Construction and analysis of a reliable five-gene prognostic signature for colon adenocarcinoma associated with the wild-type allelic state of the *COL6A6* gene

Qun Liu[1], Xiaohua Zhang[2], Yan Song[3], Junli Si[1], Zhaoshui Li[4], Quanjiang Dong[5]

[1]Second Department of Gastroenterology, Qingdao Municipal Hospital, Dalian Medical University, Qingdao, China; [2]Gastroenterology Center, Qingdao Traditional Chinese Medicine Hospital (Qingdao Hiser Hospital), Qingdao Hiser Hospital Affiliated of Qingdao University, Qingdao, China; [3]Outpatient Department, Qingdao Traditional Chinese Medicine Hospital (Qingdao Hiser Hospital), Qingdao Hiser Hospital Affiliated of Qingdao University, Qingdao, China; [4]Qingdao University, Qingdao Medical College, Qingdao, China; [5]Central Laboratories, Department of Gastroenterology, Qingdao Municipal Hospital, Dalian Medical University, Qingdao, China

Contributions: (I) Conception and design: Q Liu; (II) Administrative support: Q Dong; (III) Provision of study materials or patients: Q Liu; (IV) Collection and assembly of data: X Zhang, Y Song, Z Li; (V) Data analysis and interpretation: Q Liu, X Zhang, Y Song, J Si, Z Li; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Quanjiang Dong, MD. Central Laboratories, Department of Gastroenterology, Qingdao Municipal Hospital, Dalian Medical University, No. 5 Donghai Middle Road, Qingdao 266071, China. Email: jiangacer@126.com.

Background: Tumors emerge by acquiring a number of mutations over time. The first mutation provides a selective growth advantage compared to adjacent epithelial cells, allowing the cell to create a clone that can outgrow the cells that surround it. Subsequent mutations determine the risk of the tumor progressing to metastatic cancer. Some secondary mutations may inhibit the aggressiveness of the tumor while still increasing the survival of the clone. Meaningful mutations in genes may provide a strong molecular foundation for developing novel therapeutic strategies for cancer.

Methods: The somatic mutation and prognosis in colon adenocarcinoma (COAD) were analyzed. The copy number variation (CNV) and differentially expressed genes (DEGs) between the collagen type VI alpha 6 chain (*COL6A6*) mutation (*COL6A6*-MUT) and the *COL6A6* wild-type (*COL6A6*-WT) subgroups were evaluated. The independent prognostic signatures based on *COL6A6*-allelic state were determined to construct a Cox model. The biological characteristics and the immune microenvironment between the two risk groups were compared.

Results: *COL6A6* was found to be highly mutated in COAD at a frequency of 9%. Patients with *COL6A6*-MUT had a good overall survival (OS) compared to those with *COL6A6*-WT, who had a different CNV pattern. Significant differences in gene expression were established for 593 genes between the *COL6A6*-MUT and *COL6A6*-WT samples. Among them, *MUC16*, *ASNSP1*, *PRR18*, *PEG10*, and *RPL26P8* were determined to be independent prognostic factors. The internally validated prognostic risk model, constructed using these five genes, demonstrated its value by revealing a significant difference in patient prognosis between the high-risk and low-risk groups. Specifically, patients in the high-risk group exhibited a considerably worse prognosis than did those in the low-risk group. The high-risk group had a significantly higher proportion of patients over 60 years of age and patients in stage III. Moreover, the tumor immune dysfunction and exclusion (TIDE) score and the expression of human leukocyte antigen (HLA) family genes were all higher in the high-risk group than that in the low-risk group.

Conclusions: The allelic state of *COL6A6* and the five associated DEGs were identified as novel biomarkers for the diagnosis and prognosis of COAD and may be therapeutic targets in COAD.

Keywords: Colon adenocarcinoma (COAD); somatic mutation; collagen type VI alpha 6 chain (*COL6A6*); prognostic signature; biomarker

Introduction

Background

Colorectal cancer (CRC) is among the leading causes of mortality worldwide, causing approximately 700,000 deaths each year. Among CRC, colon adenocarcinoma (COAD) is a major subtype (1,2) and together with lung, prostate, and breast cancer, is considered among the most lethal cancers (1). In addition to surgical treatment, adjuvant (3,4) (e.g., platinum drugs) or targeted therapy (e.g., cetuximab and bevacizumab) (5) is usually used to treat this common disease. These approaches have demonstrated a good activity and efficacy, mainly when combined with other therapies; however, the 5-year overall survival (OS) rate for COAD remains low (63.5%) (6).

Literature review and study rationale

Although most of the mutations that accumulate in cells are harmless, occasionally, a mutation can alter a regulatory element. This may have phenotypic consequences (7), among which is tumor formation. Tumorigenesis is a complex evolutionary process, in which one class of genes undergoes mutation and inactivation while other genes undergo amplification and activation. A variety of chemotherapeutic drugs targeting this phenomenon have been applied in clinical treatment and obtained good

results. In-depth research into the altered expression of tumor prognostic marker genes may provide a productive perspective for improving the cure rate of cancer.

Collagen type VI alpha 6 chain (*COL6A6*) encodes the α6 (VI) protein chain of collagen VI, a component of the extracellular matrix (ECM) that is present in almost all connective tissues (8). Alterations in ECM organization or composition have been observed to be associated with human disease, such as liver fibrosis (9), cardiovascular disease (10,11), and cancer (12-14), and dysregulation of collagen VI members including *COL6A6* has been reported in cancer (15). Recent studies indicate that there is a significantly aberrant expression of *COL6A6* in breast cancer (16) and that *COL6A6* is a prognostic gene for survival among patients with lung adenocarcinoma (LUAD) (17). However, the cancer-related function of the *COL6A6* gene in COAD remains largely unknown.

Objective

In this study, the general landscape of somatic mutation and the corresponding prognosis of patients with COAD was evaluated. A Cox risk model was constructed based on the differentially expressed genes (DEGs) associated with the *COL6A6* allelic state. The putative and critical functionality of *COL6A6* in COAD was examined via an evaluation of clinical features, and the immune escape scores between high- and low-risk groups, that were compared and analyzed using a bioinformatics approach. We present this article in accordance with the TRIPOD reporting checklist (available at https://tcr.amegroups.com/article/view/10.21037/tcr-23-463/rc).

Methods

Data acquisition

Masked somatic mutation data of patients with COAD (n=433) were acquired from the Genomic Data Commons (GDC) Data Portal (https://portal.gdc.cancer.gov/) of The Cancer Genome Atlas (TCGA). VarScan software was used to preprocess the original data, and the "maftools" R package (The R Foundation for Statistical Computing, Vienna, Austria) (18) was used to visualize the landscape of

---

Highlight box

Key findings
• The collagen type VI alpha 6 chain (*COL6A6*) allelic state was associated with the overall survival of patients with colon adenocarcinoma (COAD) according to the altered expression of a five-gene prognostic signature.

What is known and what is new?
• Germ line mutation of cancer-related genes is an important factor in the development and progression of COAD.
• *COL6A6*-allelic state may be a novel diagnostic biomarker for COAD.

What is the implication, and what should change now?
• Screening for genetic mutation should be expanded, and additional genetic mutations associated with tumorigenesis and development should be analyzed in depth.

somatic mutation.

The somatic mutation data and clinical information of COAD were downloaded from the International Cancer Genome Consortium (ICGC) database (https://dcc.icgc.org/), and the data with no survival information and incomplete tumor-node-metastasis (TNM) staging information were excluded. Ultimately, the data of 331 patients were retained.

Gene expression sequencing data (HTSEQ-counts) of COAD (n=514) were acquired and converted into transcripts per million (TPM) values. The clinical data of TCGA-COAD matched patients (n=452) were downloaded and obtained with GDC software and included age, survival status, follow-up time, staging, etc. Data from patients with no survival information or incomplete TNM staging information were excluded, after which the data from 443 patients were retained for the construction of the prognostic model. For the identification of DEGs associated with *COL6A6* alteration, data from normal people (n=41) were excluded, and thus data from 474 patients were retained.

Copy number variation (CNV) analysis

Masked copy number segment (n=453) data were downloaded via the TCGA "biolinks" R package (19), and Genomic Identification of Significant Targets in Cancer (GISTIC) 2.0 analysis (20) was performed with the GenePattern platform (https://www.genepattern.org/) (21).

Somatic mutation analysis and calculation of tumor mutational burden (TMB) score

For each tumor sample, the total number of somatic mutations (except silent mutations) detected in the tumor was defined as the TMB (22). The TMB score for each sample was calculated, and the Wilcoxon rank-sum test was used to statistically analyze the difference between the *COL6A6* mutation (*COL6A6*-MUT) and *COL6A6* wild-type (*COL6A6*-WT) groups.

Identification of DEGs and functional analysis

HTSEQ-counts of COAD acquired from TCGA were used to screen the DEGs with the threshold of |fold change| ≥2 and a P value <0.05 with the "Deseq 2" R package (23). The "ggplot 2" and "Pheatmap" R packages were used to visualize the DEGs. The "clusterProfiler" R package (24) was used to conduct Gene Ontology (GO) (25) and Kyoto

Encyclopedia of Genes and Genomes (KEGG) (26) pathway enrichment analyses of the DEGs The "clusterProfiler" package was also used to conduct gene set enrichment analysis (GSEA) (27) for the gene expression matrix, with "C2.cp.all.v7.0.symbols" being selected as the reference gene set. A false discovery rate (FDR) <0.25 and a P value <0.05 were considered to indicate significant enrichment. The "Tidyverse" R package was used to conduct gene set variation analysis (GSVA) analysis (28), with "C2.cp.all.v7.0.symbols" being selected as the reference gene set. The scores of the related pathways were calculated according to the gene expression matrix of each sample via the single-sample GSEA method, and the results were visualized in a heatmap.

Establishment and validation of a COL6A6 alteration-related gene prognostic signature

Based on the TCGA-COAD dataset, univariate Cox regression, least absolute shrinkage and selection operator (LASSO) regression, and multivariate Cox regression analysis were used to screen the prognostic genes and establish the prognostic model. The "Survival" R package was used to calculate the association between the expression of each DEG and OS, and genes with a P value <0.05 were retained for the subsequent LASSO regression analysis. The "Glmnet" and "Survival" R packages were used for the LASSO regression analysis to screen out the significant variables for the univariate Cox regression analysis. In order to obtain more accurate independent prognostic factors (prognostic characteristic genes), multivariate Cox regression analysis was used for the final screening. The risk score was calculated as the follow: risk score = (exp-gene1 × coef-gene1) + (exp-gene2 × coef-gene2) + ... + (exp-gene n × coef-gene n), where "exp" refers to the gene expression, and "coef" refers to the gene coefficient. Patients were divided into high- and low-risk groups based on the median of the risk score.

The "Survival" R package was used for the Kaplan-Meier analysis and log-rank test to analyze the OS of the two risk groups. To measure the accuracy of prognostic prediction, the time-dependent receiver operating characteristic (ROC) was determined, and the "timeROC" R package (29) was used to calculate the area under the ROC curve (AUC).

Immune cell infiltration and immunotherapy correlation analysis

CIBERSORT, based on linear support vector regression,

was used to deconvolve the transcriptome expression matrix to estimate the composition and abundance of immune cells in mixed cells (30). The gene expression matrix data were uploaded to CIBERSORT, and the samples with a P value <0.05 were filtered out to obtain the immune cell infiltration matrix. A histogram was drawn using the "ggplot 2" R package to show the infiltration distribution of 22 immune cells in each sample. Stromal score, immune score, and estimate score were calculated based on messenger RNA (mRNA) expression via the "Estimate" R package (31). The potential tumor therapeutic response was predicted according to the tumor immune dysfunction and exclusion (TIDE) score, which is a type of computing algorithm based on gene expression profiles (http://tide.dfci.harvard.edu) (32). The differences of tumor immunotherapy indicators between the two risk groups were determined based on the TIDE score.

