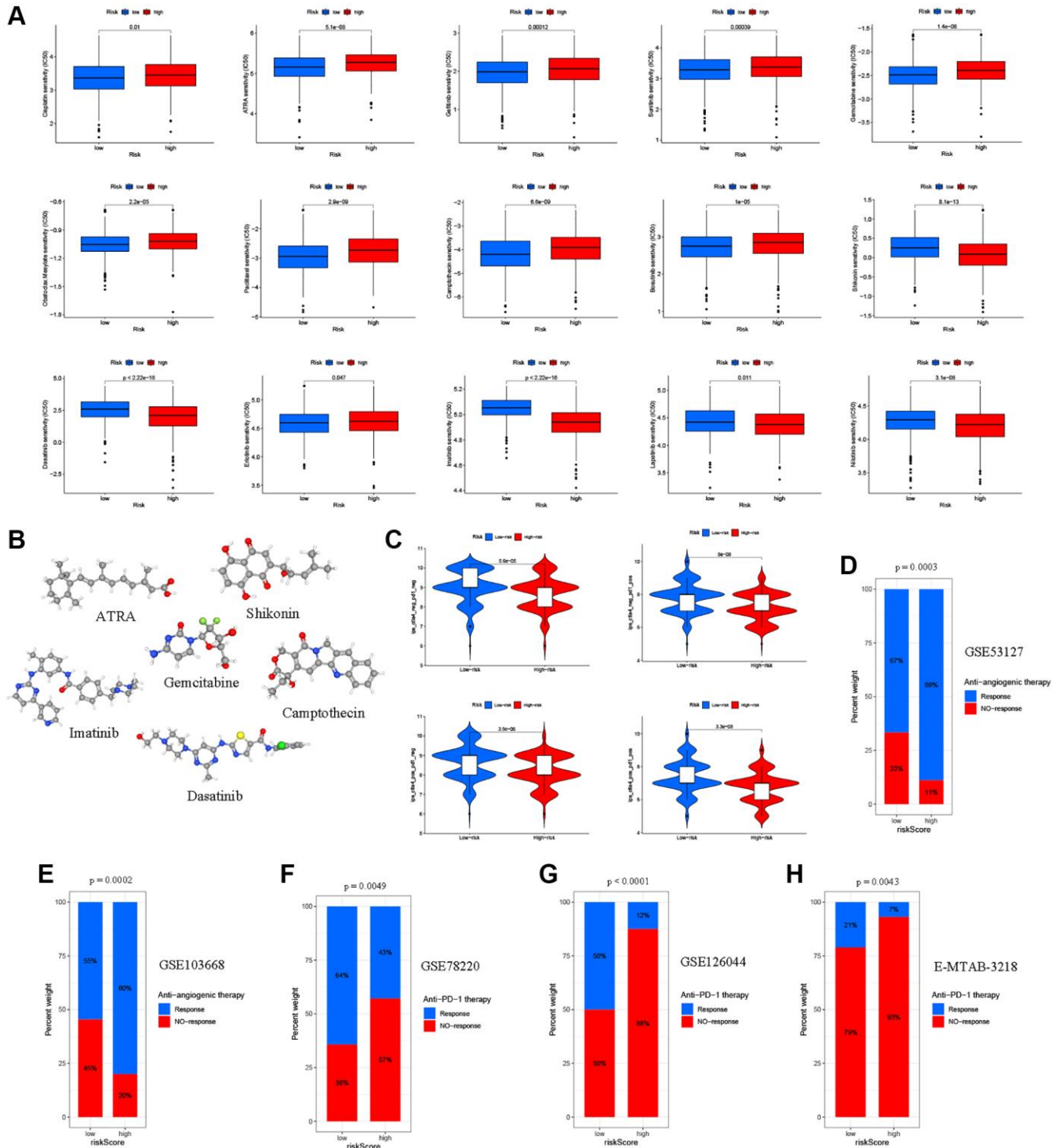


Figure 9. Relationships between the ARG_score and therapeutic sensitivity. (A) Relationships between the ARG_score and chemotherapy and sensitivity to targeted inhibitor therapy. (B) The 3D structure of 6 potential target drugs screened out from the cMap database. (C) The IPS in the different ARG_score groups. (D, E) Proportion of patients with different treatment outcomes (bevacizumab therapy) in the low- and high-ARG_score groups. The proportion of response patients in the low ARG_score groups were significantly higher than that in the high-ARG_score groups in both the GSE53127 and GSE103668 cohorts. (F–H) Patients with a high ARG_score exhibited poorer response outcomes after immunotherapy (anti-PD-1 therapy) using the GSE78220, GSE126044, and E-MTAB-3218 cohorts.

immunophenoscore (IPS) has been applied to predict the immunotherapeutic benefits to CTLA-4 and PD-1. Our results indicated that the low ARG_score group had a good response to PD-1 and CTLA-4 single positive, double positive, and double negative (Figure 9C). Further, three datasets (GSE78220, GSE126044, and E-MTAB-3218 cohorts) containing patient pretreatment and immunotherapeutic (anti-PD-1 therapy) data were collected, which had various clinical (PR/CR: responses or PD/CD: no-responses) and transcriptional (RNA-seq) information. As shown in Figure 9F–9H, the proportion of response patients in the low ARG_score groups were significantly higher than that in the high ARG_score groups ($P < 0.05$), which is consistent with the characteristics of significant vascular infiltration and immunosuppression in the high ARG_score groups. Together, our findings suggest that the ARG_score model had greater potential for predicting immunotherapy efficacy.

KLK10 was identified as a prognostic key ARG and immunotherapy target

According to the expression of 9 prognostic ARGs, a machine learning algorithm was used to predict candidate ARGs based on survival and death as binary dependent variables. As shown in Figure 10A, the three most important ARGs were CXCL10, SPINK1, and KLK10 in the RF model. Further, we analyzed the expression levels of these ARGs by combining the GTEx and TCGA databases. Our results showed that KLK10 was the most significantly different gene between normal and tumor tissues (Figure 10B, $FDR < 0.001$ and $\log_2 FC > 5$), which showed a high degree of consistency with the results of qRT-PCR analysis (Supplementary Figure 6N). To investigate the relationship between KLK10 expression and TME infiltration, we assessed the stromal score and immune score based on ESTIMATE analysis. As shown in Figure 10C, the stromal and immune score of high KLK10 expression groups were higher than those of low KLK10 expression groups, indicating that KLK10 expression was related to a hot TME. Additionally, we performed different algorithms to determine the relationship between KLK10 expression and infiltration levels of different immune cell types. The lollipop shape demonstrated that most immune cells had a positive correlation with KLK10 expression (Figure 10D). Furthermore, we used the TIDE algorithm to dissect the relationship between KLK10 expression and immunotherapy response. The results indicated that high KLK10 expression was positively correlated with the tumor immune escape (T cell dysfunction and T cell exclusion), which provides novel insights into precision immunotherapy in CRC (Figure 10E). Further, we explored the relationship between clinical and

prognostic value and KLK10 expression, and found that the expression of KLK10 was significantly related to the pathologic stage (Figure 10F). Moreover, high KLK10 expression was markedly related to poor RFS based on the median of the KLK10 expression levels (Figure 10G), highlighting its potential function as the tumor promotor in the occurrence and development of CRC.

Exploration and validation of the biological functions of KLK10 in CRC

To determine the biological characteristics of KLK10, we performed GSVA enrichment analysis between the low- and high-expression groups. The results demonstrated that KLK10 was involved in cancer development and progression mechanisms, including the TP53, MAPK, and VEGF pathways, apoptosis pathway, chemokine pathway, and tumor metastasis pathway (Figure 10H), indicating that KLK10 may play an essential role in tumor growth and invasion. To validate our results further, specific siRNAs were transfected into HT29 and HCT116 cells, and the transfection efficiency of KLK10 siRNA was verified by qRT-PCR (Figure 11A). In terms of cell proliferation, the results of the CCK8 assay indicated that a reduction in KLK10 significantly suppressed the proliferation ability of HT29 and HCT116 cells (Figure 11B). Further, the results of wound-healing assays showed that the knockdown of KLK10 markedly restrained the healing of scratched wounds (Figure 11C). As we know, the epithelial-mesenchymal transition (EMT) is one of the key mechanisms of tumor metastasis and one of the main factors leading to the poor prognosis in patients. We conducted a correlation analysis and presented our findings in a heatmap (Figure 11D). The results indicated that the KLK10 expression was positively associated with the mesenchymal markers (e.g., N-cadherin and vimentin), while being negatively related to the epithelial markers (E-cadherin). In line with these outcomes, western blot analysis confirmed that the knockdown of KLK10 downregulated the protein level of N-cadherin and vimentin in two cell lines. On the contrary, reduced KLK10 expression markedly upregulated the protein level of E-cadherin (Figure 11E, 11F). Collectively, these results highlight the pro-tumor effects of KLK10 in CRC, in accordance with the results of bioinformatics analysis.

KLK10 was significantly associated with *Fusobacterium nucleatum* (*F.n*) infection in CRC

The gut microbiota plays an important role in human health and disease. Recent studies have demonstrated that the presence of microbiota in various tumor tissues

is not accidental, but a common phenomenon [61]. Therefore, the intratumoral microbiota is also considered to be part of the TME. *F.n* is a gram-negative

anaerobe parasitic in the oral cavity, which has been found to promote CRC development through inhibits host anti-tumor immunity [62, 63]. Given the essential

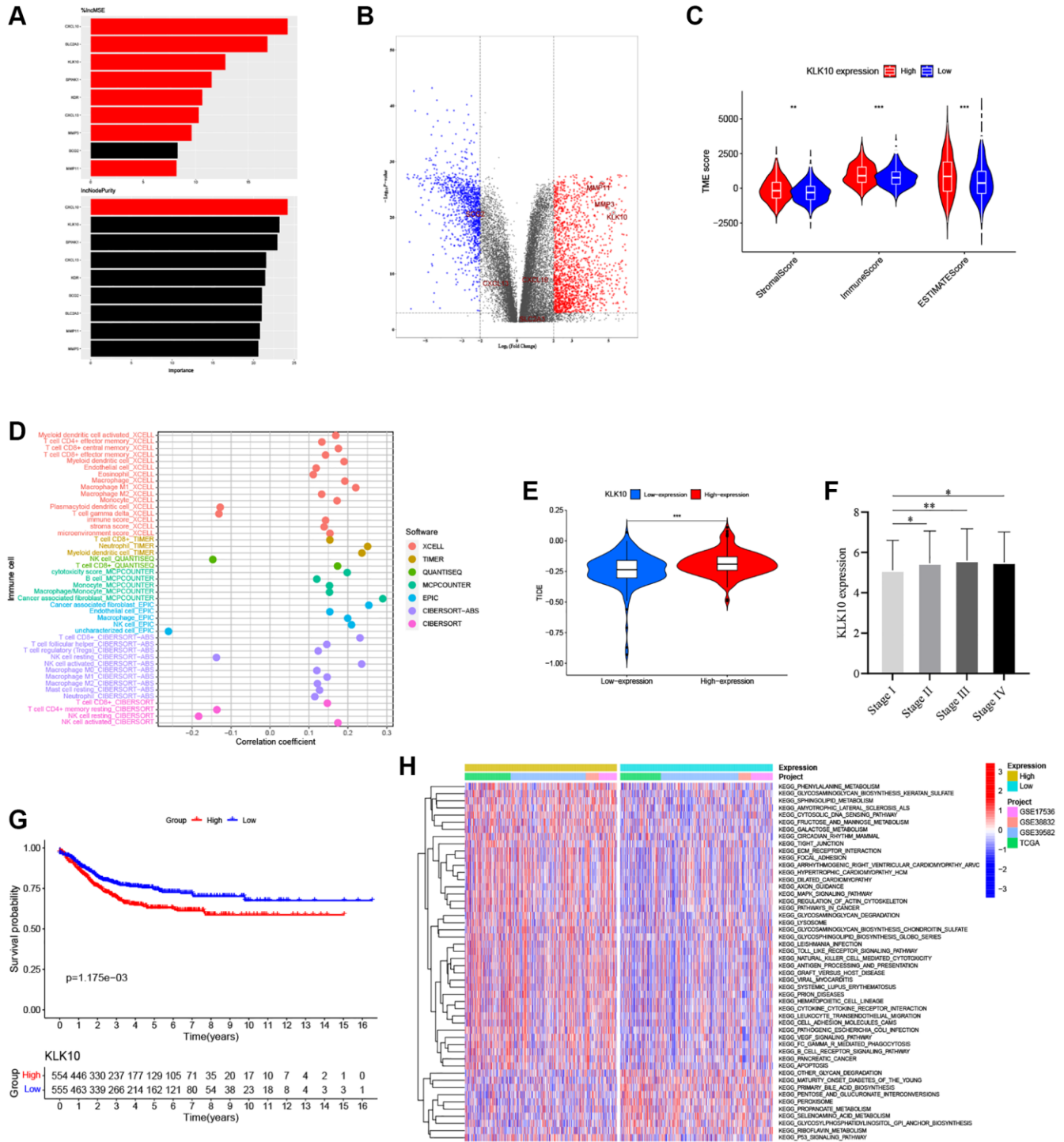


Figure 10. Identification of the key biomarkers in ARGs based on multi-methods. (A) The variable importance of 9 prognostic ARGs in RF models. (B) Volcano plot of significant DEGs between the normal and tumor tissues in four CRC cohorts (red represents upregulated genes, blue indicates downregulated genes, and gray represents no change). (C) Correlations between KLK10 expression and both the immune and stromal scores. (D) Spearman correlation analysis performed using different algorithms between KLK10 expression and immune cell infiltration. (E) The relationship between KLK10 expression and the TIDE score. (F) The correlation of KLK10 expression with clinicopathological staging characteristics. (G) Survival analysis for CRC patients with distinct KLK10 expression. (H) The GSEA pathway enrichment analysis between distinct KLK10 expression groups. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

roles played by ARGs in the tumor immune microenvironment, we speculated that *F.n* dysregulation might affect the expression of ARGs and hence

modulate local immune responses, in turn affecting immunotherapy. To determine the key ARGs associated with *F.n*, we performed whole genome microarray

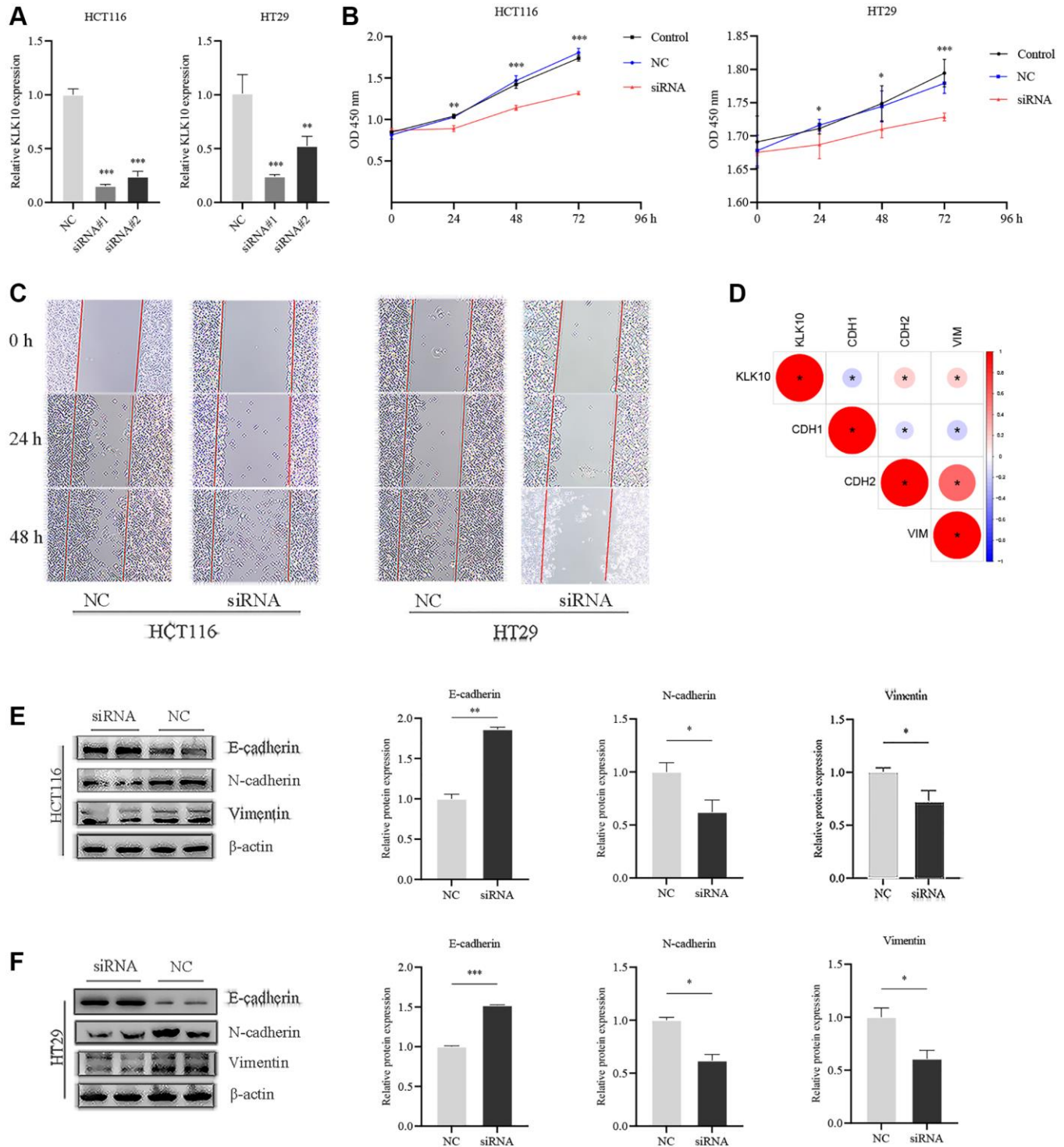


Figure 11. Influences of KLK10 expression on CRC cell proliferation, migration, and invasion. (A) The efficacy of the KLK10 transcript was detected after KLK10 knockdown in two CRC cell lines. (B) Cell viability was assessed by CCK8 array when KLK10 expression was reduced in cells. (C) Wound-healing assay of CRC cells with KLK10 knockdown monitored for 48-h with 24-h intervals. (D) Correlations between KLK10 expression and EMT markers (E-cadherin, N-cadherin, and Vimentin) using Spearman analysis. (E, F) The protein levels of E-cadherin, N-cadherin, and Vimentin were measured by western blotting after KLK10 knockdown in two CRC cell lines. Data are presented as mean \pm SD. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

analysis of HCT116 cell lines infected with *F.n* for 48 h. In addition, the gene expression profile of the RNA-seq dataset (GSE90944) from HT29 cell lines infected with *F.n* was also obtained and further analyzed. Surprisingly, KLK10 was identified as the key downstream ARG of *F.n* (DEGs was set as an FDR < 0.05 and $|\log_2 \text{FC}| \geq 1$) (Figure 12A). Moreover, the results showed that the expression of KLK10 was significantly upregulated by *F.n* in two cell lines (Figure 12B). Further, qPCR was performed to validate

the expression of KLK10 in CRC cell lines infected with *F.n*. As shown in Figure 12C, 12D, the mRNA expression of KLK10 was significantly upregulated in CRC cells by *F.n* infection for 24 h and 48 h, respectively. These findings were consistent with the results of western blot analysis (Figure 12E). Overall, our findings suggest that *F.n* infection strongly induces KLK10 expression in CRC cell lines. As such, KLK10 could serve as a promising clinical indicator and therapeutic target for CRC patients infected with *F.n*.

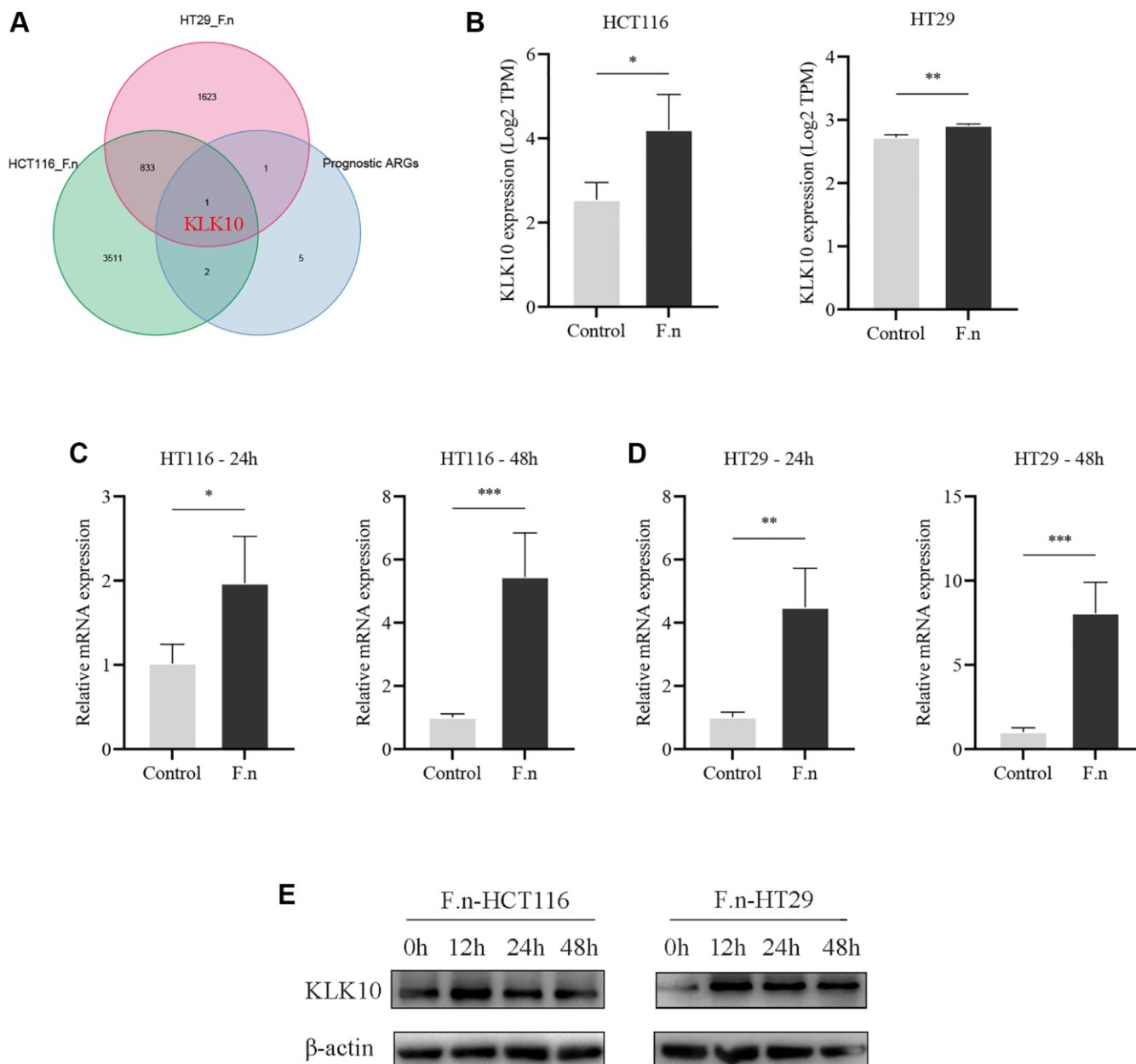


Figure 12. CRC cells co-cultured with *F.n* were found to have significantly upregulated KLK10 expression. (A) The Venn diagram demonstrating that KLK10 was shared by *F.n* infection-related DEGs and prognostic ARGs. (B) Microarray expression between HCT116 cells and *F.n* infected HCT116 cells (left panel) and expression of KLK10 in HT29 cells incubated with or without *F.n* (right panel). (C, D) The KLK10 mRNA expression in CRC cell lines infected with or without *F.n* for 24 h and 48 h, respectively, was determined by qPCR. (E) Representative western blot for KLK10 protein and β -actin protein extracted from CRC cells infected with *F.n* for 0, 12, 24, and 48 h. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

DISCUSSION

Evidence highlighting the critical role played by angiogenic factors in immune regulation and CRC progression is increasing [64–66]. Angiogenic factors excreted by CRC cells are known to stimulate endothelial cells to regulate angiogenic switches during CRC development [67]. In addition, angiogenic factors contribute to a pattern of impaired immune activation (immunosuppression) by activating suppressive immune cells or inhibiting immune effector cells, which can in turn stimulate angiogenesis and tumor progression [68, 69]. Notably, many researchers have shown that the indispensable relationship between angiogenesis and innate immunity and angiogenesis targeting has improved the immunotherapy efficacy and prognostic outcomes of tumor patients [70–73]. However, at present, there is a lack of bioinformatics analyses that demonstrate the holistic impact and TME cells infiltration features regulated by the combined role of multiple ARGs in CRC.