Statistical analysis

The statistical comparison of TMB, microsatellite instability (MSI), TIDE score, and immune escape score between the two risk groups was completed via the Wilcoxon rank-sum test. All statistical tests were bilateral, and all visualizations of statistical data were performed with R software (version 4.0.2).

Ethical statement

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) (33).

Results

The somatic mutation of COL6A6 was associated with a good prognosis in COAD

To analyze the genomic alteration level in COAD, the public biological databases of TCGA and ICGC were searched for the genomic alteration information of patients with COAD. We then analyzed the global alteration of the genome in TCGA-COAD and ICGC-COAD, and the results showed that missense mutations accounted for the majority of all variant classifications (Figure S1A-S1D). For the single CNVs, the variant of C>T was the most frequent (Figure S1E,S1F), which is possibly attributable to the local hypermethylation in cancer cells. In addition, *TTN*, *MUC16*, *APC*, *KRAS*, *TP53*, and *SYNE1* were the genes

with the highest alteration frequency (Figure S1G,S1H). This indicates that the tumorigenesis of COAD is a complex and evolutionary process involving a disorder of the regulation network in the cellular and microenvironment, which is caused by multiple genetic alterations.

According to these genomic alterations, we performed prognostic analyses of genes with somatic alterations and found that the alteration of *COL6A6* was associated with a good outcome in patients with COAD consistently across both databases (*Figure 1A,1B*). The results of COL6A6 protein amino acid analysis showed that the main mutation form of this protein was missense mutation both in TCGA-COAD and ICGC-COAD databases (*Figure 1C,1D*). As the missense mutation accounted for the largest proportion among the all alteration types of *COL6A6*, missense mutation was used to represent the alteration of this gene in the subsequent analyses. Most mutation sites were located in the functional domain of the protein (*Figure 1C,1D*). Based on the mutation level of *COL6A6*, TCGA-COAD samples were divided into the *COL6A6*-MUT (n=42, listed in table available at https://cdn.amegroups.cn/static/public/tcr-23-463-1.xlsx) and (*COL6A6*-WT, n=431, listed in table available at https://cdn.amegroups.cn/static/public/tcr-23-463-1.xlsx) groups, and the single nucleotide polymorphisms (SNPs) and CNVs in the two groups were analyzed. The result showed that patients in the *COL6A6*-MUT group exhibited a higher TMB than did those in the *COL6A6*-WT group (*Figure 1E,1F*, Figure S2). Moreover, the CNVs in the *COL6A6*-MUT group (*Figure 1G,1H*) was markedly different than were those in the *COL6A6*-WT group (*Figure 1I,1J*), suggesting that *COL6A6*-MUT has a crucial regulatory role in the SNP and CNV levels in COAD. In the *COL6A6*-WT group, copy number amplification occurred at multiple locations (*Figure 1I*), while this phenotype was not present in the *COL6A6*-MUT group (*Figure 1G*). Moreover, the chromosome locations with copy number deletion in the *COL6A6*-MUT group (*Figure 1H*) were significantly different from those in the *COL6A6*-WT group (*Figure 1J*). These findings indicate that *COL6A6* may be involved in the tumorigenesis and progression of COAD.

Identification of DEGs associated with the COL6A6 allelic state

To examine the downstream molecular events and biological functions of *COL6A6* during COAD tumorigenesis and progression, the DEGs between the *COL6A6*-MUT and
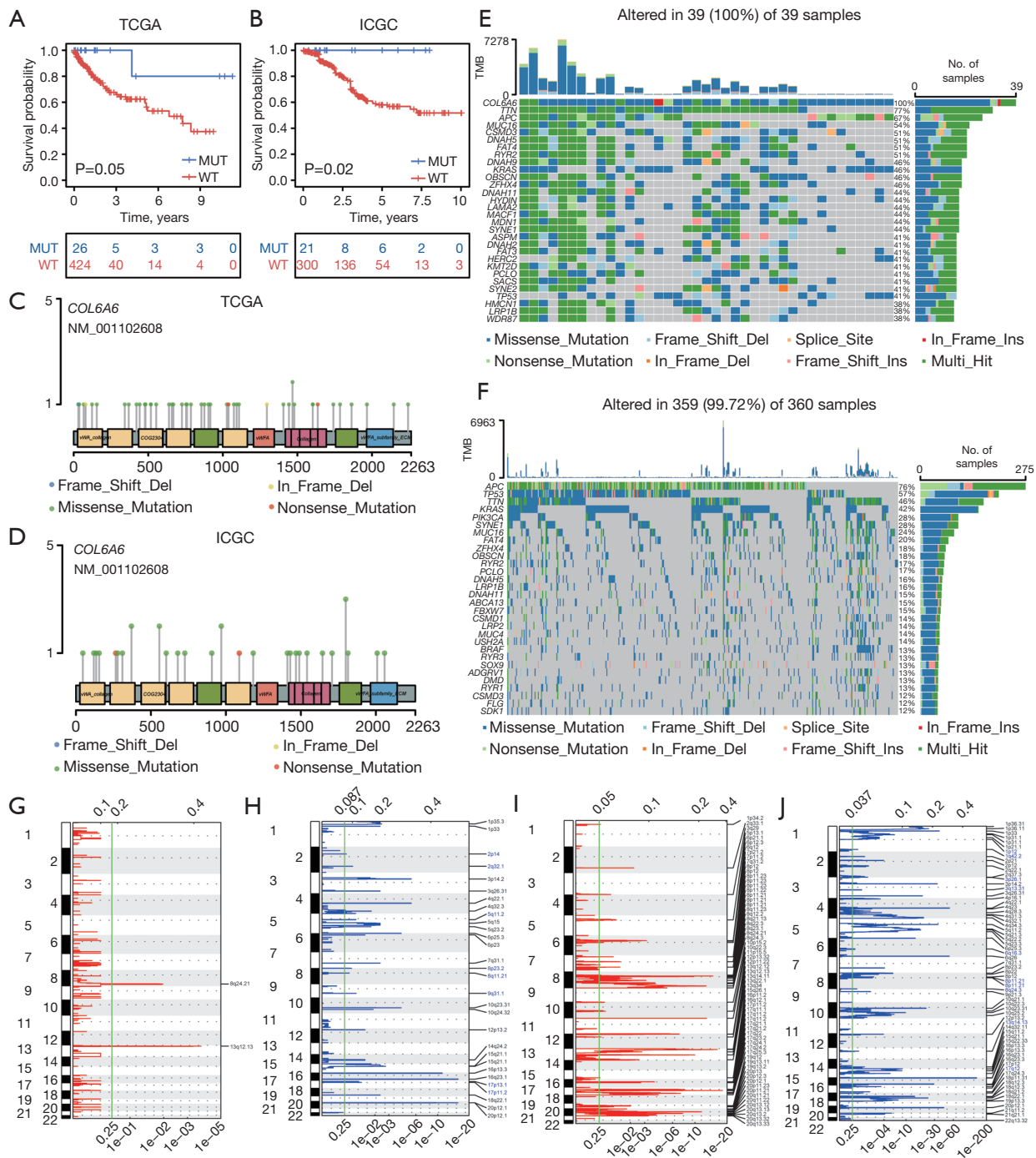
Figure 1 The genome mutation of *COL6A6* was associated with good prognosis and exhibited higher TMB. Prognostic analysis of *COL6A6*-MUT in the cohorts of TCGA-COAD (A) and ICGC-COAD (B). The mutation site information of *COL6A6* in the cohorts of TCGA-COAD (C) and ICGC-COAD (D). Top 30 mutant genes in *COL6A6*-MUT samples (E) and *COL6A6*-WT samples (F) in TCGA-COAD. (G) Amplification of the genome in *COL6A6*-MUT samples. (H) Deletion of the genome in *COL6A6*-MUT samples. (I) Amplification of the genome in *COL6A6*-WT samples. (J) Deletion of the genome in *COL6A6*-WT samples. The portion above the horizontal coordinate represents the G-score, which is used to quantify the amplification and deletion, while the bottom represents q-value. TCGA, The Cancer Genome Atlas; MUT, mutation; WT, wild-type; ICGC, International Cancer Genome Consortium; *COL6A6*, collagen type VI alpha 6 chain; TMB, tumor mutational burden; COAD, colon adenocarcinoma.

Transl Cancer Res 2024;13(5):2475-2496 | https://dx.doi.org/10.21037/tcr-23-463
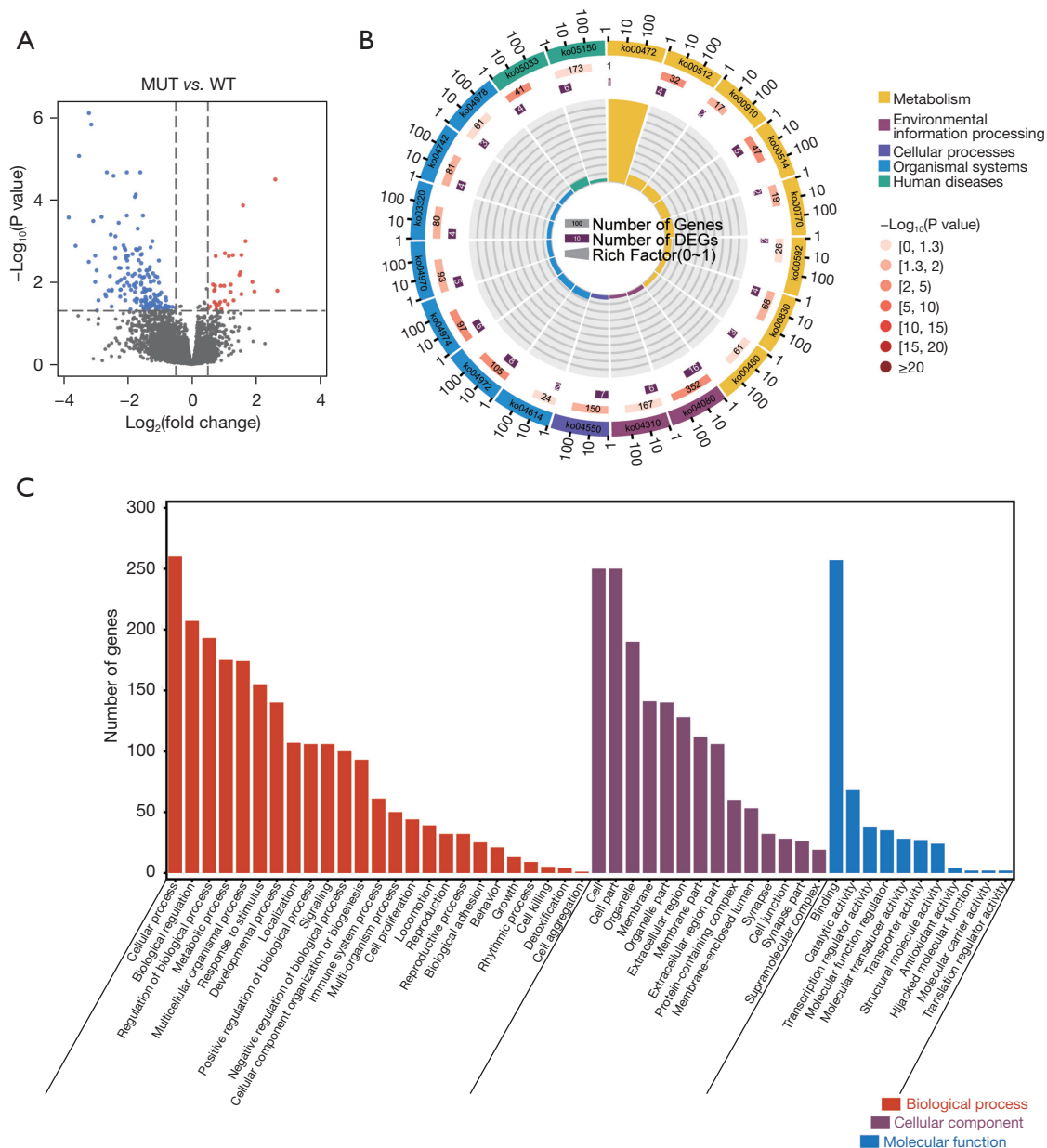
Figure 2 Identification of DEGs associated with the *COL6A6* allelic state. (A) Volcano plot showing the DEGs in between the *COL6A6*-MUT *COL6A6*-WT groups. (B) KEGG analysis of the *COL6A6* allelic state-associated genes. (C) Enriched GO terms of *COL6A6* allelic state-associated genes. MUT, mutation; WT, wild-type; DEGs, differentially expressed genes; *COL6A6*, collagen type VI alpha 6 chain; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology.