Herein, we demonstrated global alterations in ARGs at the transcriptional and genetic level. Most of them were elevated in advanced CRC patients (stage III–IV) and related to RFS. Next, we identified two distinct angiogenesis subgroups (Cluster A and B) based on 36 ARGs, and Cluster B was found to have more advanced clinicopathological characteristics and worse RFS compared to Cluster A. Furthermore, significant discrepancies were observed in TME infiltrations and biological functions between the two subgroups. Given the crucial roles played by gene mutations in the efficacy of immunotherapy, we further identified three gene Clusters with different clinicopathological characteristics, immune cell abundance, and functional characterization based on the DEGs between the two angiogenesis subgroups. To quantify the angiogenesis subgroups, a prognostic ARG_score with robust predictive ability was constructed. The Cluster B and gene Cluster C with the poorest RFS had the greatest ARG_score among the two ARG subgroups and three gene subgroups. Patients with low- and high-ARG_score exhibited significantly distinct clinical outcomes, indicating that the ARG_score could predict poor clinical outcomes. Angiogenesis is confirmed that play a part in the aggressiveness of CRC [74, 75]. Interestingly, differences in mRNA transcriptomes between different angiogenesis subgroups were markedly enriched in cancer- and metastasis-related pathways, consistent with the existing conclusions.

The clinicopathological characteristics differed significantly between the low- and high-ARG_score groups. Multivariate Cox regression analysis showed that ARG_score could be used as an independent

predictor of RFS in CRC patients. Its predictive robustness for 1-, 5-, and 10-year RFS were validated by ROCs. Moreover, we confirmed that ARG_score could be used for the prognosis stratification of CRC patients, implying that the ARG_score may have a reliable predictive capacity. Further, a quantitative nomogram was established to improve the performance and facilitate the use of the model by integrating the ARG_score and tumor stage. Recently, chemoresistance has become a major challenge, and the prognosis for patients with CRC can be very poor due to disease recurrence [76]. To address this, we investigated the sensitivity of patients in different ARG_score groups to different drugs. Our findings indicated that combination with targeting angiogenesis was conducive to ameliorating drug resistance and improving prognosis. In this context, TMB and MSI are considered as response predictors for tumor immunotherapy, including CRC [77, 78]. Our results revealed that there were significant differences between TMB and MSI between the low- and high-ARG_score groups. Furthermore, the high TMB and MSI-H groups have been previously demonstrated to be associated with a better prognosis for CRC patients [14, 79], consistent with our expected results. Results from clinical trials have demonstrated that anti-PD-1 antibodies can ultimately achieve a lasting complete response in CRC patients [80, 81]. In the present study, the expression of immune checkpoints was found to be significantly different in the high- and low-ARG_score groups. Thus, our results suggested that the ARG_score can be used to predict the efficacy of targeting angiogenesis therapy and immunotherapy, and in conjunction with targeted therapy, may be a significant adjunctive strategy for CRC immunotherapy. Taken together, these findings indicate that the ARG_score can be used as a predictor of response to immunotherapy and clinical outcomes independent of the TMB, and may also be useful in assessing the MSI status of CRC patients effectively.

Evidence has increasingly shown that the TME plays a vital role in carcinogenesis, tumor progression, and drug resistance [82]. Currently, CRC patients continue to exhibit immunotherapy heterogeneity in their clinical outcomes, highlighting the important effects of the TME on CRC progression. The TME surrounding tumor cells is mainly composed of stromal cells and immune cells, such as tumor infiltrating immune cells, blood vessels, and tumor-associated fibroblasts [83]. Further, we evaluated the immune and stromal scores using the ESTIMATE algorithm and found that the high ARG_score groups significantly presented higher stromal scores than the low ARG_score groups, while the immune scores were poorer than those of the low ARG_score groups. These findings suggest that angiogenesis could be related to the involvement of the

TME, thus regulating CRC tumorigenesis and progression. Granulocytes, lymphocytes, and macrophages are primary tumor-infiltrating immune cells elements of the TME, and have been proven to take part in many host immune pathways, including inflammatory responses mediated by the tumor to improve survival [84]. Tumor-associated macrophages are mainly categorized into M1 macrophages and M2 macrophages. M1 macrophages are characterized by anti-tumor function due to their ability to produce type I pro-inflammatory cytokines, while M2 macrophages contribute to the formation of immunosuppressive microenvironment and matrix remodeling, hence accelerating tumor growth [85, 86]. We noticed that the low ARG_score groups demonstrated favorable survival with more infiltrations of M1 macrophages, while increased infiltrations of M2 macrophages were observed in the high ARG_score groups with a worse prognosis. These findings are consistent with several previous studies. In addition, effector T cells, memory T cells, and T cell differentiation have been shown to play a crucial role in the immune defense of CRC, and higher densities of tumor-infiltrating T cells in CRC tissues indicate a good prognosis [87, 88]. We demonstrated that the low ARG_score groups showed a higher infiltration of activated memory CD4+, CD8 + T, and follicular helper T cells, indicating that they have a positive effect on CRC prognosis. Recent studies have revealed that the infiltration of Tregs can suppress anti-cancer immunoreactivity, and hence favor tumor growth [89]. This is consistent with our finding of a higher Treg infiltration in the TME of patients with a high ARG_score.

Kallikrein-related peptidase 10 (KLK10) is a member of the KLKs family with tryptic or chymotryptic activity [90, 91]. Several studies have found that CRC patients with high KLK10 expression have a poorer prognosis [92, 93]. In this study, we confirmed that KLK10 expression was significantly upregulated at the mRNA level in CRC tissues compared to adjacent normal tissues. Prognosis analysis showed that higher levels of KLK10 expression in patients was significantly associated with a shorter RFS, which was consistent with the results of previously published studies. However, the roles of KLK10 in CRC development, immune infiltration, and immunotherapy response remain poorly understood. Functional enrichment analysis showed that higher KLK10 expression was involved in tumor cells adhesion, invasion, metastasis, and immune cell infiltration, which highlighted the potential mining value of KLK10 in CRC progression and the tumor immune microenvironment. Further *in vitro* experiments were performed to confirm that the knockdown of KLK10 in CRC cells markedly decreased in cell viability and proliferation. In addition,

this study was the first to suggest that the downregulation of KLK10 suppressed the migration and invasion of CRC cells through the inhibition of the EMT. Recently, KLK10 was identified as a potential target for immunotherapy based on the immunopeptidome analysis of ovarian cancer antigens [94]. The results of different algorithms in immune cell infiltration showed that high KLK10 expression may be conducive to an immune hot type. Moreover, the TIDE analysis implied that patients with high KLK10 expression were more susceptible to tumor immune escape, in accordance with previous studies. *F.n* has been confirmed to play an important role in the occurrence and development of CRC. Interestingly, we explored the sequencing results obtained after CRC cells were infected with *F.n*, and discovered that KLK10 expression was significantly upregulated. *F.n* may affect CRC cell proliferation, invasion, metastasis, and drug resistance via the upregulation of KLK10, and hence change the clinical outcomes of patients. At present, there are no studies on the impact of *F.n* on the survival and development of CRC regulated by the expression of ARGs. However, further systematic molecular experiments may elucidate novel pathogenic pathways and therapeutic targets.

This study has some limitations. Namely, our results were obtained by analyzing public databases (TCGA and GEO), and future studies will need to incorporate the latest clinical samples for further verification. Furthermore, multicenter clinical immunotherapy cohorts should be performed to confirm the robustness and consistency of the ARG_score in predicting prognosis and immunotherapy efficacy, which will be the emphasis of future research. Lastly, complementary *in vivo* and *in vitro* studies will be needed to verify our findings. Our laboratory is further investigating the important links between these factors.

CONCLUSIONS

In this study, a novel angiogenesis-related molecular subtype based on ARGs in CRC was identified, providing a comprehensive assessment of the heterogeneity and complexity of TME. We also identified the therapeutic orientation of ARGs in targeting therapy and immunotherapy. Although these findings highlight the vital clinical implications of ARGs, future studies in multi-centers on a larger scale are essential. This study is the first to describe the potential contribution of *F.n* to the transcriptional changes in ARGs, and provides specific directions for further research. The results presented here suggest that ARGs could be used as an effective biomarker to predict prognosis and immunotherapy response in patients with CRC, highlighting their potential involvement in tumor

development in *F.n*-infected CRC patients. Further *in vitro* or *in vivo* experiments should be conducted to verify the relationships between these factors.

AUTHOR CONTRIBUTIONS

ZL designed the research route and conducted the experimental verification. ZL executed the majority of data collation, analysis and writing. YL contributed to some data analysis and figures. PG and YW performed supervision and funding acquisition. All authors have read and approved the final manuscript.

ACKNOWLEDGMENTS

We would like to extend our thanks to the GEO and TCGA network for providing the data. The authors also gratefully acknowledge contributions from the GEPIA 2 database, PubChem database, TCIA database, and TIDE website.

CONFLICTS 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.

ETHICAL STATEMENT AND CONSENT

All procedures for the colorectal cancer validation cohort were done in agreement with the Declaration of Helsinki and approved by the Ethical Committee of Ningbo Second Hospital, also known as Ningbo HwaMei Hospital of the University of Chinese Academy of Science (application number SL-NBEY-KYSB-2022-009-01). The patients/participants provided their written informed consent to participate in this study.

FUNDING

This research was funded by the National Nature Science Foundation of China (NSFC) (Grant No. 81970466), Ningbo Top Medical and Health Research Program (Grant No. 2022010101) and Ningbo Digestive System Tumor Clinical Medical Research Center (Grant No. 2019A21003).

REFERENCES

1. Cartwright TH. Treatment decisions after diagnosis of metastatic colorectal cancer. *Clin Colorectal Cancer*. 2012; 11:155–66. <https://doi.org/10.1016/j.clcc.2011.11.001> PMID:22192364
2. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA Cancer J Clin*. 2013; 63:11–30. <https://doi.org/10.3322/caac.21166> PMID:23335087
3. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021; 71:209–49. <https://doi.org/10.3322/caac.21660> PMID:33538338
4. Li N, Lu B, Luo C, Cai J, Lu M, Zhang Y, Chen H, Dai M. Incidence, mortality, survival, risk factor and screening of colorectal cancer: A comparison among China, Europe, and northern America. *Cancer Lett*. 2021; 522:255–68. <https://doi.org/10.1016/j.canlet.2021.09.034> PMID:34563640
5. Ganesh K, Stadler ZK, Cercek A, Mendelsohn RB, Shia J, Segal NH, Diaz LA Jr. Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nat Rev Gastroenterol Hepatol*. 2019; 16:361–75. <https://doi.org/10.1038/s41575-019-0126-x> PMID:30886395
6. Zhang L, Cao F, Zhang G, Shi L, Chen S, Zhang Z, Zhi W, Ma T. Trends in and Predictions of Colorectal Cancer Incidence and Mortality in China From 1990 to 2025. *Front Oncol*. 2019; 9:98. <https://doi.org/10.3389/fonc.2019.00098> PMID:30847304
7. Moriarity A, O'Sullivan J, Kennedy J, Mehigan B, McCormick P. Current targeted therapies in the treatment of advanced colorectal cancer: a review. *Ther Adv Med Oncol*. 2016; 8:276–93. <https://doi.org/10.1177/1758834016646734> PMID:27482287
8. Xie YH, Chen YX, Fang JY. Comprehensive review of targeted therapy for colorectal cancer. *Signal Transduct Target Ther*. 2020; 5:22. <https://doi.org/10.1038/s41392-020-0116-z> PMID:32296018
9. Lugano R, Ramachandran M, Dimberg A. Tumor angiogenesis: causes, consequences, challenges and opportunities. *Cell Mol Life Sci*. 2020; 77:1745–70. <https://doi.org/10.1007/s00018-019-03351-7> PMID:31690961
10. Azoitei N, Pusapati GV, Kleger A, Möller P, Küfer R, Genze F, Wagner M, van Lint J, Carmeliet P, Adler G, Seufferlein T. Protein kinase D2 is a crucial regulator of tumour cell-endothelial cell communication in gastrointestinal tumours. *Gut*. 2010; 59:1316–30.