COL6A6-MUT groups were extracted and identified from TCGA-COAD gene expression matrix via R software. After screening (fold change =2, P value <0.05) was applied, there were 531 downregulated (*COL6A6*-MUT *vs.* *COL6A6*-WT) (*Figure 2A,2B*, listed in table available at https:// cdn.amegroups.cn/static/public/tcr-23-463-2.xlsx) and

62 upregulated genes (*COL6A6*-MUT *vs.* *COL6A6*-WT) (*Figure 2A*, table available at listed in https://cdn.amegroups. cn/static/public/tcr-23-463-3.xlsx). The subsequent KEGG analysis showed that the DEGs present in the *COL6A6*-MUT group were related to metabolism (e.g., D-arginine and D-ornithine metabolism, mucin type O-glycan

**Table 1** KEGG pathways of DEGs in the *COL6A6*-MUT and *COL6A6*-WT groups

ID	Description	Gene ratio	P value
Ko00472	D-arginine and D-ornithine metabolism	1/1	0.01
Ko00512	Mucin type O-glycan biosynthesis	4/32	0.001
Ko00910	Nitrogen metabolism	2/17	0.03
Ko00514	Other types of O-glycan biosynthesis	5/47	0.001
Ko00770	Pantothenate and CoA biosynthesis	2/19	0.04
Ko00592	Alpha-linolenic acid metabolism	2/26	0.07
Ko00830	Retinol metabolism	4/68	0.02
Ko00480	Glutathione metabolism	3/61	0.08
Ko04080	Neuroactive ligand-receptor interaction	16/352	<0.001
Ko04310	Wnt signaling pathway	6/167	0.06
Ko04550	Signaling pathways regulating pluripotency of stem cells	7/150	0.01
Ko04614	Renin-angiotensin system	2/24	0.06
Ko04972	Pancreatic secretion	8/105	<0.001
Ko04974	Protein digestion and absorption	6/97	0.005
Ko04970	Salivary secretion	5/93	0.02
Ko03320	PPAR signaling pathway	4/80	0.04
Ko04742	Taste transduction	4/81	0.04
Ko04978	Mineral absorption	3/61	0.08
Ko05033	Nicotine addiction	4/41	0.004
Ko05150	Staphylococcus aureus infection	6/173	0.08

KEGG, Kyoto Encyclopedia of Genes and Genomes; DEGs, differentially expressed genes; *COL6A6*, collagen type VI alpha 6 chain; MUT, mutation; WT, wild-type; PPAR, peroxisome proliferator activated receptor.

biosynthesis), environmental information processing (e.g., neuroactive ligand-receptor interaction and Wnt signaling pathway), cellular processes (e.g., signaling pathways regulating the pluripotency of stem cells), organismal system (e.g., renin-angiotensin system and pancreatic secretion), and human disease (e.g., nicotine addiction and *Staphylococcus aureus* infection) (*Figure 2B*, *Table 1*). GO analysis showed that the DEGs present in the *COL6A6*-MUT group were related to cellular process, biological regulation, metabolic process, multicellular organismal process, immune system process, and cell proliferation (*Figure 2C*). These findings suggest that *COL6A6* is responsible for the intercellular signal transduction and immune regulation of cancer cells.

Establishment of COL6A6 allelic state-related gene prognostic signature in TCGA cohort

We next used the related DEGs to establish the prognostic signature. A total of 593 identified genes (*Figure 2*) were used in univariate Cox regression analysis to screen for prognostic genes associated with the *COL6A6* allelic state (tables available at https://cdn.amegroups.cn/static/public/tcr-23-463-4.xlsx), and 60 genes were eligible after screening (P value <0.05, table available at https://cdn.amegroups.cn/static/public/tcr-23-463-5.xlsx). Among these, 29 genes (P value <0.05, table available at https://cdn.amegroups.cn/static/public/tcr-23-463-6.xlsx) were chosen with zero LASSO regression coefficients (*Figure 3A,3B*),
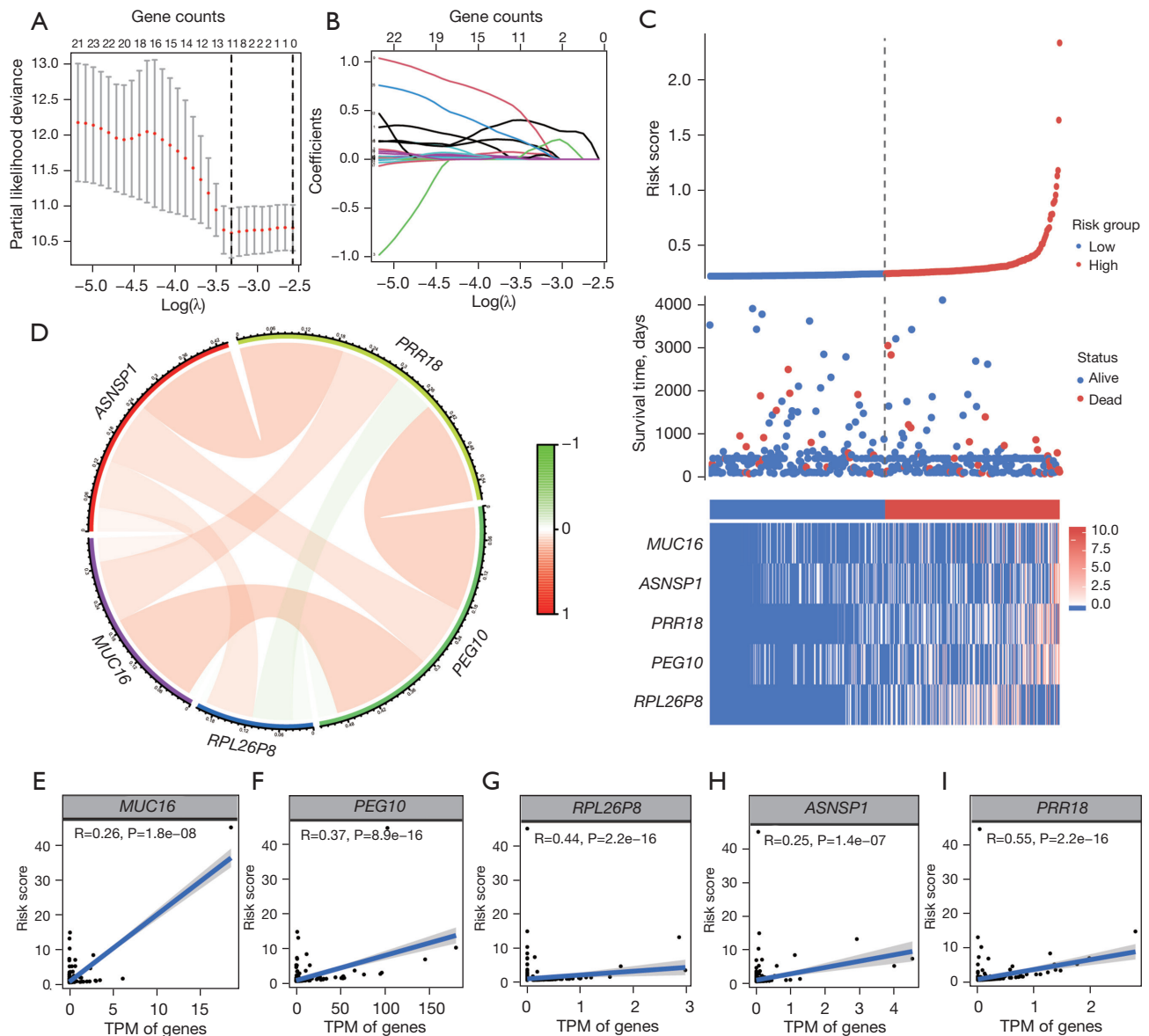
Figure 3 Establishment of the *COL6A6* allelic state-associated gene prognostic signature in TCGA cohort. (A) A stepwise Cox proportional risk regression model was used to screen the prognostic genes. (B) The LASSO coefficient spectrum of prognostic gene screening. (C) The rick score distribution, survival status of patients, and heatmap of prognostic gene distribution in the training cohort. (D) Correlation analysis of the prognostic genes in the *COL6A6* allelic state model. (E-I) Correlation analysis between the prognostic genes and risk score. TPM, transcripts per million; *COL6A6*, collagen type VI alpha 6 chain; TCGA, The Cancer Genome Atlas; LASSO, least absolute shrinkage and selection operator.

and an 11-gene risk score was built via multivariate Cox regression analysis (table available at https://cdn.amegroups.cn/static/public/tcr-23-463-7.xlsx). Finally, five genes that were all highly expressed in the high-risk group compared with the low-risk group (*Figure 3C*) were identified as independent prognostic factors: *MUC16*,

ASNSP1, *PRR18*, *PEG10*, and *RPL26P8* (*Figure 3C*, tables available at https://cdn.amegroups.cn/static/public/tcr-23-463-8.xlsx). Subsequently, we analyzed the expression correlation of these five genes, and found that *PRR18* was negatively correlated with *ROL26P8*, while the other genes were positively correlated with *ROL26P8* (*Figure 3D*, table

Transl Cancer Res 2024;13(5):2475-2496 | https://dx.doi.org/10.21037/tcr-23-463

available at https://cdn.amegroups.cn/static/public/tcr-23-463-9.xlsx). The five genes were all significantly and positively associated with the risk score (*Figure 3E-3I*), indicating that these prognostic genes are of great significance for the evaluation of COAD outcome.

Internal validation of the prognostic signature

According to the risk score, the OS analysis was conducted, and the prognosis of patients with a high-risk score was significantly poorer than that of those with a low-risk score (*Figure 4A*). A nomogram was drawn based on multivariate Cox analysis (*Figure 4B*), and the results showed that the expression levels of the five-gene prognostic signature predicted a poor outcome in COAD. The ROC curve (*Figure 4C*) was used to predict the prognosis at 1, 2, and 3 years, which showed that the model had good prediction efficacy (1-year AUC =0.635; 2-year AUC =0.687; 3-year AUC =0.646) (*Figure 4C*). The calibration analysis of the nomogram predicted probability also supported the accuracy of the Cox model (*Figure 4D*). Moreover, the clinical factors including age, tumor stage, and gender were analyzed via the risk score (*Figure 4E-4G*, table available at https://cdn.amegroups.cn/static/public/tcr-23-463-10.xlsx), which suggested that the high-risk group had a higher proportion of patients over 60 years of age (*Figure 4E*) and a slightly higher proportion of males compared with the low-risk group (*Figure 4G*). Of note, there was a significantly higher proportion of stage III patients in the high-risk group than in the low-risk group (*Figure 4F*).