- <https://doi.org/10.1136/gut.2009.206813>
PMID:20732914
11. Hurwitz H, Fehrenbacher L, Novotny W, Cartwright T, Hainsworth J, Heim W, Berlin J, Baron A, Griffing S, Holmgren E, Ferrara N, Fyfe G, Rogers B, et al. Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *N Engl J Med*. 2004; 350:2335–42.
<https://doi.org/10.1056/NEJMoa032691>
PMID:15175435
 12. Mitsuhashi A, Goto H, Saijo A, Trung VT, Aono Y, Ogino H, Kuramoto T, Tabata S, Uehara H, Izumi K, Yoshida M, Kobayashi H, Takahashi H, et al. Fibrocyte-like cells mediate acquired resistance to anti-angiogenic therapy with bevacizumab. *Nat Commun*. 2015; 6:8792.
<https://doi.org/10.1038/ncomms9792>
PMID:26635184
 13. Smyth EC, Nilsson M, Grabsch HI, van Grieken NC, Lordick F. Gastric cancer. *Lancet*. 2020; 396:635–48.
[https://doi.org/10.1016/S0140-6736\(20\)31288-5](https://doi.org/10.1016/S0140-6736(20)31288-5)
PMID:32861308
 14. Zhao W, Jin L, Chen P, Li D, Gao W, Dong G. Colorectal cancer immunotherapy-Recent progress and future directions. *Cancer Lett*. 2022; 545:215816.
<https://doi.org/10.1016/j.canlet.2022.215816>
PMID:35810989
 15. Jahanafrooz Z, Mosafer J, Akbari M, Hashemzaei M, Mokhtarzadeh A, Baradaran B. Colon cancer therapy by focusing on colon cancer stem cells and their tumor microenvironment. *J Cell Physiol*. 2020; 235:4153–66.
<https://doi.org/10.1002/jcp.29337>
PMID:31647128
 16. Saleh R, Taha RZ, Toor SM, Sasidharan Nair V, Murshed K, Khawar M, Al-Dhaheri M, Petkar MA, Abu Nada M, Elkord E. Expression of immune checkpoints and T cell exhaustion markers in early and advanced stages of colorectal cancer. *Cancer Immunol Immunother*. 2020; 69:1989–99.
<https://doi.org/10.1007/s00262-020-02593-w>
PMID:32393998
 17. Bader JE, Voss K, Rathmell JC. Targeting Metabolism to Improve the Tumor Microenvironment for Cancer Immunotherapy. *Mol Cell*. 2020; 78:1019–33.
<https://doi.org/10.1016/j.molcel.2020.05.034>
PMID:32559423
 18. Kaymak I, Williams KS, Cantor JR, Jones RG. Immuno-metabolic Interplay in the Tumor Microenvironment. *Cancer Cell*. 2021; 39:28–37.
<https://doi.org/10.1016/j.ccell.2020.09.004>
PMID:33125860
 19. Jin MZ, Jin WL. The updated landscape of tumor microenvironment and drug repurposing. *Signal Transduct Target Ther*. 2020; 5:166.
<https://doi.org/10.1038/s41392-020-00280-x>
PMID:32843638
 20. Du WW, Yang W, Yee AJ. Roles of versican in cancer biology--tumorigenesis, progression and metastasis. *Histol Histopathol*. 2013; 28:701–13.
<https://doi.org/10.14670/HH-28.701>
PMID:23519970
 21. Rahmani M, Wong BW, Ang L, Cheung CC, Carthy JM, Walinski H, McManus BM. Versican: signaling to transcriptional control pathways. *Can J Physiol Pharmacol*. 2006; 84:77–92.
<https://doi.org/10.1139/y05-154>
PMID:16845893
 22. Sluiter NR, de Cuba EM, Kwakman R, Meijerink WJ, Delis-van Diemen PM, Coupé VM, Beliën JA, Meijer GA, de Hingh IH, te Velde EA. Versican and vascular endothelial growth factor expression levels in peritoneal metastases from colorectal cancer are associated with survival after cytoreductive surgery and hyperthermic intraperitoneal chemotherapy. *Clin Exp Metastasis*. 2016; 33:297–307.
<https://doi.org/10.1007/s10585-016-9779-9>
PMID:26873137
 23. Takeshita S, Kikuno R, Tezuka K, Amann E. Osteoblast-specific factor 2: cloning of a putative bone adhesion protein with homology with the insect protein fasciclin I. *Biochem J*. 1993; 294:271–8.
<https://doi.org/10.1042/bj2940271>
PMID:8363580
 24. Ueki A, Komura M, Koshino A, Wang C, Nagao K, Homochi M, Tsukada Y, Ebi M, Ogasawara N, Tsuzuki T, Kasai K, Kasugai K, Takahashi S, Inaguma S. Stromal POSTN Enhances Motility of Both Cancer and Stromal Cells and Predicts Poor Survival in Colorectal Cancer. *Cancers (Basel)*. 2023; 15:606.
<https://doi.org/10.3390/cancers15030606>
PMID:36765564
 25. Baniias L, Gurzu S, Kovacs Z, Bara T, Bara T Jr, Jung I. Nuclear maspin expression: A biomarker for budding assessment in colorectal cancer specimens. *Pathol Res Pract*. 2017; 213:1227–30.
<https://doi.org/10.1016/j.prp.2017.07.025>
PMID:28780084
 26. Li Z, Shi HY, Zhang M. Targeted expression of maspin in tumor vasculatures induces endothelial cell apoptosis. *Oncogene*. 2005; 24:2008–19.
<https://doi.org/10.1038/sj.onc.1208449>
PMID:15688005
 27. Bettstetter M, Woenckhaus M, Wild PJ, Rümmele P, Blaszyk H, Hartmann A, Hofstädter F, Dietmaier W.

- Elevated nuclear maspin expression is associated with microsatellite instability and high tumour grade in colorectal cancer. *J Pathol.* 2005; 205:606–14.
<https://doi.org/10.1002/path.1732>
 PMID:15714592
28. Dietmaier W, Bettstetter M, Wild PJ, Woenckhaus M, Rümmele P, Hartmann A, Dechant S, Blaszyk H, Pauer A, Klinkhammer-Schalke M, Hofstädter F. Nuclear Maspin expression is associated with response to adjuvant 5-fluorouracil based chemotherapy in patients with stage III colon cancer. *Int J Cancer.* 2006; 118:2247–54.
<https://doi.org/10.1002/ijc.21620>
 PMID:16331619
 29. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczesniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A. A survey of best practices for RNA-seq data analysis. *Genome Biol.* 2016; 17:13.
<https://doi.org/10.1186/s13059-016-0881-8>
 PMID:26813401
 30. Sabah A, Tiun S, Sani NS, Ayob M, Taha AY. Enhancing web search result clustering model based on multiview multirepresentation consensus cluster ensemble (mmcc) approach. *PLoS One.* 2021; 16:e0245264.
<https://doi.org/10.1371/journal.pone.0245264>
 PMID:33449949
 31. Seiler M, Huang CC, Szalma S, Bhanot G. ConsensusCluster: a software tool for unsupervised cluster discovery in numerical data. *OMICS.* 2010; 14:109–13.
<https://doi.org/10.1089/omi.2009.0083>
 PMID:20141333
 32. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 2013; 14:7.
<https://doi.org/10.1186/1471-2105-14-7>
 PMID:23323831
 33. Rich JT, Neely JG, Paniello RC, Voelker CC, Nussenbaum B, Wang EW. A practical guide to understanding Kaplan-Meier curves. *Otolaryngol Head Neck Surg.* 2010; 143:331–6.
<https://doi.org/10.1016/j.otohns.2010.05.007>
 PMID:20723767
 34. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* 2015; 12:453–7.
<https://doi.org/10.1038/nmeth.3337>
 PMID:25822800
 35. Huang L, Wu C, Xu D, Cui Y, Tang J. Screening of Important Factors in the Early Sepsis Stage Based on the Evaluation of ssGSEA Algorithm and ceRNA Regulatory Network. *Evol Bioinform Online.* 2021; 17:11769343211058463.
<https://doi.org/10.1177/11769343211058463>
 PMID:34866895
 36. Mahoney KM, Rennert PD, Freeman GJ. Combination cancer immunotherapy and new immunomodulatory targets. *Nat Rev Drug Discov.* 2015; 14:561–84.
<https://doi.org/10.1038/nrd4591>
 PMID:26228759
 37. Wang JB, Li P, Liu XL, Zheng QL, Ma YB, Zhao YJ, Xie JW, Lin JX, Lu J, Chen QY, Cao LL, Lin M, Liu LC, et al. An immune checkpoint score system for prognostic evaluation and adjuvant chemotherapy selection in gastric cancer. *Nat Commun.* 2020; 11:6352.
<https://doi.org/10.1038/s41467-020-20260-7>
 PMID:33311518
 38. Zhang B, Wu Q, Li B, Wang D, Wang L, Zhou YL. m⁶A regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in gastric cancer. *Mol Cancer.* 2020; 19:53.
<https://doi.org/10.1186/s12943-020-01170-0>
 PMID:32164750
 39. Zhao H, Chen Y, Shen P, Gong L. Identification of Immune Cell Infiltration Landscape and Their Prognostic Significance in Uveal Melanoma. *Front Cell Dev Biol.* 2021; 9:713569.
<https://doi.org/10.3389/fcell.2021.713569>
 PMID:34513843
 40. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012; 16:284–7.
<https://doi.org/10.1089/omi.2011.0118>
 PMID:22455463
 41. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol.* 2008; 26:1364–70.
<https://doi.org/10.1200/JCO.2007.12.9791>
 PMID:18323559
 42. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 2018; 28:1747–56.
<https://doi.org/10.1101/gr.239244.118>
 PMID:30341162
 43. Zhang H, Meltzer P, Davis S. RCircos: an R package for Circos 2D track plots. *BMC Bioinformatics.* 2013; 14:244.
<https://doi.org/10.1186/1471-2105-14-244>
 PMID:23937229

44. Geleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One*. 2014; 9:e107468. <https://doi.org/10.1371/journal.pone.0107468> PMID:25229481
45. Pentheroudakis G, Kotoula V, Fountzilas E, Kouvatseas G, Basdanis G, Xanthakis I, Makatsoris T, Charalambous E, Papamichael D, Samantas E, Papakostas P, Bafaloukos D, Razis E, et al. A study of gene expression markers for predictive significance for bevacizumab benefit in patients with metastatic colon cancer: a translational research study of the Hellenic Cooperative Oncology Group (HeCOG). *BMC Cancer*. 2014; 14:111. <https://doi.org/10.1186/1471-2407-14-111> PMID:24555920
46. Birkbak NJ, Li Y, Pathania S, Greene-Colozzi A, Dreze M, Bowman-Colin C, Sztupinski Z, Krzystanek M, Diossy M, Tung N, Ryan PD, Garber JE, Silver DP, et al. Overexpression of BLM promotes DNA damage and increased sensitivity to platinum salts in triple-negative breast and serous ovarian cancers. *Ann Oncol*. 2018; 29:903–9. <https://doi.org/10.1093/annonc/mdy049> PMID:29452344
47. Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hui-Lieskovan S, Berent-Maoz B, Pang J, Chmielowski B, Cherry G, Seja E, Lomeli S, Kong X, et al. Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell*. 2016; 165:35–44. <https://doi.org/10.1016/j.cell.2016.02.065> PMID:26997480
48. Cho JW, Hong MH, Ha SJ, Kim YJ, Cho BC, Lee I, Kim HR. Genome-wide identification of differentially methylated promoters and enhancers associated with response to anti-PD-1 therapy in non-small cell lung cancer. *Exp Mol Med*. 2020; 52:1550–63. <https://doi.org/10.1038/s12276-020-00493-8> PMID:32879421
49. Choueiri TK, Fishman MN, Escudier B, McDermott DF, Drake CG, Kluger H, Stadler WM, Perez-Gracia JL, McNeel DG, Curti B, Harrison MR, Plimack ER, Appleman L, et al. Immunomodulatory Activity of Nivolumab in Metastatic Renal Cell Carcinoma. *Clin Cancer Res*. 2016; 22:5461–71. <https://doi.org/10.1158/1078-0432.CCR-15-2839> PMID:27169994
50. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26:139–40. <https://doi.org/10.1093/bioinformatics/btp616> PMID:19910308
51. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015; 43:e47. <https://doi.org/10.1093/nar/gkv007> PMID:25605792
52. Yu T, Guo F, Yu Y, Sun T, Ma D, Han J, Qian Y, Kryczek I, Sun D, Nagarsheth N, Chen Y, Chen H, Hong J, et al. *Fusobacterium nucleatum* Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. *Cell*. 2017; 170:548–63.e16. <https://doi.org/10.1016/j.cell.2017.07.008> PMID:28753429
53. Zhang S, Yang Y, Weng W, Guo B, Cai G, Ma Y, Cai S. *Fusobacterium nucleatum* promotes chemoresistance to 5-fluorouracil by upregulation of BIRC3 expression in colorectal cancer. *J Exp Clin Cancer Res*. 2019; 38:14. <https://doi.org/10.1186/s13046-018-0985-y> PMID:30630498
54. Grada A, Otero-Vinas M, Prieto-Castrillo F, Obagi Z, Falanga V. Research Techniques Made Simple: Analysis of Collective Cell Migration Using the Wound Healing Assay. *J Invest Dermatol*. 2017; 137:e11–6. <https://doi.org/10.1016/j.jid.2016.11.020> PMID:28110712
55. Xu F, He L, Zhan X, Chen J, Xu H, Huang X, Li Y, Zheng X, Lin L, Chen Y. DNA methylation-based lung adenocarcinoma subtypes can predict prognosis, recurrence, and immunotherapeutic implications. *Aging (Albany NY)*. 2020; 12:25275–93. <https://doi.org/10.18632/aging.104129> PMID:33234739
56. Picard E, Verschoor CP, Ma GW, Pawelec G. Relationships Between Immune Landscapes, Genetic Subtypes and Responses to Immunotherapy in Colorectal Cancer. *Front Immunol*. 2020; 11:369. <https://doi.org/10.3389/fimmu.2020.00369> PMID:32210966
57. Topalian SL, Taube JM, Anders RA, Pardoll DM. Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy. *Nat Rev Cancer*. 2016; 16:275–87. <https://doi.org/10.1038/nrc.2016.36> PMID:27079802
58. Biller LH, Schrag D. Diagnosis and Treatment of Metastatic Colorectal Cancer: A Review. *JAMA*. 2021; 325:669–85. <https://doi.org/10.1001/jama.2021.0106> PMID:33591350