Identification of DEGs associated with different risk group based on the COL6A6 allelic state

We then performed further analysis of DEGs in the high- and low-risk groups. The volcano plot showed that a total 2,921 genes (2,818 upregulated and 103 downregulated genes, listed in table available at https://cdn.amegroups.cn/static/public/tcr-23-463-11.xlsx) were significantly different in terms of expression level (*Figure 5A*). The KEGG analysis showed that the DEGs between the two groups were involved in the pathways of environment information processing (e.g., neuroactive ligand-receptor interaction and cAMP signaling pathway), organismal systems (e.g., taste transduction and mineral absorption), and human disease (e.g., nicotine addiction and Maturity onset diabetes of the young) (*Figure 5B*, *Table 2*). The GO analysis showed that the DEGs found between the two groups were involved

in the biological processes of cellular process, biological regulation, response to stimulus, multicellular organismal process, metabolic process, signaling, and immune system (*Figure 5C*).

We explored the functional enrichment of DEGs associated with different risk groups based on *COL6A6* alteration via GSEA. The results showed that the high-risk group was significantly activated for chylomicron assembly; plasma lipoprotein assembly; chylomicron remodeling; and metabolic pathway of low-density lipoprotein, high-density lipoprotein, and triglyceride disease (*Figure 6A-6D*). Meanwhile, the low-risk group was significantly activated for cholesterol biosynthesis pathway; terpenoid backbone biosynthesis; checkpoint kinase 1 (CHK1)-, checkpoint kinase 2 (CHK2)-, and CDP-diacylglycerol synthase 1 (CDS1)-mediated inactivation of cyclin b CDK1 complex; and transcription of E2F targets under negative control by dream complex (*Figure 6E-6H*). These findings suggest that the two groups with different risk levels based on *COL6A6*-MUT have different gene expression patterns, which may play a vital role in the OS of those with COAD.

GSVA of DEGs associated with different risk group based on the COL6A6 allelic state

To 10 determine the enrichment function of related genes in the high- and low-risk groups, GSVA was conducted (table available at https://cdn.amegroups.cn/static/public/tcr-23-463-12.xlsx). The results showed that pathways enriched in the low-risk group were intracellular metabolic regulation, mitophagy, ataxia telangiectasia mutated (ATM) pathway, and major histocompatibility complex (MHC)-mediated antigen-processing presentation, while pathways enriched in high-risk group were metabolism of amine-derived hormones, thyrocine biosynthesis, regulation of tp53 activity through association with cofactors, and neurotoxicity of *Clostridium* toxins (Figure S3), which was also verified by the subsequent correlation analysis with the risk score (*Figure 7*). These findings all indicate that the regulatory networks in the high-risk group are significantly different from those of the low-risk group in relation to the *COL6A6*-MUT, which may ultimately cause differences in therapeutic outcomes and tumor development.

### *Comparison of TMB, MSI, and TIDE scores between different risk groups*

According to the above risk score, COAD samples were
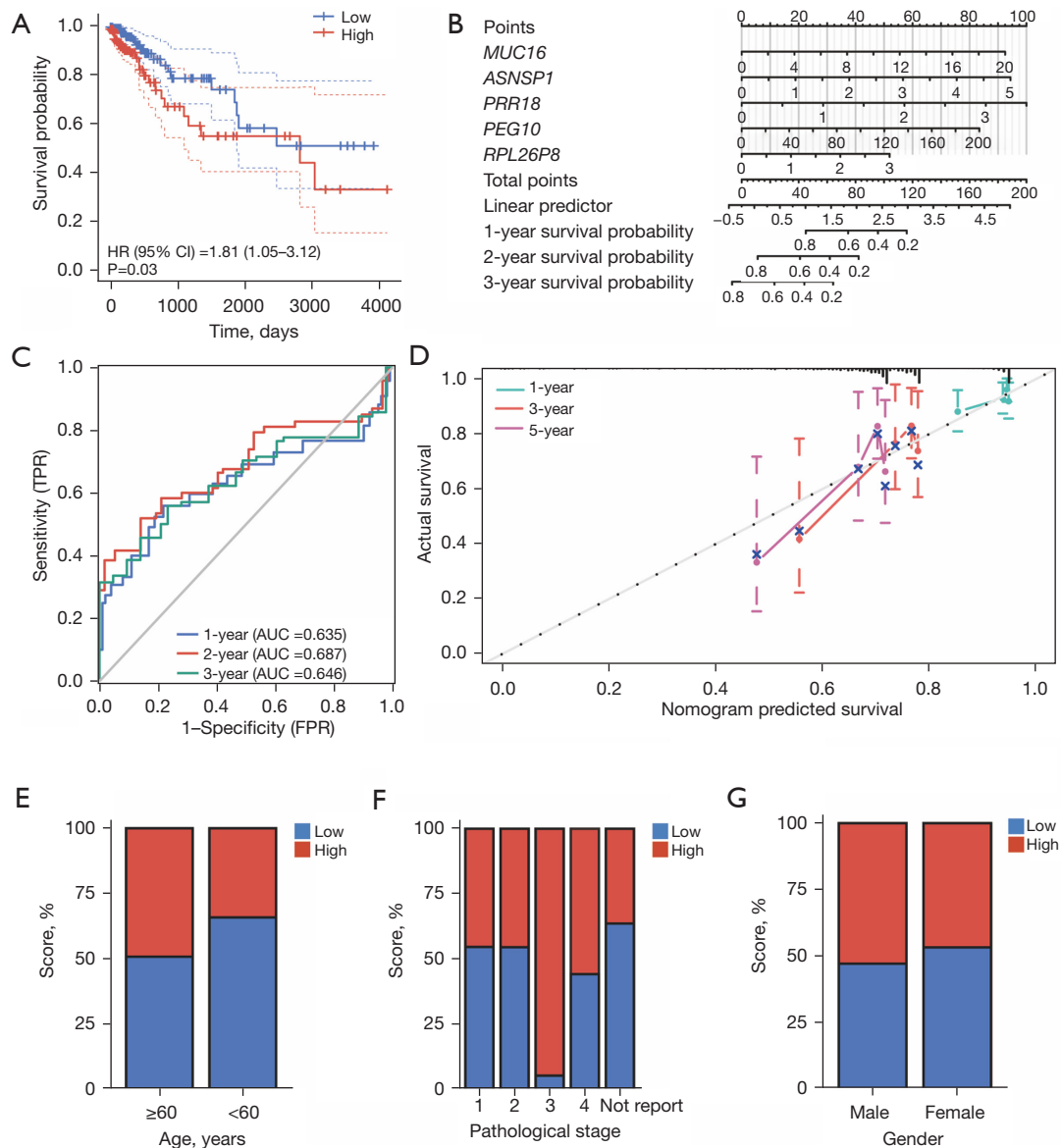
Figure 4 Internal validation of the prognostic signature. (A) The OS of the high- and low-risk group. (B) Nomogram result of the risk model. (C) ROC analyses of the model at 1, 2, and 3 years. (D) Risk groups in relation to clinical index (age, stage, and gender). (E) The percentage of the high- and low-risk groups by age. (F) The percentage of the high- and low-risk groups by stage. (G) The percentage of the high- and low-risk groups by gender. HR, hazard ratio; CI, confidence interval; TPR, true-positive rate; FPR, false-positive rate; AUC, area under the ROC curve; ROC, receiver operating characteristic; OS, overall survival.

divided into high-risk and low-risk groups. The TMB, total number of somatic mutations (except silent mutations), and MSI were further analyzed and showed that TMB (P value =0.51) and MSI (P value =0.71) scores were not statistically different between the high- and low-risk groups (*Figure 8A,8B*). However, TIDE scores, which reflect the effect of potential immunotherapy, were significantly

different between the high- and low-risk groups (P value =0.01) (*Figure 8C*). Moreover, other immune marker molecules, including CD274 (P value =0.17), CD8 (P value =0.07), and exclusion (an indicator of reactive immune escape) (P value =0.76) showed no significant difference (*Figure 8D-8F*). This indicates that the prognostic signature based on the *COL6A6*-MUT could predict the tumor
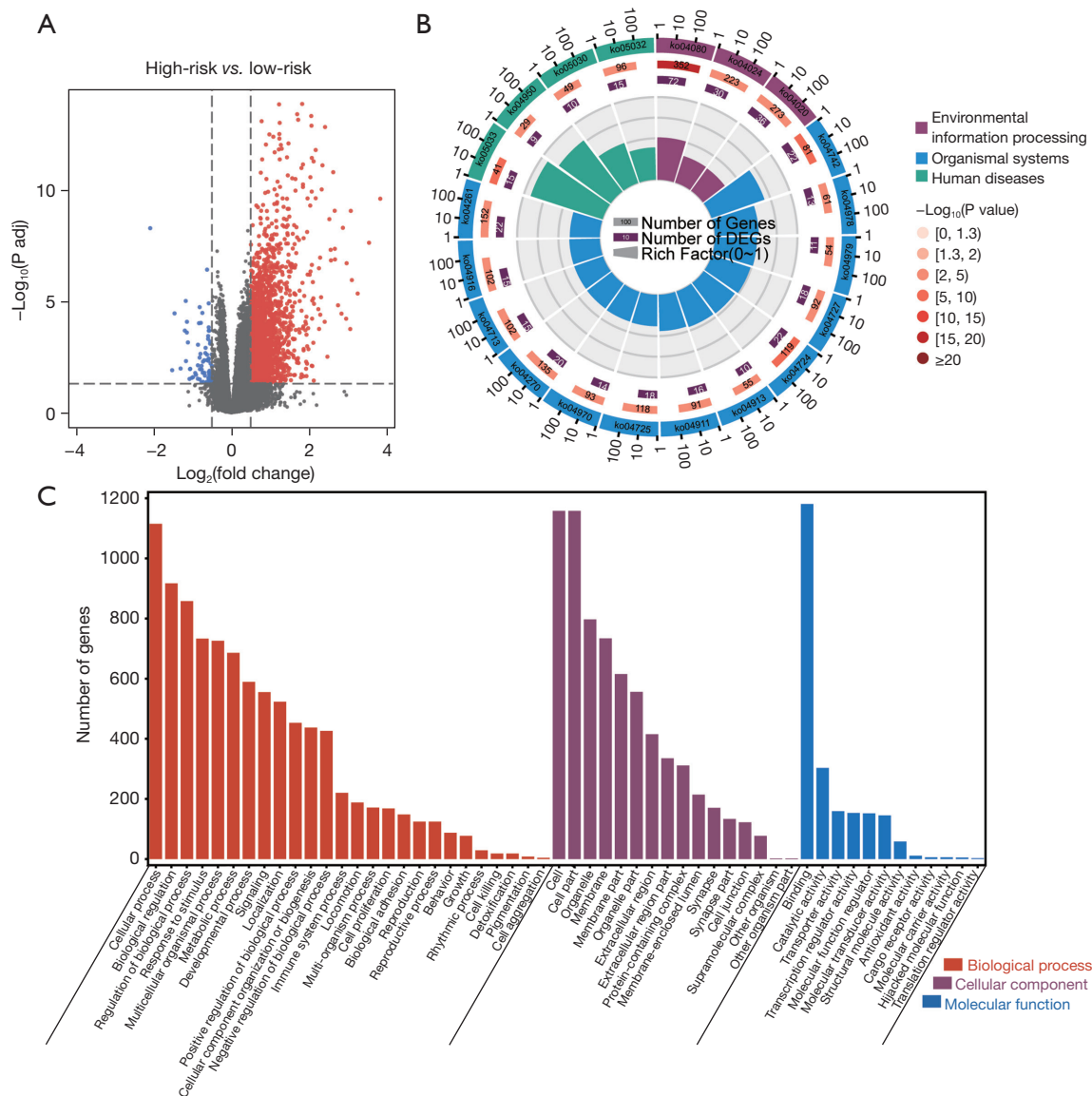
Figure 5 Identification of DEGs associated with the different risk groups based on the *COL6A6* allelic state. (A) Volcano plot showing the DEGs between the high-risk and low-risk group. (B) KEGG analysis of the DEGs. (C) Enriched GO terms of the DEGs. DEGs, differentially expressed genes; *COL6A6*, collagen type VI alpha 6 chain; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology.

immunotherapy effect, demonstrating the indispensable role of *COL6A6* in COAD.