59. Rmali KA, Puntis MC, Jiang WG. Tumour-associated angiogenesis in human colorectal cancer. *Colorectal Dis.* 2007; 9:3–14.
<https://doi.org/10.1111/j.1463-1318.2006.01089.x>
PMID:17181841
60. Yeung Y, Tebbutt NC. Bevacizumab in colorectal cancer: current and future directions. *Expert Rev Anticancer Ther.* 2012; 12:1263–73.
<https://doi.org/10.1586/era.12.104>
PMID:23113577
61. Nejman D, Livyatan I, Fuks G, Gavert N, Zwang Y, Geller LT, Rotter-Maskowitz A, Weiser R, Mallel G, Gigi E, Meltser A, Douglas GM, Kamer I, et al. The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science.* 2020; 368:973–80.
<https://doi.org/10.1126/science.aay9189>
PMID:32467386
62. Gur C, Ibrahim Y, Isaacson B, Yamin R, Abed J, Gamliel M, Enk J, Bar-On Y, Stanietsky-Kaynan N, Copenhagen-Glazer S, Shussman N, Almogy G, Cuapio A, et al. Binding of the Fap2 protein of *Fusobacterium nucleatum* to human inhibitory receptor TIGIT protects tumors from immune cell attack. *Immunity.* 2015; 42:344–55.
<https://doi.org/10.1016/j.immuni.2015.01.010>
PMID:25680274
63. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, Clancy TE, Chung DC, Lochhead P, Hold GL, El-Omar EM, Brenner D, Fuchs CS, et al. *Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe.* 2013; 14:207–15.
<https://doi.org/10.1016/j.chom.2013.07.007>
PMID:23954159
64. Lin Y, Bai L, Chen W, Xu S. The NF-kappaB activation pathways, emerging molecular targets for cancer prevention and therapy. *Expert Opin Ther Targets.* 2010; 14:45–55.
<https://doi.org/10.1517/14728220903431069>
PMID:20001209
65. Mody K, Baldeo C, Bekaii-Saab T. Antiangiogenic Therapy in Colorectal Cancer. *Cancer J.* 2018; 24:165–70.
<https://doi.org/10.1097/PPO.0000000000000328>
PMID:30119079
66. Pedrosa L, Esposito F, Thomson TM, Maurel J. The Tumor Microenvironment in Colorectal Cancer Therapy. *Cancers (Basel).* 2019; 11:1172.
<https://doi.org/10.3390/cancers11081172>
PMID:31416205
67. Cubillo A, Álvarez-Gallego R, Muñoz M, Pond G, Perea S, Sánchez G, Martín M, Rodríguez-Pascual J, Garralda E, Vega E, de Vicente E, Quijano Y, Muñoz C, et al. Dynamic Angiogenic Switch as Predictor of Response to Chemotherapy-Bevacizumab in Patients With Metastatic Colorectal Cancer. *Am J Clin Oncol.* 2019; 42:56–9.
<https://doi.org/10.1097/COC.0000000000000474>
PMID:29975196
68. O'Byrne KJ, Dalgleish AG, Browning MJ, Steward WP, Harris AL. The relationship between angiogenesis and the immune response in carcinogenesis and the progression of malignant disease. *Eur J Cancer.* 2000; 36:151–69.
[https://doi.org/10.1016/s0959-8049\(99\)00241-5](https://doi.org/10.1016/s0959-8049(99)00241-5)
PMID:10741273
69. Rahma OE, Hodi FS. The Intersection between Tumor Angiogenesis and Immune Suppression. *Clin Cancer Res.* 2019; 25:5449–57.
<https://doi.org/10.1158/1078-0432.CCR-18-1543>
PMID:30944124
70. Finn RS, Qin S, Ikeda M, Galle PR, Ducreux M, Kim TY, Kudo M, Breder V, Merle P, Kaseb AO, Li D, Verret W, Xu DZ, et al, and IMbrave150 Investigators. Atezolizumab plus Bevacizumab in Unresectable Hepatocellular Carcinoma. *N Engl J Med.* 2020; 382:1894–905.
<https://doi.org/10.1056/NEJMoa1915745>
PMID:32402160
71. Rini BI, Powles T, Atkins MB, Escudier B, McDermott DF, Suarez C, Bracarda S, Stadler WM, Donskov F, Lee JL, Hawkins R, Ravaud A, Alekseev B, et al, and IMmotion151 Study Group. Atezolizumab plus bevacizumab versus sunitinib in patients with previously untreated metastatic renal cell carcinoma (IMmotion151): a multicentre, open-label, phase 3, randomised controlled trial. *Lancet.* 2019; 393:2404–15.
[https://doi.org/10.1016/S0140-6736\(19\)30723-8](https://doi.org/10.1016/S0140-6736(19)30723-8)
PMID:31079938
72. Rivera LB, Bergers G. Intertwined regulation of angiogenesis and immunity by myeloid cells. *Trends Immunol.* 2015; 36:240–9.
<https://doi.org/10.1016/j.it.2015.02.005>
PMID:25770923
73. Trenti A, Tedesco S, Boscaro C, Trevisi L, Bolego C, Cignarella A. Estrogen, Angiogenesis, Immunity and Cell Metabolism: Solving the Puzzle. *Int J Mol Sci.* 2018; 19:859.
<https://doi.org/10.3390/ijms19030859>
PMID:29543707
74. Kasprzak A. Angiogenesis-Related Functions of Wnt Signaling in Colorectal Carcinogenesis. *Cancers (Basel).* 2020; 12:3601.

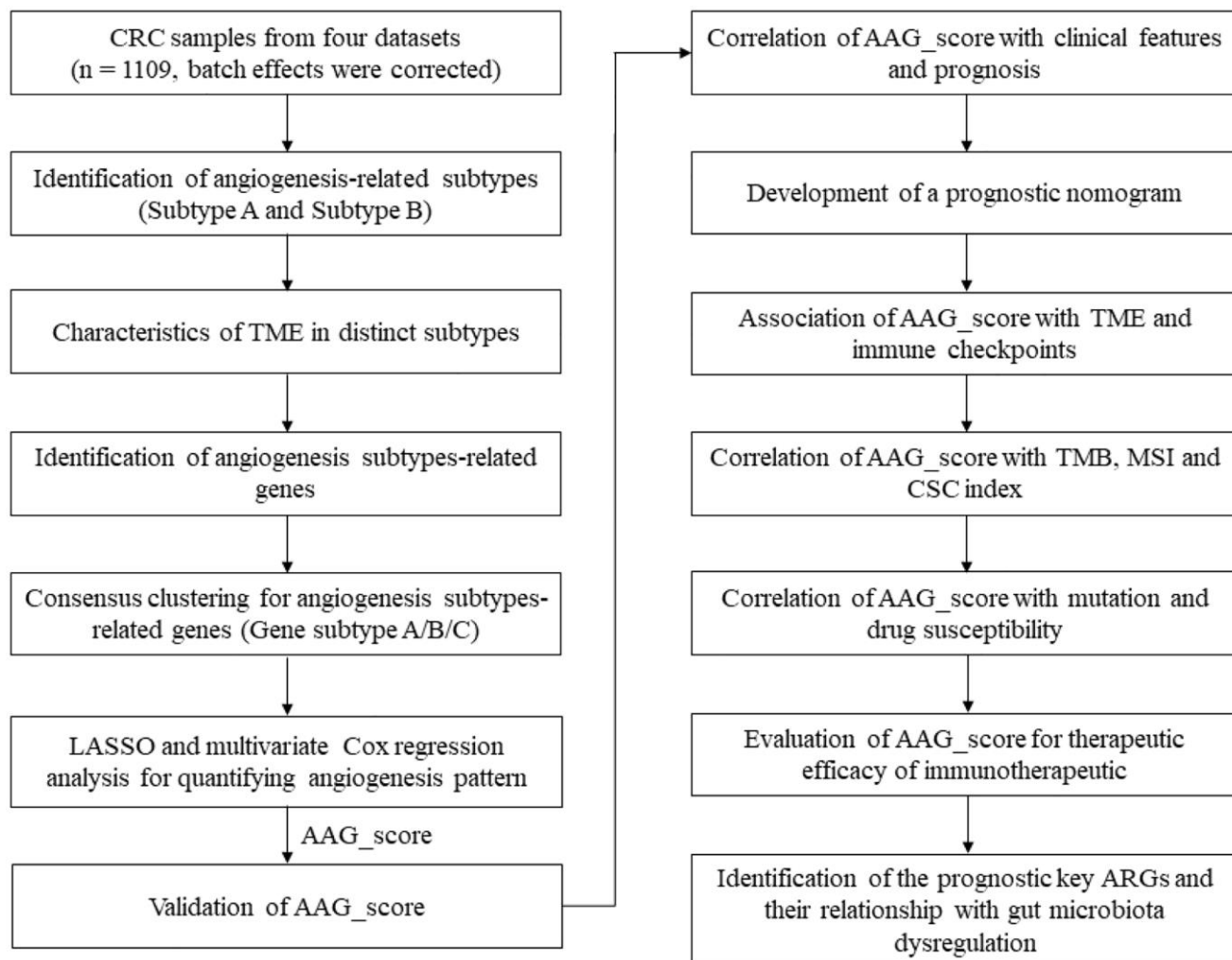
- <https://doi.org/10.3390/cancers12123601>
PMID:33276489
75. Zeng Z, Li Y, Pan Y, Lan X, Song F, Sun J, Zhou K, Liu X, Ren X, Wang F, Hu J, Zhu X, Yang W, et al. Cancer-derived exosomal miR-25-3p promotes pre-metastatic niche formation by inducing vascular permeability and angiogenesis. *Nat Commun.* 2018; 9:5395.
<https://doi.org/10.1038/s41467-018-07810-w>
PMID:30568162
76. Brenner H, Kloor M, Pox CP. Colorectal cancer. *Lancet.* 2014; 383:1490–502.
[https://doi.org/10.1016/S0140-6736\(13\)61649-9](https://doi.org/10.1016/S0140-6736(13)61649-9)
PMID:24225001
77. Li K, Luo H, Huang L, Luo H, Zhu X. Microsatellite instability: a review of what the oncologist should know. *Cancer Cell Int.* 2020; 20:16.
<https://doi.org/10.1186/s12935-019-1091-8>
PMID:31956294
78. Saeed A, Salem ME. Prognostic value of tumor mutation burden (TMB) and INDEL burden (IDB) in cancer: current view and clinical applications. *Ann Transl Med.* 2020; 8:575.
<https://doi.org/10.21037/atm-2020-75>
PMID:32566602
79. Ganesh K. Optimizing immunotherapy for colorectal cancer. *Nat Rev Gastroenterol Hepatol.* 2022; 19:93–4.
<https://doi.org/10.1038/s41575-021-00569-4>
PMID:34907331
80. Brahmer JR, Drake CG, Wollner I, Powderly JD, Picus J, Sharfman WH, Stankevich E, Pons A, Salay TM, McMiller TL, Gilson MM, Wang C, Selby M, et al. Phase I study of single-agent anti-programmed death-1 (MDX-1106) in refractory solid tumors: safety, clinical activity, pharmacodynamics, and immunologic correlates. *J Clin Oncol.* 2010; 28:3167–75.
<https://doi.org/10.1200/JCO.2009.26.7609>
PMID:20516446
81. Shan J, Han D, Shen C, Lei Q, Zhang Y. Mechanism and strategies of immunotherapy resistance in colorectal cancer. *Front Immunol.* 2022; 13:1016646.
<https://doi.org/10.3389/fimmu.2022.1016646>
PMID:36238278
82. Hinshaw DC, Shevde LA. The Tumor Microenvironment Innately Modulates Cancer Progression. *Cancer Res.* 2019; 79:4557–66.
<https://doi.org/10.1158/0008-5472.CAN-18-3962>
PMID:31350295
83. Turley SJ, Cremasco V, Astarita JL. Immunological hallmarks of stromal cells in the tumour microenvironment. *Nat Rev Immunol.* 2015; 15:669–82.
<https://doi.org/10.1038/nri3902>
PMID:26471778
84. Seager RJ, Hajal C, Spill F, Kamm RD, Zaman MH. Dynamic interplay between tumour, stroma and immune system can drive or prevent tumour progression. *Converg Sci Phys Oncol.* 2017; 3:034002.
<https://doi.org/10.1088/2057-1739/aa7e86>
PMID:30079253
85. Pan Y, Yu Y, Wang X, Zhang T. Tumor-Associated Macrophages in Tumor Immunity. *Front Immunol.* 2020; 11:583084.
<https://doi.org/10.3389/fimmu.2020.583084>
PMID:33365025
86. Sousa S, Brion R, Lintunen M, Kronqvist P, Sandholm J, Mönkkönen J, Kellokumpu-Lehtinen PL, Lauttia S, Tynnen O, Joensuu H, Heymann D, Määttä JA. Human breast cancer cells educate macrophages toward the M2 activation status. *Breast Cancer Res.* 2015; 17:101.
<https://doi.org/10.1186/s13058-015-0621-0>
PMID:26243145
87. Ma R, Yuan D, Guo Y, Yan R, Li K. Immune Effects of $\gamma\delta$ T Cells in Colorectal Cancer: A Review. *Front Immunol.* 2020; 11:1600.
<https://doi.org/10.3389/fimmu.2020.01600>
PMID:33013819
88. Zhang L, Yu X, Zheng L, Zhang Y, Li Y, Fang Q, Gao R, Kang B, Zhang Q, Huang JY, Konno H, Guo X, Ye Y, et al. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature.* 2018; 564:268–72.
<https://doi.org/10.1038/s41586-018-0694-x>
PMID:30479382
89. Göschl L, Scheinecker C, Bonelli M. Treg cells in autoimmunity: from identification to Treg-based therapies. *Semin Immunopathol.* 2019; 41:301–14.
<https://doi.org/10.1007/s00281-019-00741-8>
PMID:30953162
90. Kalinska M, Meyer-Hoffert U, Kantyka T, Potempa J. Kallikreins - The melting pot of activity and function. *Biochimie.* 2016; 122:270–82.
<https://doi.org/10.1016/j.biochi.2015.09.023>
PMID:26408415
91. Srinivasan S, Kryza T, Batra J, Clements J. Remodelling of the tumour microenvironment by the kallikrein-related peptidases. *Nat Rev Cancer.* 2022; 22:223–38.
<https://doi.org/10.1038/s41568-021-00436-z>
PMID:35102281
92. Alexopoulou DK, Papadopoulos IN, Scorilas A. Clinical significance of kallikrein-related peptidase (KLK10) mRNA expression in colorectal cancer. *Clin Biochem.* 2013; 46:1453–61.
<https://doi.org/10.1016/j.clinbiochem.2013.03.002>
PMID:23499583