The relationship between the tumor microenvironment (TME) and prognostic signature based on the COL6A6 allelic state

To analyze the relationship between the risk score and

immune cell infiltration in the COAD microenvironment, we calculated the proportion of immune cell infiltration in TME with the CIBERSORT algorithm. TME immune cell infiltration in COAD showed that M0, M1, M2 macrophages, T cells (CD8, naïve CD4, resting memory CD4, activated memory CD4), natural killer (NK) cells (resting and activated), and monocytes accounted for the main immune cells (Figure S4A). The subsequent immune

Table 2 KEGG pathways of DEGs between the high- and low-risk groups

ID	Description	Gene ratio	Adjusted P value
Ko04080	Neuroactive ligand-receptor interaction	72/352	<0.001
Ko04024	cAMP signaling pathway	30/223	0.005
Ko04020	Calcium signaling pathway	36/273	0.002
Ko04742	Taste transduction	22/81	<0.001
Ko04978	Mineral absorption	13/61	0.005
Ko04979	Cholesterol metabolism	11/54	0.01
Ko04727	GABAergic synapse	18/92	0.001
Ko04724	Glutamatergic synapse	22/119	<0.001
Ko04913	Ovarian steroidogenesis	10/55	0.04
Ko04911	Insulin secretion	16/91	0.04
Ko04725	Mineral absorption	3/61	0.08
Ko04970	Nicotine addiction	4/41	0.004
Ko04270	Staphylococcus aureus infection	6/173	0.008
Ko04713	Circadian entrainment	15/102	0.04
Ko04916	Melanogenesis	15/102	0.04
Ko04261	Adrenergic signaling in cardiomyocytes	22/152	0.01
Ko05033	Nicotine addiction	15/41	<0.001
Ko04950	Maturity onset diabetes of the young	9/29	0.002
Ko05030	Cocaine addiction	10/49	0.02
Ko05032	Morphine addiction	15/96	0.02

KEGG, Kyoto Encyclopedia of Genes and Genomes; DEGs, differentially expressed genes.

cell score correlation analysis showed that there were significant and positive correlations between activated NK cells and regulatory T cells (Tregs), M1 macrophages and follicular helper T cells, activated NK cells and M1 macrophages, and between activated NK cells and CD8 T cells (Figure S4B). We then compared levels of immune cell infiltration in the TME between the two risk groups, and the results showed that the high-risk group had a higher proportion of plasma cells and Tregs than did the low-risk group but lower proportion of resting CD4 memory T cells and M2 macrophages (*Figure 9A,9B*). We also applied the ESTIMATE algorithm to determine the stromal score and immune score to investigate their association with the risk score, but no significant differences in immune or stromal score between the two risk groups were found (Figure S4C,S4D).

Subsequently, we analyzed the expressional levels of human leukocyte antigen (HLA) family genes in the two risk groups. HLA family genes are the most complex and polymorphic genes. consisting of the most concentrated genes related to immune regulation and demonstrating a close correlation with multiple diseases. The results showed that among the MHC class I (MHC I) genes, *HLA-G* had a significantly higher expression in the high-risk group than in the low-risk group (*Figure 9C*). Among MHC II genes, the expression of MHC II, DP alpha 1 (*HLA-DPA1*) and MHC II, DP beta 1 (*HLA-DPB1*) were higher in the high-risk group than in the low-risk group (*Figure 9C*). These findings suggest that the prognostic signature based on the *COL6A6*-MUT can reflect the tumor immunotherapy effect to a certain extent.

Discussion

ECM is the noncellular component of TME stroma that forms a scaffold in the tumor (13,34,35). The alterations in
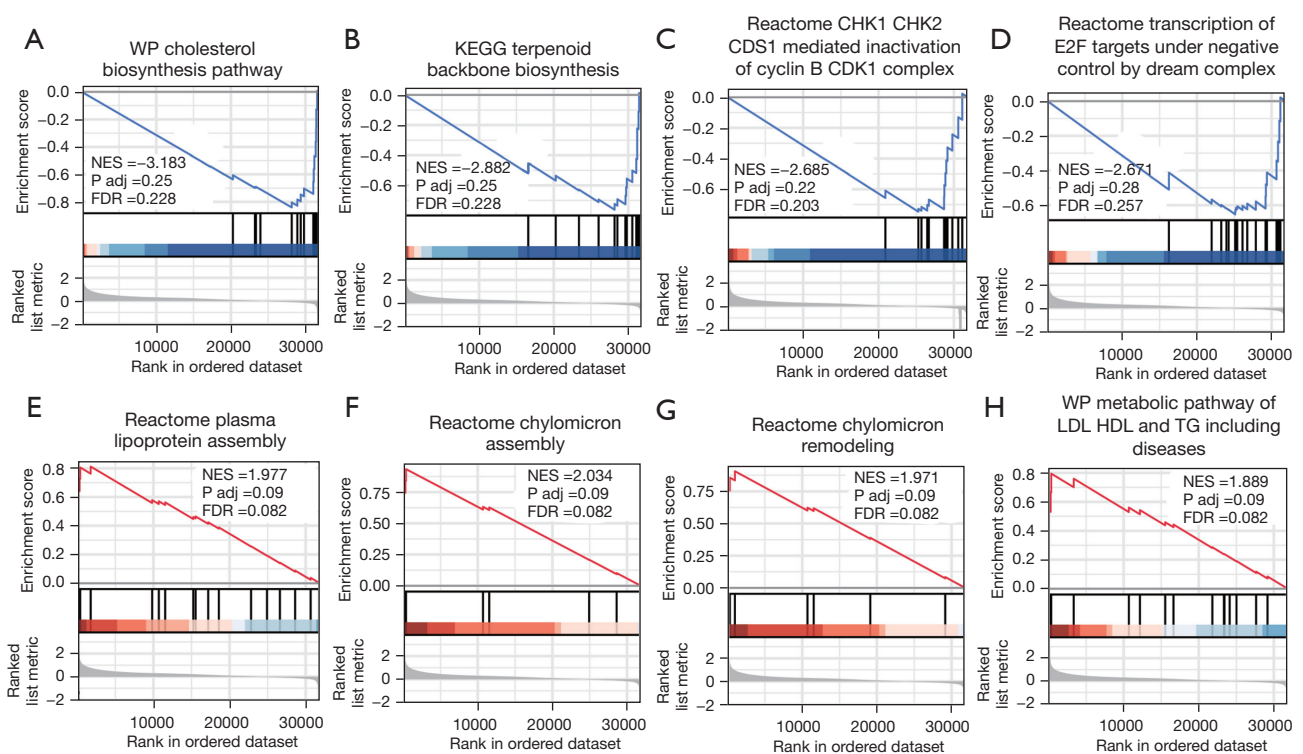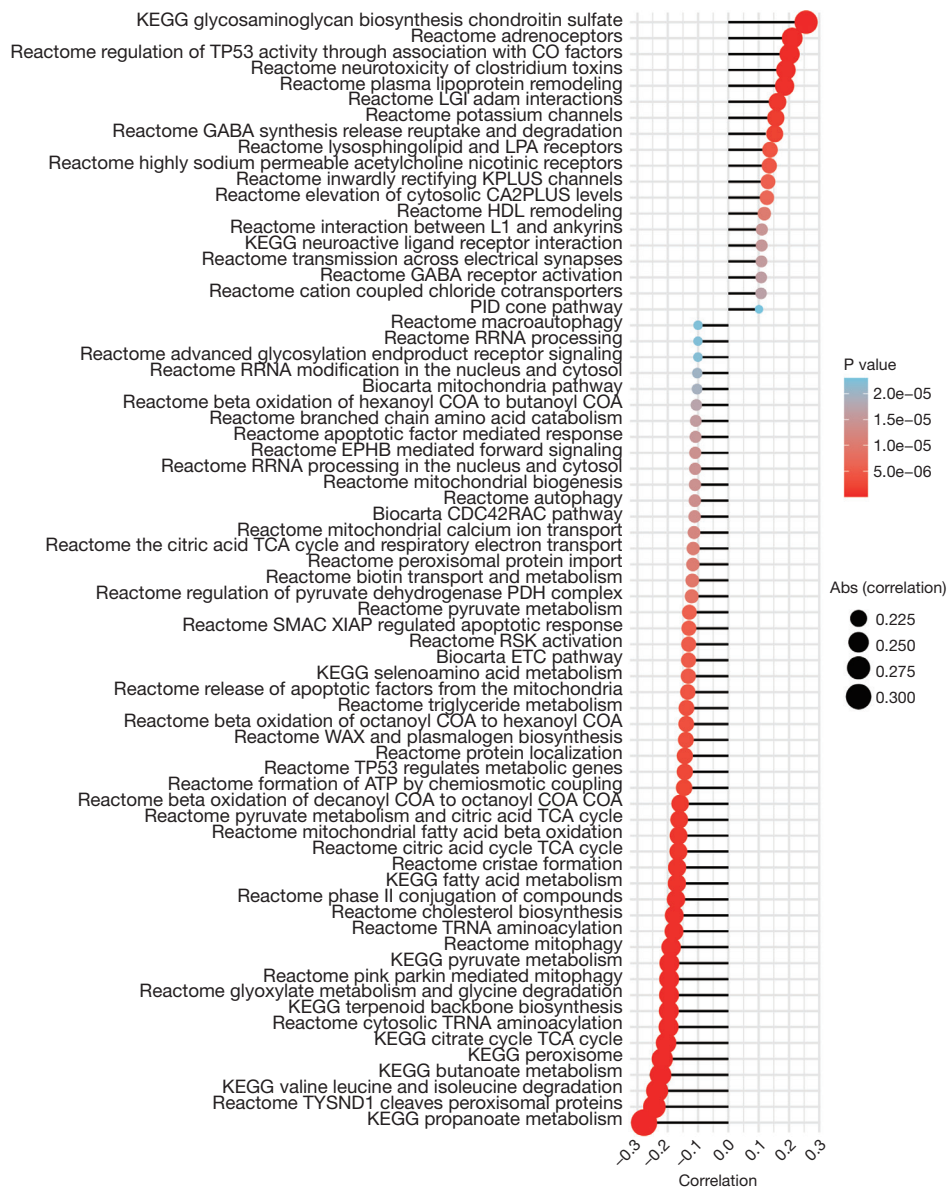
Figure 6 GSEA analysis of the DEGs associated with the different risk groups based on the *COL6A6* allelic state. (A-D) GSEA showing the top four most significantly activated pathways in the high-risk group. (E-H) GSEA showing the top four most significantly activated pathways in the low-risk group. NES, normalized enrichment score; FDR, false discovery rate; KEGG, Kyoto Encyclopedia of Genes and Genomes; GSEA, gene set enrichment analysis; DEGs, differentially expressed genes; *COL6A6*, collagen type VI alpha 6 chain.

both stiffness and the degradation of ECM contribute to the promotion of a malignant cancer phenotype (36). Collagen is the most frequent ECM scaffolding protein within the stroma that is related to ECM stiffness enhancement and tumor malignancy (13,37). The collagen family is classified into different groups based on triple helix domain, length, and molecular weight (15) and includes fibril collagens; fibril-associated collagens; network-forming collagens; collagens VI, VII, XXVI, and XXVIII; membrane collagen; and multiplexins (38). Collagen VI is a major ECM protein composed of six polypeptide chains—a1 (VI), a2 (VI), a3 (VI), a4 (VI), a5 (VI), and a6 (VI)—encoded by different genes—*COL6A1*, *COL6A2*, *COL6A3*, *COL6A4*, *COL6A5*, and *COL6A6*, respectively (39). The major function of collagen VI in ECM is to contribute to the property of the local ECM microenvironment and provide structural support for cells (39). In addition, it is involved in regulating multiple cancer-related signaling pathways, such as apoptosis, autophagy, proliferation, angiogenesis, fibrosis, chemotherapy resistance, and inflammation

(39-44). However, the molecular mechanism underlying how collagen VI functions in tumorigenesis and the oncotherapy has not been clarified.