93. Wei H, Dong C, Shen Z. Kallikrein-related peptidase (KLK10) cessation blunts colorectal cancer cell growth and glucose metabolism by regulating the PI3K/Akt/mTOR pathway. *Neoplasma*. 2020; 67:889–97.
https://doi.org/10.4149/neo_2020_190814N758
PMID:[32386481](https://pubmed.ncbi.nlm.nih.gov/32386481/)

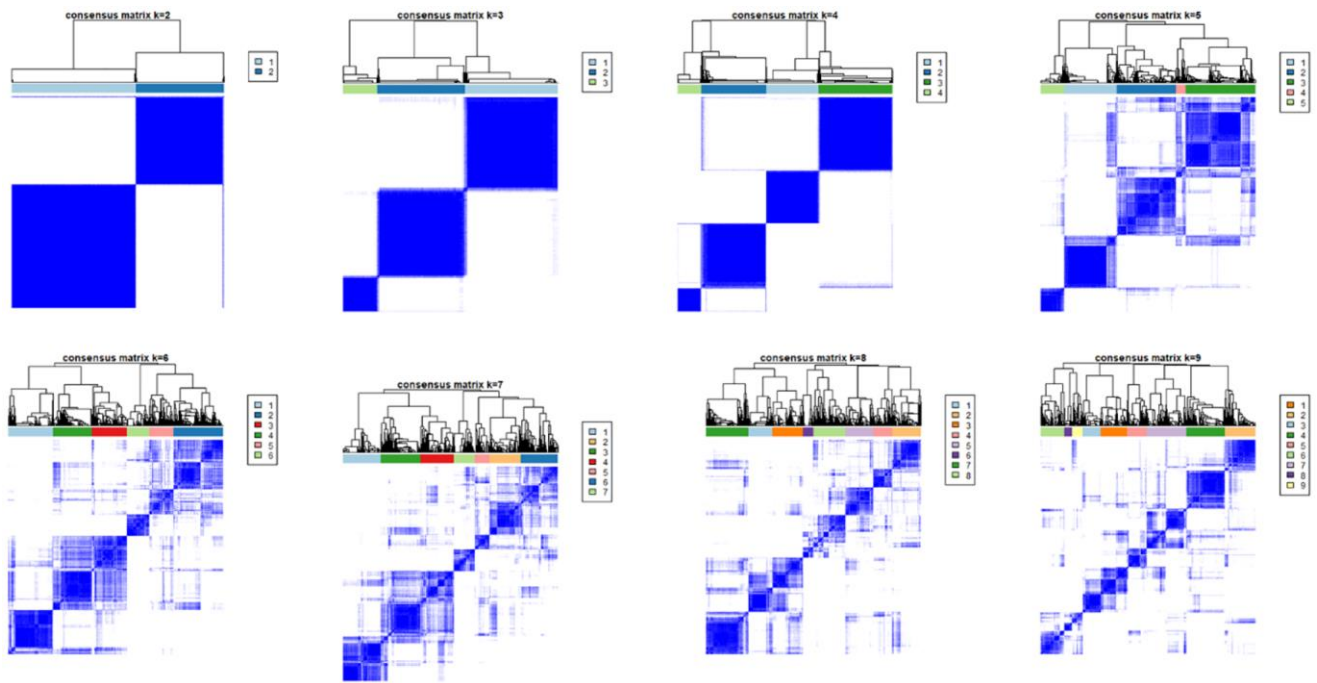
94. Schuster H, Peper JK, Bösmüller HC, Röhle K, Backert L, Bilich T, Ney B, Löffler MW, Kowalewski DJ, Trautwein N, Rabsteyn A, Engler T, Braun S, et al. The immunopeptidomic landscape of ovarian carcinomas. *Proc Natl Acad Sci U S A*. 2017; 114:E9942–51.
<https://doi.org/10.1073/pnas.1707658114>
PMID:[29093164](https://pubmed.ncbi.nlm.nih.gov/29093164/)

SUPPLEMENTARY MATERIALS

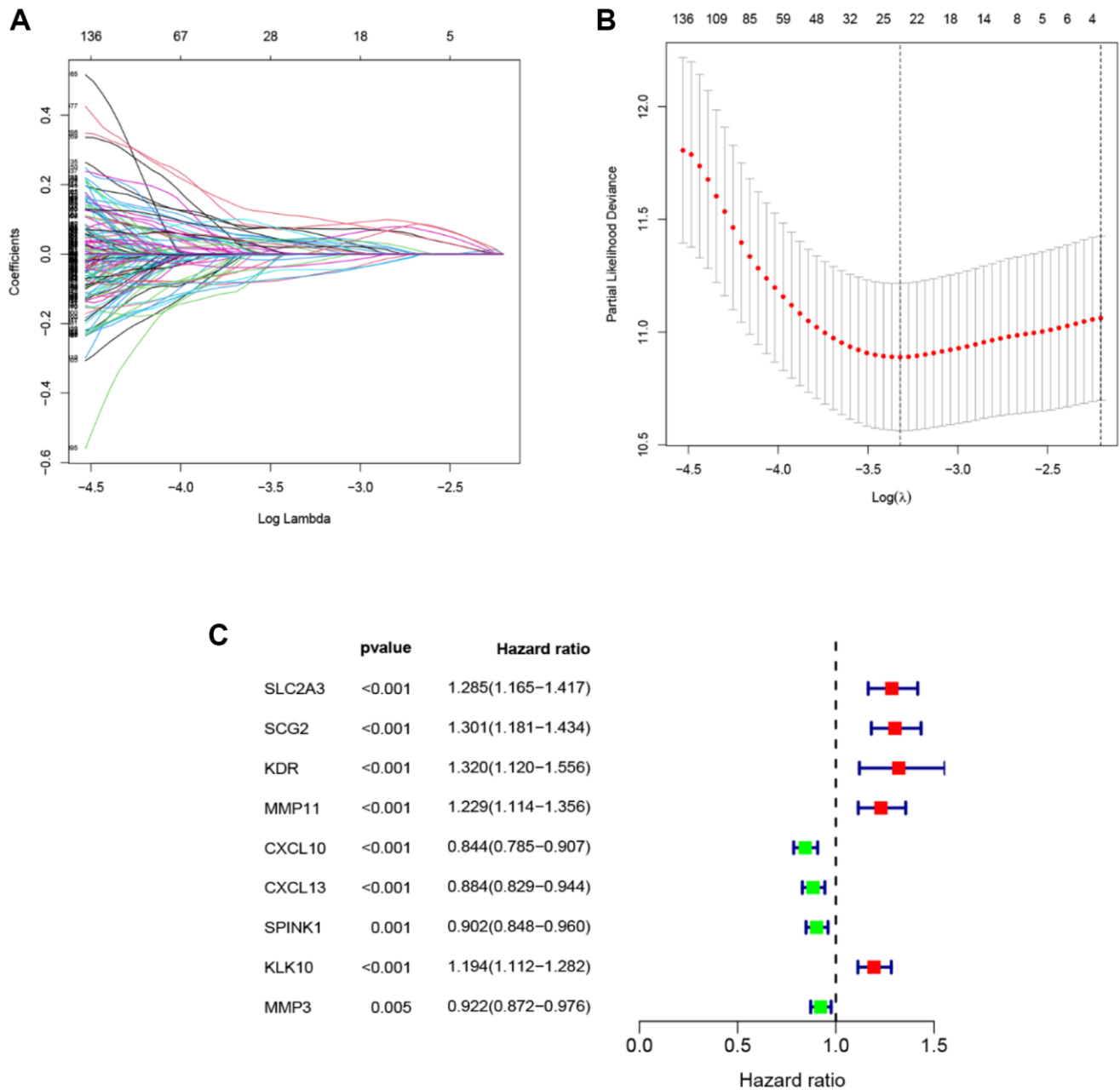
Supplementary Figures



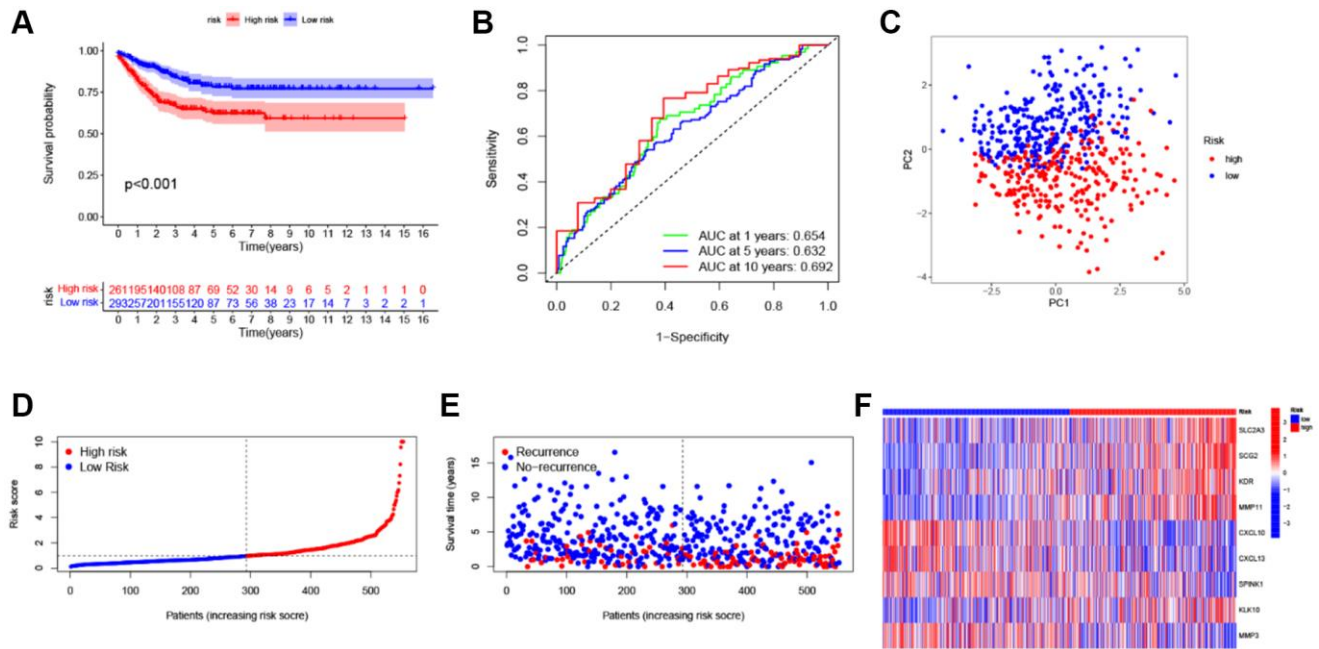
Supplementary Figure 1. The specific workflow graph of data analysis.