Principal findings

Our study indicated that the somatic mutation of *COL6A6* has obvious significance for the prognosis in the ICGC-COAD and TCGA-COAD cohorts. We analyzed the somatic mutation and the OS-related genes in both the ICGC-COAD and TCGA-COAD cohorts, and the *COL6A6*-MUT was found to be associated with patient survival, and the change in CNV and SNP patterns of cancer cells. This suggests that *COL6A6*-MUT may influence the fate of the cancer cells by changing the gene expression pattern. Meanwhile, 593 DEGs were found between the *COL6A6*-MUT and *COL6A6*-WT samples and were involved mainly in pathways of metabolism, environmental information processing, cellular processes, and human disease. Collectively, these findings indicate that

Figure 7 GSVA analysis of the DEGs associated with different risks group based on the *COL6A6* allelic state. Correlation analysis between the differential pathways and the risk score. KEGG, Kyoto Encyclopedia of Genes and Genomes; GSVA, gene set variation analysis; DEGs, differentially expressed genes; *COL6A6*, collagen type VI alpha 6 chain.

role of *COL6A6* in COAD is substantial.

After univariate Cox regression, LASSO regression, and multivariate Cox regression analyses were applied to the 593 DEGs, five genes (*MUC16*, *ASNSP1*, *PRR18*, *PEG10*, and *RPL26P8*) were identified as independent prognostic factors. Among these genes, the expression of *MUC16*, *ASNSP1*, *PRR18*, and *PEG10* expression was significantly reduced in patients with the *COL6A6*-MUT compared to those without it, while *RPL26P8* expression

was increased. Meanwhile, the five prognostic significance genes were all positively correlated with the risk score, especially *MUC16*. *MUC16*, a widely known a tumor biomarker and a novel target in cancer therapy (45-47), is overexpressed in multiple malignancies, including ovarian, pancreatic, breast, and lung cancer (46,48-51), and induces an antitumor immune response (47,52-54). In our study, *MUC16* expression was significantly lower in patients with COAD and the *COL6A6*-MUT than it was
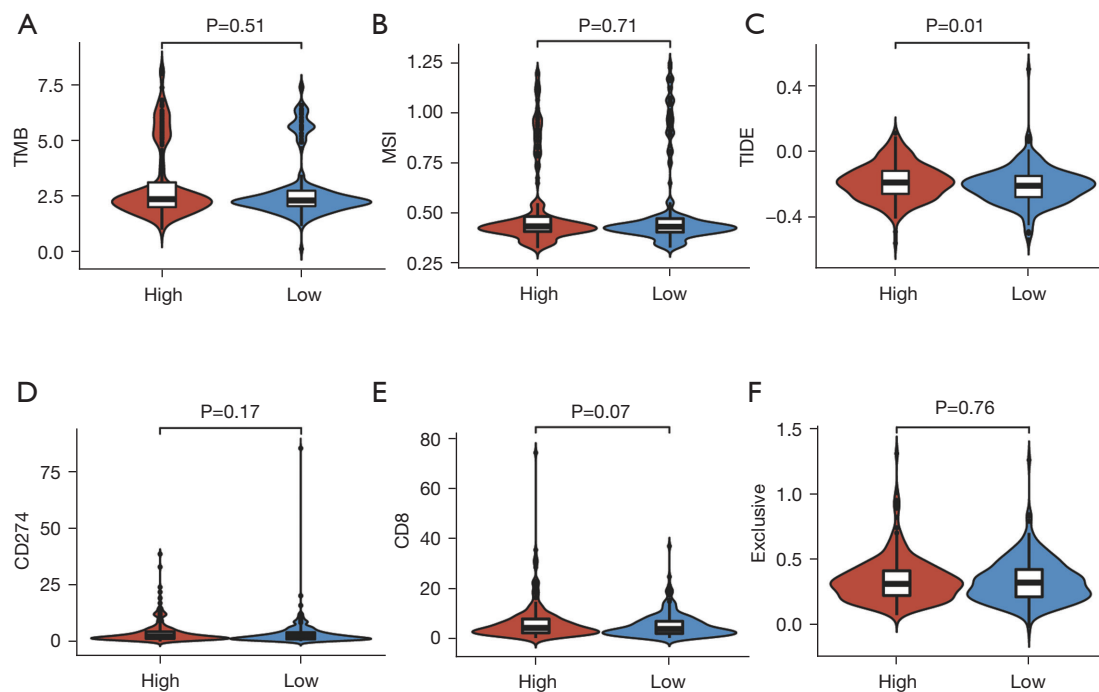
Figure 8 The TMB, MSI, and TIDE scores of the different risk groups. (A) The TMB score was not statistically different between the two risk groups. (B) The MSI score was not statistically different between the two risk groups. (C) The TIDE score was higher in the high-risk group than in the low-risk group. (D) The CD274 level was not statistically different between the two risk groups. (E) The CD8 level was not statistically different between the two risk groups. (F) The exclusion score was not statistically different between the two risk groups. TMB, tumor mutation burden; MSI, microsatellite instability; TIDE, tumor immune dysfunction and exclusion.

in those with *COL6A6*-WT, suggesting that in *COL6A6-MUT* samples, *MUC16* expression was inhibited to some extent. This could be explained by the enhanced alteration of *MUC16* (54%) accompanied by the *COL6A6*-MUT, with *COL6A6* potentially reverse regulating *MUC16* expression. Furthermore, *COL6A6*-MUT was also associated with the reduced expression of another oncogene, *PEG10*. *PEG10* has also been reported to be overexpressed in multiple cancers, such as hepatocellular carcinoma (55), pancreatic carcinoma (56), breast cancer (57,58), prostate cancer (59), gallbladder carcinoma (60), and colon cancer (61). The lower expression of this gene in the *COL6A6*-MUT samples might indicate greater benefit for patient survival.

The role of the immune system in cancer is well established (62-66). Besides conventional therapy, tumor immune infiltrate has been considered and used for several different solid tumor types to achieve a good prognosis, and the abundance of tumor-infiltrating immune cells is an independent predictor of prognosis (62). Previous studies have demonstrated that an increased density of multiple types of immune cells, such as memory CD45RO[+] and CD8[+]

T cells (62,67), cytotoxic lymphocytes (68), NK cells (69,70), FoxP3[+] Tregs (71), and dendritic cells (72-74) can serve as independent predictors of increased OS for patient with CRC. We found that there was no significant differences in immune score, stromal score, and exclusive score between the two risk groups based on the five prognostic genes associated with the *COL6A6*-WT allelic state. However, the TIDE score in the high-risk group was higher than that in the low-risk group, indicating a better outcome of immunotherapy in the low-risk group. Moreover, some of the HLA family genes (*HLA-G* from MHC I, *HLA-DPA1* and *HLA-DPB1* from MHC II) were more highly expressed in the high-risk group than in the low-risk group. MHC I and MHC II genes are known for their antitumor and proimmune evasion function and have been used as targets in cancer immunotherapy (75-79). The higher expression of MHC I and MHC II genes in the high-risk group predicted a poorer outcome for patients. Notably, our results also showed that the high-risk group had a higher proportion of plasma cell and Treg infiltration than did the low-risk group but a lower proportion of resting memory CD4 T cell and
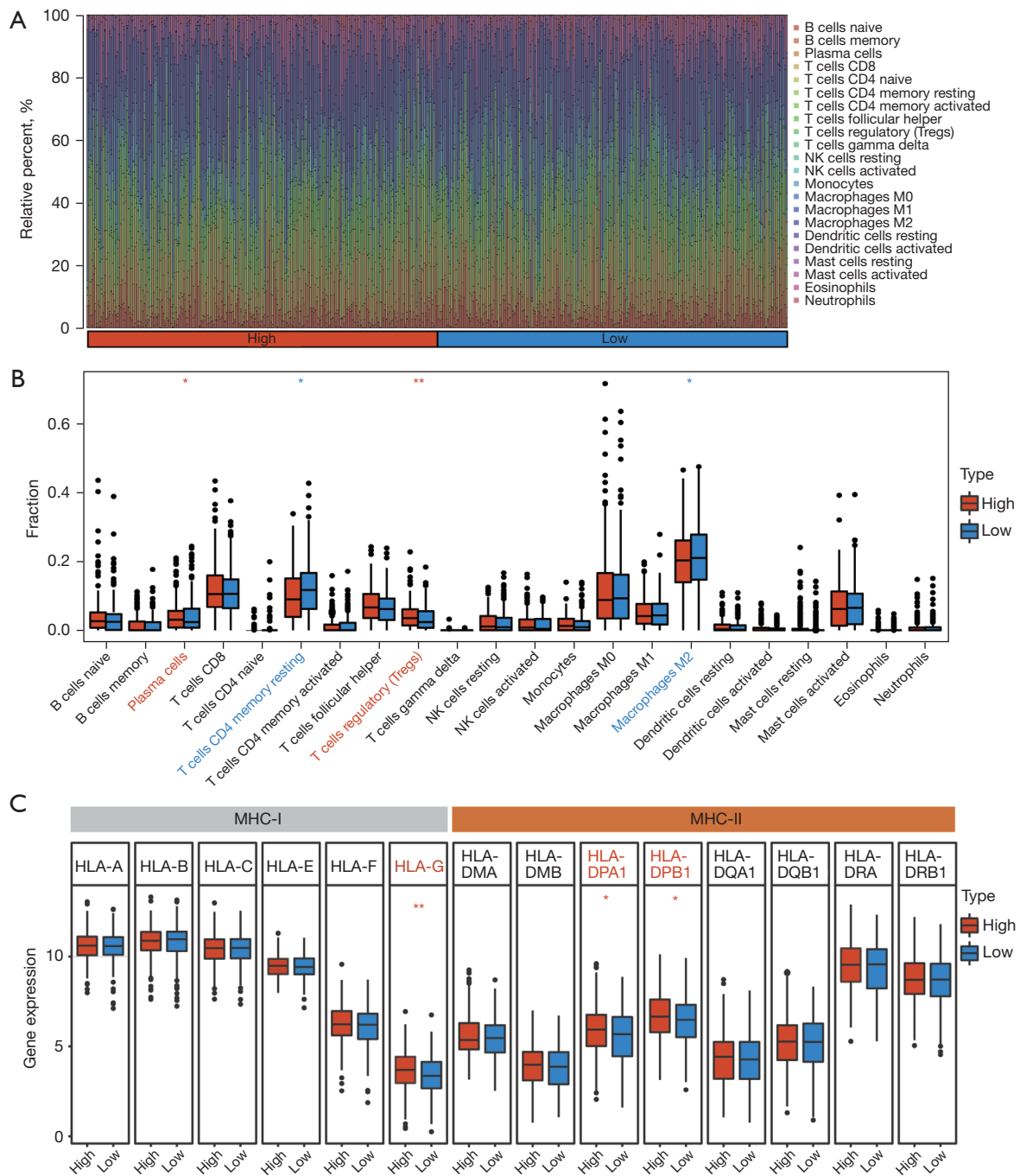
Figure 9 The relationship between the TME and the prognostic signature based on the *COL6A6* allelic state. (A) The landscape of immune cell infiltration in the TME of the high- and low-risk groups in TCGA-COAD cohort. (B) Comparison of the immune cell infiltration in the TME of the two risk groups. (C) Comparison of HLA family genes in the two risk groups. *, P<0.05; **, P<0.01. NK, natural killer; MHC, major histocompatibility complex; HLA, human leukocyte antigen; TME, tumor microenvironment; *COL6A6*, collagen type VI alpha 6 chain; TCGA, The Cancer Genome Atlas; COAD, colon adenocarcinoma.

M2 macrophage infiltration. Tregs have been reported to be critical for preventing autoimmunity and suppressing effective tumor immunity (80). These demonstrate that the five-gene signature that was correlated with the *COL6A6* allelic state, can serve as an independent prognostic indicator for tumor immunotherapy.

Finally, we identified 2,921 DEGs between the high- and low-risk groups performed according to the five independent prognostic genes based on the *COL6A6* allelic state. Subsequently, KEGG analysis indicated that these DEGs affected a variety of cellular biological processes, including environmental information processing, organismal systems, and human disease, which is consistent with the pathways that the *COL6A6*-MUT-induced DEGs were involved in. Ensuing GSEA and GSVA also indicated that *COL6A6*-MUT affected multiple biological processes in cancer cells, including intracellular metabolic regulation, mitophagy, ATM pathway, and MHC-mediated antigen-processing presentation, among others. The ATM pathway has been reported to be closely related to cancer immunotherapy (81).

Strengths and limitations

This study examined the prognostic factors of COAD from the perspective of gene mutation. The possible mechanism underlying tumor immune microenvironment remodeling was investigated at the genomic level. This may provide a novel platform for clarifying the immune invasion mechanism of COAD. However, there are some notable limitations in this study. First, the large number of datasets might have resulted in batch differences that could not be avoided or removed during analysis. Moreover, despite the fact that the bioinformatic analysis based on high-throughput sequencing enabled us to efficiently deduce the related molecular mechanisms, further experimental validations should be completed to verify and illuminate the specific regulatory mechanism of *COL6A6*. Finally, correlated clinical research was lacking for this study and thus could not be included in the study's analysis. Subsequent research should collect the relevant clinical tissues and corresponding information to further clarify the molecular role of *COL6A6* at the protein level.

Comparison with other research

Much has been reported about the effects of genetic mutations in COAD, in particular, genes involved in DNA

damage response. Our study analyzed gene mutations from a genomic perspective using big data mining, which identified additional genes relevant to the prognosis of COAD. It is hoped these findings can contribute to further insights into the pathogenesis and treatment resistance of COAD.

Interpretation of findings

The DEGs identified in our analysis can be used to further characterize the underlying mechanisms of COAD tumor occurrence and progression. These genes show significant differences in expression between different mutant groups, and changes in their expression levels suggest that they may be involved in the disease process. We performed functional enrichment analyses of these DEGs to gain insight into their biological significance. These analyses included pathway enrichment and GO analysis, which allowed us to identify the key biological processes, molecular functions, and cellular components associated with DEGs. In addition, many of these DEGs are known to participate in pathways related to processes such as lipid metabolism, cell cycle regulation, and immune response. Alterations in these pathways are associated with the onset and progression of cancer. Upregulation or down-regulation of specific genes in these pathways may contribute to various aspects of COAD pathogenesis, such as cell proliferation, invasion, and immune evasion. It is important to note that the exact contribution of individual DEGs to the progress of COAD may vary, and a full understanding requires further experimental validation. However, our analysis provides a basis for identifying potential molecular factors that can be targeted in further COAD research and therapeutic interventions.

In this analysis, we accounted for the potential confounding factors that could have affected our results. We carefully considered patient age, tumor stage, and treatment options and performed a stratified analysis to assess how these factors affected the results. Specifically, we performed analyses across different age groups and tumor stages to examine the consistency of our findings across these subgroups. In addition, in the multivariate Cox regression analysis from which the predictive model was constructed, we included these factors as covariates to adjust for their potential impact on the risk assessment. This approach allowed us to better understand the independent contribution of the *COL6A6* allelic state while controlling for the influence of patient age, tumor stage, and treatment

regimen. To gain a comprehensive understanding, we also performed a subgroup analysis based on these confounding factors. For example, we evaluated the relationship between risk scores and patient outcomes for different age groups and tumor stages, which permitted us to determine how risk assessments might vary in the different patient subgroups. Overall, we adopted measures to address potential confounding factors to ensure the robustness and reliability of our analysis.

The five prognostic genes may not function individually but rather collectively to influence COAD prognosis as components of a complex molecular network. Although the exact mechanism of their interaction requires further study, we can suggest some potential effects. (I) In pathway crosstalk, the identified genes may be involved in common pathways or networks associated with cancer progression. They may participate in shared signaling cascades, transcriptional regulation, or metabolic pathways that have synergistic or antagonistic effects on tumor development and therapeutic response. (II) Concerning photosynthesis, some of these genes may cooperate to regulate specific cellular processes. For example, they may work in combination to regulate cell cycle progression, apoptosis, immune escape, or angiogenesis, which are critical for cancer progression. (III) Within the context of immune response regulation, these genes may play a role in shaping the TME and immune response. The interaction between these genes may affect the recruitment and activity of immune cells within the tumor, thereby affecting overall immune surveillance and response to therapy. (IV) Regarding therapeutic potential, understanding the interactions between these genes could have therapeutic implications. Targeting multiple genes in a network may be more effective at altering disease trajectories than targeting a single gene. (V) Finally, as it relates to biomarker characteristics, combinations of these genes may form prognostic or predictive biomarker signatures that can provide a more comprehensive COAD prognosis than can individual genes used alone.

Regarding the higher expression of HLA family genes, particularly MHC I and II, in high-risk populations, there are several potential effects these may have on COAD. (I) For immune recognition and tumor surveillance, increased expression of MHC I and II molecules may lead to enhanced antigen presentation to immune cells, such as T cells. This may enhance immune recognition of tumor-specific antigens, thereby facilitating immune surveillance and targeted destruction of cancer cells. (II) Regarding

tumor antigen presentation, MHC I molecules present intracellularly derived antigens to cytotoxic T lymphocytes, eliminating cells that present abnormal or tumor-associated antigens. Higher MHC I expression in high-risk populations may indicate an active immune response against tumor antigens, which may imply a more aggressive antitumor immune response. (III) Concerning tumor-infiltrating immune cells, elevated MHC expression may attract and stimulate immune cell infiltration into the TME. Although the specific immune cell composition needs to be further studied, an increase in immune cell types, such as CD8+ T cells, may be associated with a more powerful antitumor response. (IV) Related to the potential of immunotherapy, high MHC expression may indicate that tumors in high-risk populations are more immunogenic, potentially making them more sensitive to immunotherapy approaches such as immune checkpoint inhibitor or adoptive T-cell therapy. (V) In terms of immune escape mechanism, elevated MHC expression may conversely also reflect an attempt by tumor cells to evade immune surveillance by overexpressing antigens as a form of antigen masking. Tumor cells may "exhaust" immune responses by presenting antigens that lead to immune tolerance or dysfunction. (VI) Finally, concerning tumor heterogeneity, variability in MHC expression may reflect tumor heterogeneity, with some subclones within the high-risk group having a more active immune response and others possessing immune escape mechanisms. It is important to note that the effect of higher MHC expression in high-risk populations is complex and may be influenced by a number of factors, including the specific TME, interactions between tumor cells and immune cells, and the overall immune response. Further experimental and clinical studies are needed to confirm these effects and determine the functional consequences of MHC expression in the progression of COAD.

Implications and future directions

Screening for genetic mutations should be expanded, and additional genetic mutations associated with tumorigenesis and development should be analyzed in depth.

Conclusions

This paper systematically analyzed the genomic mutation landscape in COAD. The prognosis of all mutated genes was evaluated as a whole. Specifically, the *COL6A6* allelic state was found to be associated with the OS of patients

with COAD, which could be mainly attributed to the influence of the five genes, which further alter the immune microenvironment of COAD. Overall, these findings indicate the role of *COL6A6* in COAD, which may be used to identify those patients most suitable for immunotherapy.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://tcr.amegroups.com/article/view/10.21037/tcr-23-463/rc

Peer Review File: Available at https://tcr.amegroups.com/article/view/10.21037/tcr-23-463/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tcr.amegroups.com/article/view/10.21037/tcr-23-463/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study conformed to the provisions of the Declaration of Helsinki (as revised in 2013).

References

1. Labianca R, Beretta GD, Kildani B, et al. Colon cancer. Crit Rev Oncol Hematol 2010;74:106-33.
2. Brody H. Colorectal cancer. Nature 2015;521:S1.
3. Mody K, Baldeo C, Bekaii-Saab T. Antiangiogenic Therapy in Colorectal Cancer. Cancer J 2018;24:165-70.
4. Miyamoto Y, Oki E, Saeki H, et al. Recent Advances in Systemic Chemotherapy for Metastatic Colorectal Cancer. Gan To Kagaku Ryoho 2016;43:15-23.
5. Piawah S, Venook AP. Targeted therapy for colorectal cancer metastases: A review of current methods of molecularly targeted therapy and the use of tumor biomarkers in the treatment of metastatic colorectal cancer. Cancer 2019;125:4139-47.
6. Erratum. Arq Gastroenterol 2020;57:341.
7. Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. Science 2015;349:1483-9.
8. Necula L, Matei L, Dragu D, et al. Collagen Family as Promising Biomarkers and Therapeutic Targets in Cancer. Int J Mol Sci 2022;23:12415.
9. Parola M, Pinzani M. Liver fibrosis: Pathophysiology, pathogenetic targets and clinical issues. Mol Aspects Med 2019;65:37-55.
10. Goldfracht I, Efraim Y, Shinnawi R, et al. Engineered heart tissue models from hiPSC-derived cardiomyocytes and cardiac ECM for disease modeling and drug testing applications. Acta Biomater 2019;92:145-59.
11. Frangogiannis NG. The Extracellular Matrix in Ischemic and Nonischemic Heart Failure. Circ Res 2019;125:117-46.
12. Naba A, Pearce OMT, Del Rosario A, et al. Characterization of the Extracellular Matrix of Normal and Diseased Tissues Using Proteomics. J Proteome Res 2017;16:3083-91.
13. Najafi M, Farhood B, Mortezaee K. Extracellular matrix (ECM) stiffness and degradation as cancer drivers. J Cell Biochem 2019;120:2782-90.
14. Mohan V, Das A, Sagi I. Emerging roles of ECM remodeling processes in cancer. Semin Cancer Biol 2020;62:192-200.
15. Haq F, Ahmed N, Qasim M. Comparative genomic analysis of collagen gene diversity. 3 Biotech 2019;9:83.
16. Makoukji J, Makhoul NJ, Khalil M, et al. Gene expression profiling of breast cancer in Lebanese women. Sci Rep 2016;6:36639.
17. Ma Y, Qiu M, Guo H, et al. Comprehensive Analysis of the Immune and Prognostic Implication of COL6A6 in Lung Adenocarcinoma. Front Oncol 2021;11:633420.
18. Mayakonda A, Koeffler HP. Maftools: Efficient analysis, visualization and summarization of MAF files from large-scale cohort based cancer studies. BioRxiv 2016. doi: 10.1101/052662.
19. Colaprico A, Silva TC, Olsen C, et al. TCGAbiolinks: an

R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res 2016;44:e71.