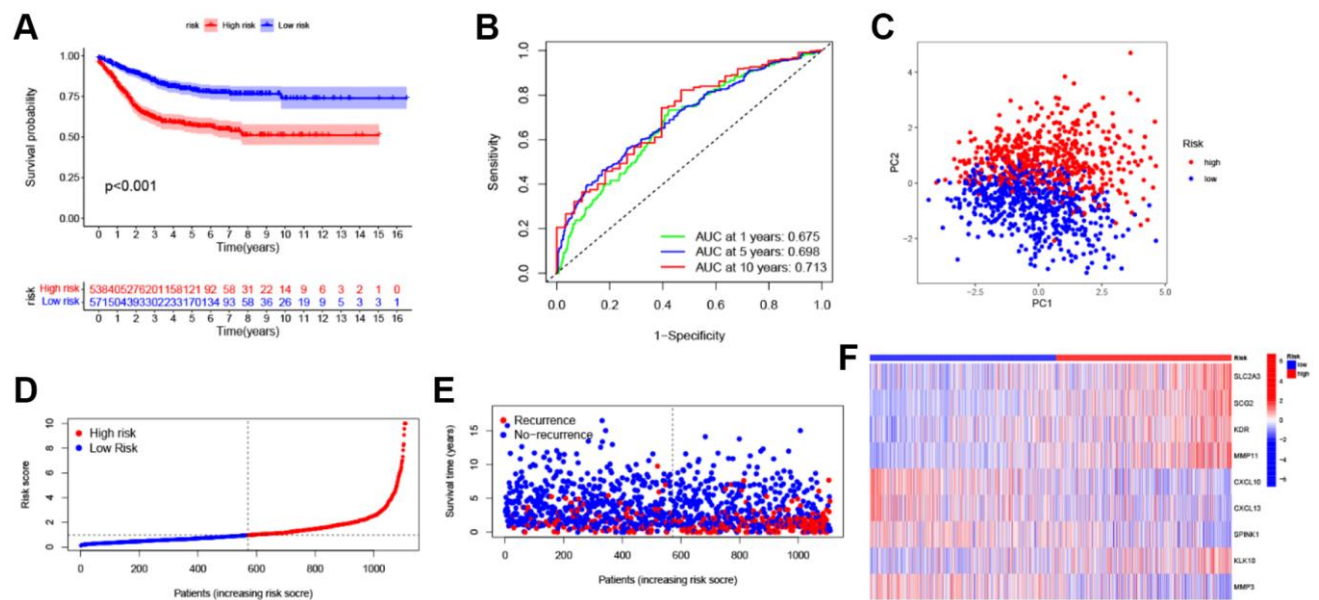
Supplementary Figure 2. Unsupervised clustering of gene subtypes based on DEGs and Consensus matrix heatmaps for $k = 2-9$.



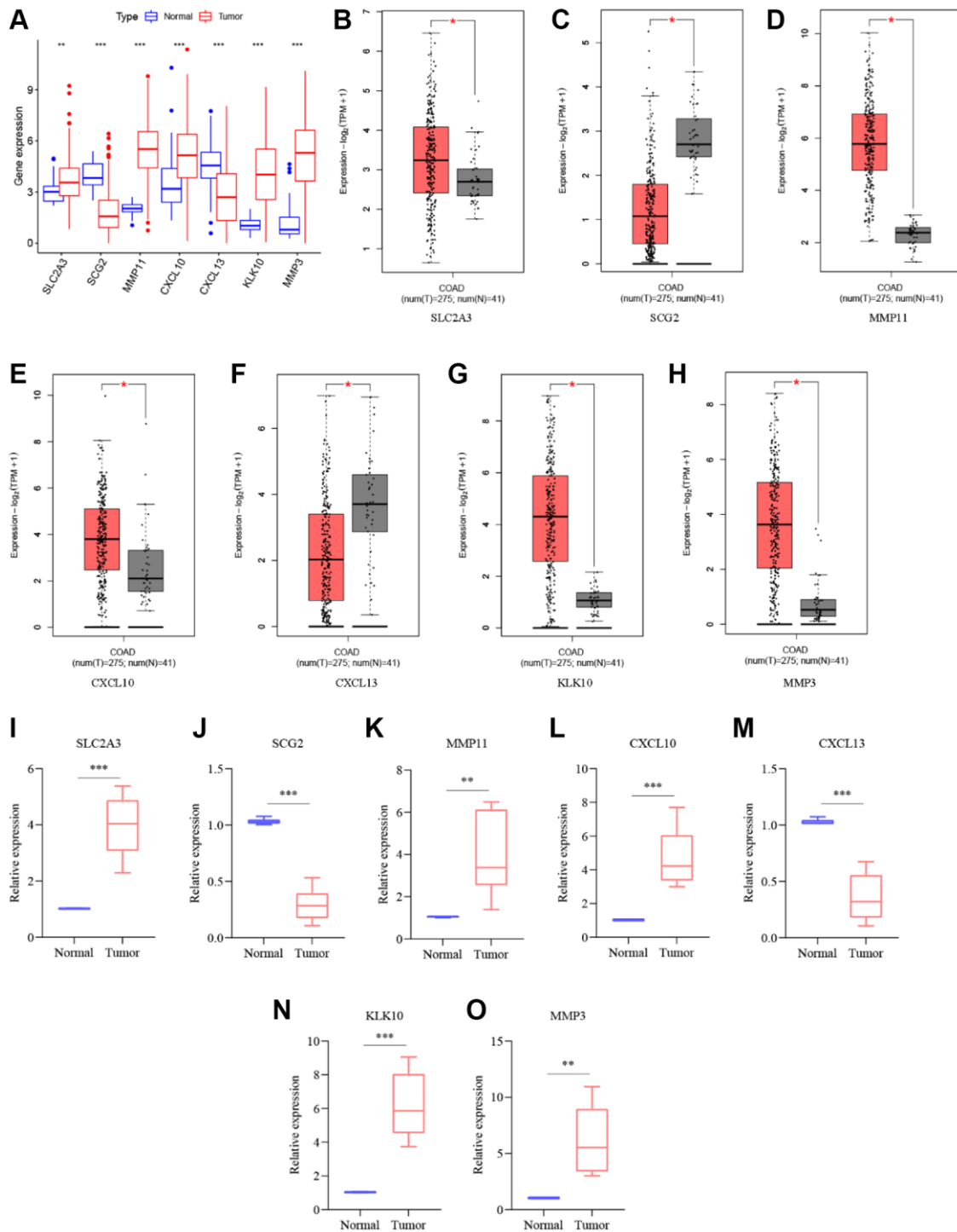
Supplementary Figure 3. Identifying representative candidate prognostic genes. (A, B) The LASSO regression analysis and partial likelihood deviance on the prognostic genes. **(C)** Forest plot of multivariate cox regression analysis for prognostic genes.



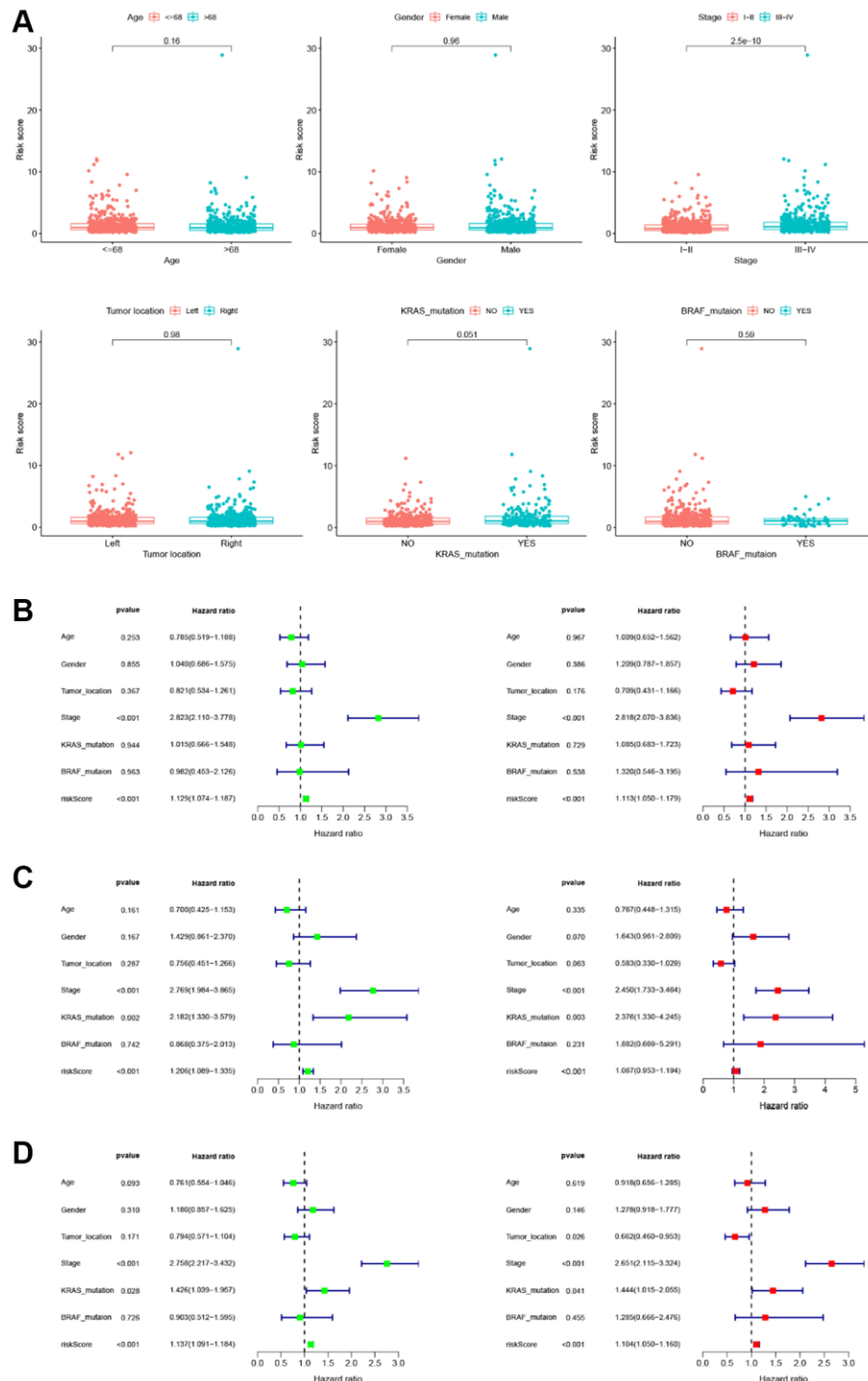
Supplementary Figure 4. Validation of ARG_score in testing cohort. (A) K-M analysis of the RFS between the two ARG_score groups. (B) ROC curves to predict the sensitivity and specificity of 1-, 5-, and 10-year survival according to the ARG_score. (C) The PCA analysis demonstrated that the patients in the different ARG_score groups were distributed in two directions. (D, E) The ranked dot plot indicates the ARG_score distribution and scatter plot presenting the patients' survival status. (F) Expression patterns of 9 selected prognostic genes in high- and low-ARG_score groups.