20. Mermel CH, Schumacher SE, Hill B, et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol 2011;12:R41.

21. Coletta A, Molter C, Duqué R, et al. InSilico DB genomic datasets hub: an efficient starting point for analyzing genome-wide studies in GenePattern, Integrative Genomics Viewer, and R/Bioconductor. Genome Biol 2012;13:R104.

22. Merino DM, McShane LM, Fabrizio D, et al. Establishing guidelines to harmonize tumor mutational burden (TMB): in silico assessment of variation in TMB quantification across diagnostic platforms: phase I of the Friends of Cancer Research TMB Harmonization Project. J Immunother Cancer 2020;8:e000147.

23. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.

24. Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012;16:284-7.

25. Gene Ontology Consortium; Aleksander SA, Balhoff J, et al. The Gene Ontology knowledgebase in 2023. Genetics 2023;224:iyad031.

26. Kanehisa M, Furumichi M, Sato Y, et al. KEGG: integrating viruses and cellular organisms. Nucleic Acids Res 2021;49:D545-51.

27. Canzler S, Hackermüller J. multiGSEA: a GSEA-based pathway enrichment analysis for multi-omics data. BMC Bioinformatics 2020;21:561.

28. Ferreira MR, Santos GA, Biagi CA, et al. GSVA score reveals molecular signatures from transcriptomes for biomaterials comparison. J Biomed Mater Res A 2021;109:1004-14.

29. Beyene KM, El Ghouch A. Time-dependent ROC curve estimation for interval-censored data. Biom J 2022;64:1056-74.

30. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015;12:453-7.

31. Scire J, Huisman JS, Grosu A, et al. estimateR: an R package to estimate and monitor the effective reproductive number. BMC Bioinformatics 2023;24:310.

32. Fu J, Li K, Zhang W, et al. Large-scale public data reuse to model immunotherapy response and resistance. Genome Med 2020;12:21.

33. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013;310:2191-4.

34. Fu R, Zhang YW, Li HM, et al. LW106, a novel indoleamine 2,3-dioxygenase 1 inhibitor, suppresses tumour progression by limiting stroma-immune crosstalk and cancer stem cell enrichment in tumour micro-environment. Br J Pharmacol 2018;175:3034-49.

35. Yamaguchi R, Perkins G. Animal models for studying tumor microenvironment (TME) and resistance to lymphocytic infiltration. Cancer Biol Ther 2018;19:745-54.

36. Chen J, Luo Y, Zhou Y, et al. Promotion of Tumor Growth by ADAMTS4 in Colorectal Cancer: Focused on Macrophages. Cell Physiol Biochem 2018;46:1693-703.

37. Patrawalla NY, Kajave NS, Albanna MZ, et al. Collagen and Beyond: A Comprehensive Comparison of Human ECM Properties Derived from Various Tissue Sources for Regenerative Medicine Applications. J Funct Biomater 2023;14:363.

38. Devos H, Zoidakis J, Roubelakis MG, et al. Reviewing the Regulators of COL1A1. Int J Mol Sci 2023;24:10004.

39. Li X, Li Z, Gu S, et al. A pan-cancer analysis of collagen VI family on prognosis, tumor microenvironment, and its potential therapeutic effect. BMC Bioinformatics 2022;23:390.

40. Cescon M, Rampazzo E, Bresolin S, et al. Collagen VI sustains cell stemness and chemotherapy resistance in glioblastoma. Cell Mol Life Sci 2023;80:233.

41. Abbonante V, Malara A, Chrisam M, et al. Lack of COL6/collagen VI causes megakaryocyte dysfunction by impairing autophagy and inducing apoptosis. Autophagy 2023;19:984-99.

42. Hong X, Zhang J, Zou J, et al. Role of COL6A2 in malignant progression and temozolomide resistance of glioma. Cell Signal 2023;102:110560.

43. Zhang X, Liu J, Yang X, et al. High expression of COL6A1 predicts poor prognosis and response to immunotherapy in bladder cancer. Cell Cycle 2023;22:610-8.

44. Zhang Y, Liu Z, Yang X, et al. H3K27 acetylation activated-COL6A1 promotes osteosarcoma lung metastasis by repressing STAT1 and activating pulmonary cancer-associated fibroblasts. Theranostics 2021;11:1473-92.

45. Felder M, Kapur A, Gonzalez-Bosquet J, et al. MUC16 (CA125): tumor biomarker to cancer therapy, a work in progress. Mol Cancer 2014;13:129.

46. Aithal A, Rauth S, Kshirsagar P, et al. MUC16 as a novel target for cancer therapy. Expert Opin Ther Targets 2018;22:675-86.

47. Zhai Y, Lu Q, Lou T, et al. MUC16 affects the biological functions of ovarian cancer cells and induces an antitumor immune response by activating dendritic cells. Ann Transl Med 2020;8:1494.

48. Olson MT, Aguilar EN, Brooks CL, et al. Preclinical Evaluation of a Humanized, Near-Infrared Fluorescent Antibody for Fluorescence-Guided Surgery of MUC16-Expressing Pancreatic Cancer. Mol Pharm 2022;19:3586-99.

49. Chaudhary S, Appadurai MI, Maurya SK, et al. MUC16 promotes triple-negative breast cancer lung metastasis by modulating RNA-binding protein ELAVL1/HUR. Breast Cancer Res 2023;25:25.

50. Saad HM, Tourky GF, Al-Kuraishy HM, et al. The Potential Role of MUC16 (CA125) Biomarker in Lung Cancer: A Magic Biomarker but with Adversity. Diagnostics (Basel) 2022;12:2985.

51. Lakshmanan I, Salfity S, Seshacharyulu P, et al. MUC16 Regulates TSPYL5 for Lung Cancer Cell Growth and Chemoresistance by Suppressing p53. Clin Cancer Res 2017;23:3906-17.

52. Felder M, Kapur A, Rakhmilevich AL, et al. MUC16 suppresses human and murine innate immune responses. Gynecol Oncol 2019;152:618-28.

53. Menon BB, Kaiser-Marko C, Spurr-Michaud S, et al. Suppression of Toll-like receptor-mediated innate immune responses at the ocular surface by the membrane-associated mucins MUC1 and MUC16. Mucosal Immunol 2015;8:1000-8.

54. Zhang L, Han X, Shi Y. Association of MUC16 Mutation With Response to Immune Checkpoint Inhibitors in Solid Tumors. JAMA Netw Open 2020;3:e2013201.

55. Bang H, Ha SY, Hwang SH, et al. Expression of PEG10 Is Associated with Poor Survival and Tumor Recurrence in Hepatocellular Carcinoma. Cancer Res Treat 2015;47:844-52.

56. Peng YP, Zhu Y, Yin LD, et al. PEG10 overexpression induced by E2F-1 promotes cell proliferation, migration, and invasion in pancreatic cancer. J Exp Clin Cancer Res 2017;36:30.

57. Li X, Xiao R, Tembo K, et al. PEG10 promotes human breast cancer cell proliferation, migration and invasion. Int J Oncol 2016;48:1933-42.

58. Deng X, Hu Y, Ding Q, et al. PEG10 plays a crucial role in human lung cancer proliferation, progression, prognosis and metastasis. Oncol Rep 2014;32:2159-67.

59. Shapovalova M, Lee JK, Li Y, et al. PEG10 Promoter-Driven Expression of Reporter Genes Enables Molecular Imaging of Lethal Prostate Cancer. Cancer Res 2019;79:5668-80.

60. Liu Z, Yang Z, Liu D, et al. TSG101 and PEG10 are prognostic markers in squamous cell/adenosquamous carcinomas and adenocarcinoma of the gallbladder. Oncol Lett 2014;7:1128-38.

61. Li B, Shi C, Li B, et al. The effects of Curcumin on HCT-116 cells proliferation and apoptosis via the miR-491/PEG10 pathway. J Cell Biochem 2018;119:3091-8.

62. Christofides A, Strauss L, Yeo A, et al. The complex role of tumor-infiltrating macrophages. Nat Immunol 2022;23:1148-56.

63. Fujii SI, Shimizu K. Immune Networks and Therapeutic Targeting of iNKT Cells in Cancer. Trends Immunol 2019;40:984-97.

64. Sabado RL, Balan S, Bhardwaj N. Dendritic cell-based immunotherapy. Cell Res 2017;27:74-95.

65. Sabry M, Lowdell MW. Killers at the crossroads: The use of innate immune cells in adoptive cellular therapy of cancer. Stem Cells Transl Med 2020;9:974-84.

66. Dumauthioz N, Labiano S, Romero P. Tumor Resident Memory T Cells: New Players in Immune Surveillance and Therapy. Front Immunol 2018;9:2076.

67. Zhang L, Yu X, Zheng L, et al. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. Nature 2018;564:268-72.

68. Fionda C, Scarno G, Stabile H, et al. NK Cells and Other Cytotoxic Innate Lymphocytes in Colorectal Cancer Progression and Metastasis. Int J Mol Sci 2022;23:7859.

69. Sorrentino C, D'Antonio L, Fieni C, et al. Colorectal Cancer-Associated Immune Exhaustion Involves T and B Lymphocytes and Conventional NK Cells and Correlates With a Shorter Overall Survival. Front Immunol 2021;12:778329.

70. Courau T, Bonnereau J, Chicoteau J, et al. Cocultures of human colorectal tumor spheroids with immune cells reveal the therapeutic potential of MICA/B and NKG2A targeting for cancer treatment. J Immunother Cancer 2019;7:74.

71. Saito T, Nishikawa H, Wada H, et al. Two FOXP3(+) CD4(+) T cell subpopulations distinctly control the prognosis of colorectal cancers. Nat Med 2016;22:679-84.

72. Li E, Yang X, Du Y, et al. CXCL8 Associated Dendritic Cell Activation Marker Expression and Recruitment as Indicators of Favorable Outcomes in Colorectal Cancer. Front Immunol 2021;12:667177.

73. Mosińska P, Gabryelska A, Zasada M, et al. Dual Functional Capability of Dendritic Cells - Cytokine-Induced Killer Cells in Improving Side Effects of

Colorectal Cancer Therapy. Front Pharmacol 2017;8:126.

74. Chistiakov DA, Orekhov AN, Bobryshev YV. Dendritic Cells in Colorectal Cancer and a Potential for their Use in Therapeutic Approaches. Curr Pharm Des 2016;22:2431-8.

75. Burr ML, Sparbier CE, Chan KL, et al. An Evolutionarily Conserved Function of Polycomb Silences the MHC Class I Antigen Presentation Pathway and Enables Immune Evasion in Cancer. Cancer Cell 2019;36:385-401.e8.

76. Cornel AM, Mimpen IL, Nierkens S. MHC Class I Downregulation in Cancer: Underlying Mechanisms and Potential Targets for Cancer Immunotherapy. Cancers (Basel) 2020;12:1760.

77. Shukla A, Cloutier M, Appiya Santharam M, et al. The

MHC Class-I Transactivator NLRC5: Implications to Cancer Immunology and Potential Applications to Cancer Immunotherapy. Int J Mol Sci 2021;22:1964.

78. Axelrod ML, Cook RS, Johnson DB, et al. Biological Consequences of MHC-II Expression by Tumor Cells in Cancer. Clin Cancer Res 2019;25:2392-402.

79. Goodman AM, Castro A, Pyke RM, et al. MHC-I genotype and tumor mutational burden predict response to immunotherapy. Genome Med 2020;12:45.

80. Tanaka A, Sakaguchi S. Targeting Treg cells in cancer immunotherapy. Eur J Immunol 2019;49:1140-6.

81. Hu M, Zhou M, Bao X, et al. ATM inhibition enhances cancer immunotherapy by promoting mtDNA leakage and cGAS/STING activation. J Clin Invest 2021;131:e139333.