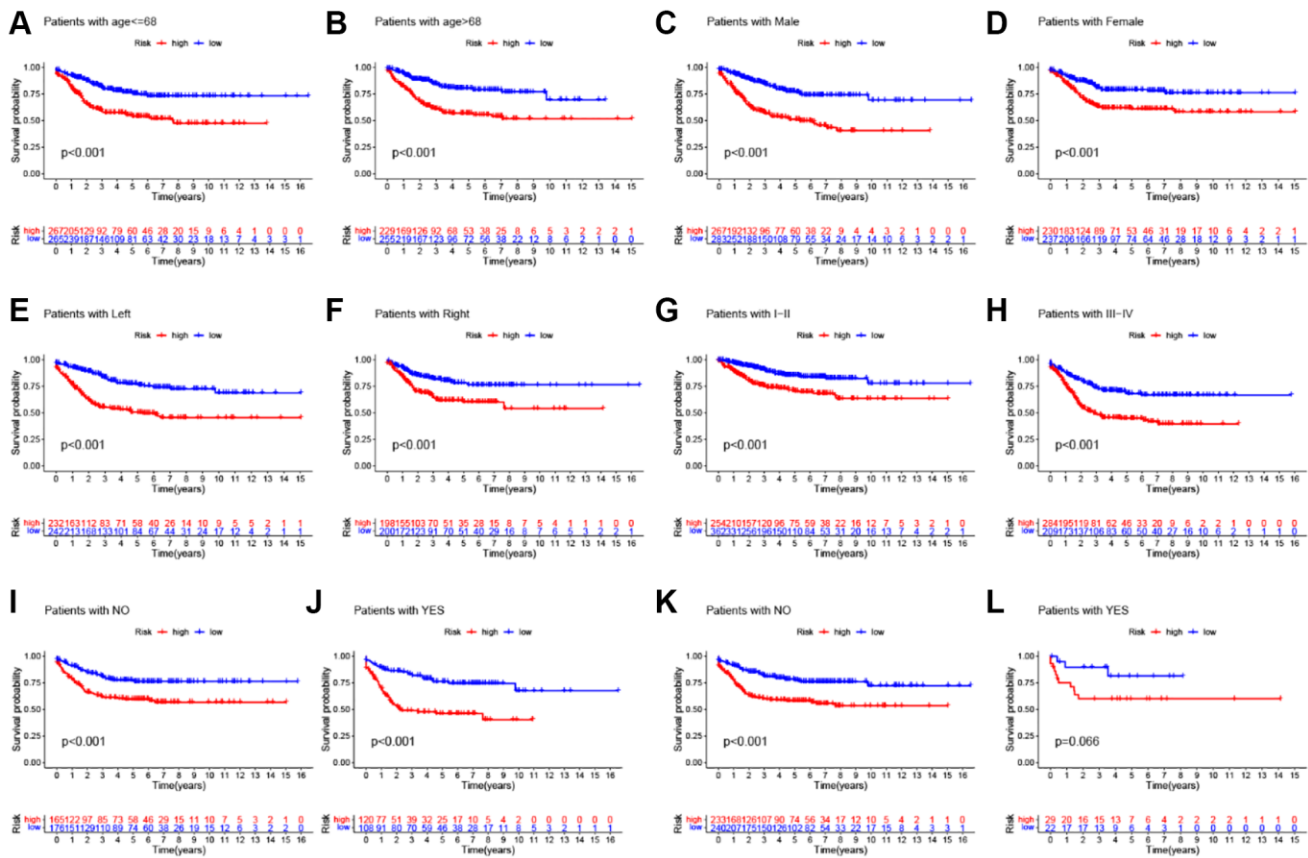
Supplementary Figure 5. Validation of ARG_score in entire cohort. (A) K-M analysis of the RFS between the two ARG_score groups. (B) ROC curves to predict the sensitivity and specificity of 1-, 5-, and 10-year survival according to the ARG_score. (C) The PCA analysis demonstrated that the patients in the different ARG_score groups were distributed in two directions. (D, E) The ranked dot plot indicates the ARG_score distribution and scatter plot presenting the patients' survival status. (F) Expression patterns of 9 selected prognostic genes in high- and low-ARG_score groups.



Supplementary Figure 6. Expression levels of 9 selected angiogenesis-related prognostic genes in CRC tissues and corresponding normal tissues. (A) Expression of 9 prognostic genes in CRC tissues and normal human colon tissues (TCGA cohort). **(B–H)** Expression of 9 prognostic genes in CRC tissues and normal human colon tissues (GEPIA 2 cohort). **(I–O)** Expression levels of 9 prognostic genes in CRC tissues and corresponding normal tissues by RT-PCR. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.



Supplementary Figure 7. The correlation and independent prognosis analysis of ARG_score and clinicopathological variables in CRC. (A) The correlation between the ARG_score and clinical characteristics, including age, gender, TNM stage, tumor location, KRAS mutation status, and BRAF mutation status. **(B)** Univariate and multivariate analyses showed the prognostic value of the ARG_score in the training cohort. **(C)** Univariate and multivariate analyses showed the prognostic value of the ARG_score in the testing cohort. **(D)** Univariate and multivariate analyses showed the prognostic value of the ARG_score in the entire cohort.



Supplementary Figure 8. Stratification analysis of the ARG_score in CRC. (A, B) Age (age ≤ 68 and age > 68 years old). (C, D) Gender (female and male). (E, F) Tumor location (left-side or right-side). (G, H) Tumor stage (I-II or III-IV). (I, J) KRAS mutation (yes and no). (K, L) BRAF mutation (yes and no).

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 3, 5 and 6.

Supplementary Table 1. Summary of 36 recognized ARGs.

Gene	Type
VCAN	Angiogenesis
POSTN	Angiogenesis
FSTL1	Angiogenesis
LRPAP1	Angiogenesis
STC1	Angiogenesis
LPL	Angiogenesis
VEGFA	Angiogenesis
PF4	Angiogenesis
THBD	Angiogenesis
FGFR1	Angiogenesis
TNFRSF21	Angiogenesis
CCND2	Angiogenesis
COL5A2	Angiogenesis
ITGAV	Angiogenesis
SERPINA5	Angiogenesis
KCNJ8	Angiogenesis
APP	Angiogenesis
JAG1	Angiogenesis
COL3A1	Angiogenesis
SPP1	Angiogenesis
NRP1	Angiogenesis
OLR1	Angiogenesis
PDGFA	Angiogenesis
PTK2	Angiogenesis
SLCO2A1	Angiogenesis
PGLYRP1	Angiogenesis
VAV2	Angiogenesis
S100A4	Angiogenesis
MSX1	Angiogenesis
VTN	Angiogenesis
TIMP1	Angiogenesis
APOH	Angiogenesis
PRG2	Angiogenesis
JAG2	Angiogenesis
LUM	Angiogenesis
CXCL6	Angiogenesis

Supplementary Table 2. The primer sequences for qRT-PCR.

Gene	Primer Sequence
CXCL13	F: 5'-GGTCAGCAGCCTCTCTCCAGTC-3' R: 5'-TTGAATTCGATCAATGAAGCGT-3'
SLC2A3	F: 5'-GAAGAGGAGAATGCTAAG-3' R: 5'-CAATGGAAATGATGATGG-3'
CXCL10	F: 5'-GTGGCATTCAAGGAGTACCTC-3' R: 5'-GCCTTCGATTCTGGATTCAGACA-3'
KDR	F: 5'-CTACTGATTTTTTGCCCTTGTTTC-3' R: 5'-TAGTCATTGTTCCCAGCATTTC-3'
KLK10	F: 5'-TCTACCCTGGCGTGGTCACC-3' R: 5'-GCAGAGCCACAGGGGTAAACAC-3'
MMP11	F: 5'-AAGAGGTTTCGTGCTTTCTGG-3' R: 5'-CCATGGGAACCGAAGGAT-3'
MMP3	F: 5'-CCTGCTTTGTCTTTGATGC-3' R: 5'-TGAGTCAATCCCTGGAAAGTC-3'
SCG2	F: 5'-ACCAGACCTCAGGTTGGAAAA-3' R: 5'-ACCAGACCTCAGGTTGGAAAA-3'
SPINK1	F: 5'-TGTCTGTGGACTGATGGAA-3' R: 5'-GCCCAGATTTTTGAATGAGG-3'
β -actin	F: 5'-GCATGGAGTCCTGTGGCAT-3' R: 5'-CTAGAAGCATTTGCGGTGG-3'
siKLK10#1	5'-CCTCCACACCTCTAAACATCTC-3'
siKLK10#2	5'-TTGTTGTACTTCACTCTCCGG-3'
siKLK10-NC	5'-UUCUCCGAACGUGUCACGUTT-3'

Supplementary Table 3. Clinical information of 1109 colorectal cancer patients.**Supplementary Table 4. The prognostic values of 36 ARGs in patients with CRC.**

id	HR	HR.95L	HR.95H	P-value	km
VCAN	1.224994274	1.11397604	1.347076523	2.83E-05	1.19E-06
POSTN	1.181163635	1.081386881	1.29014653	0.000217649	1.66E-06
FSTL1	1.273841191	1.118843592	1.450311189	0.0002558	5.61E-06
LRPAP1	0.67653805	0.510799323	0.896053914	0.006421073	0.000186841
STC1	1.201646156	1.066394221	1.354052241	0.002568988	0.000953075
LPL	1.19409106	1.094679196	1.302530881	6.34E-05	2.38E-05
VEGFA	1.466477431	1.219837129	1.762986225	4.60E-05	2.23E-06
PF4	0.947950341	0.865382536	1.038396098	0.250296529	0.048438335
THBD	1.222489097	1.040031194	1.436956507	0.01485467	0.005430823
FGFR1	1.31569863	1.108971297	1.56096275	0.001655545	0.000456971
TNFRSF21	1.165541157	0.937967838	1.448329178	0.166925139	0.008317478
CCND2	0.905160617	0.809083578	1.012646611	0.081779674	0.000854507
COL5A2	1.237035349	1.121895833	1.363991568	1.98E-05	5.01E-07

ITGAV	1.249982819	1.078549708	1.448664847	0.00303005	0.003795799
SERPINA5	1.197708641	1.027155793	1.396580732	0.021345418	0.001044201
KCNJ8	1.389973359	1.213983433	1.591476364	1.87E-06	1.27E-08
APP	1.062376145	0.833601656	1.353935739	0.624820263	0.078671415
JAG1	1.016142325	0.838207208	1.231849612	0.870487009	0.030863117
COL3A1	1.241335579	1.121609276	1.373842079	2.94E-05	1.20E-07
SPP1	1.116360262	1.047764384	1.189447029	0.000668835	7.87E-05
NRP1	1.317014905	1.13423922	1.529243769	0.000303391	2.38E-05
OLR1	1.17399583	1.078246613	1.278247659	0.000219433	0.00037697
PDGFA	0.913533642	0.77915275	1.071091278	0.265285904	0.110789116
PTK2	1.307826879	1.020229861	1.676495868	0.034172484	0.009839317
SLCO2A1	1.076135281	0.939249225	1.232971093	0.290480775	0.053553315
PGLYRP1	0.916923323	0.703850283	1.194498888	0.52036185	0.141275291
VAV2	1.139619561	0.966985174	1.343074102	0.118902964	0.008376223
S100A4	1.198276931	1.084160975	1.324404436	0.00039635	7.44E-05
MSX1	0.959960251	0.833859773	1.105130278	0.569546779	0.069710173
TIMP1	1.390530524	1.186145976	1.630132527	4.81E-05	2.03E-05
APOH	0.978210944	0.820736312	1.165900225	0.805686095	0.116002365
PRG2	0.982731271	0.71948208	1.342299939	0.912807991	0.133462913
JAG2	1.088750856	0.937652326	1.264198246	0.26465296	0.009779996
LUM	1.140747631	1.03752703	1.254237354	0.006503175	0.001225002
CXCL6	0.942957149	0.863324081	1.029935575	0.191984201	0.002126871

Supplementary Table 5. Spearman correlation analysis of the 36 ARGs.

Supplementary Table 6. Prognostic analysis of 620 subtype-related genes using a univariate Cox regression model.