Exploring Cancer Dependency Map genes and immune subtypes in colon cancer, in which *TIGD1* contributes to colon cancer progression

Guoyang Zhang[1,2,*], Zongfeng Feng[1,2,*], Qingwen Zeng[1,2,*], Ping Huang[3]

[1]Department of General Surgery, First Affiliated Hospital of Nanchang University, Nanchang, China
[2]Medical Innovation Center, The First Affiliated Hospital of Nanchang University, Nanchang, China
[3]Department of Nutrition, The First Affiliated Hospital of Nanchang University, Nanchang, China
*Equal contribution

Correspondence to: Ping Huang; **email:** hp17872@126.com, https://orcid.org/0000-0001-5348-9979

ABSTRACT

Background: Tumour-dependent genes identified in CRISPR–Cas9 screens have been widely reported in Cancer Dependency Maps (CDMs). CDM-derived tumour-dependent genes play an important role in tumorigenesis and progression; however, they have not been investigated in colon cancer (CC).
Methods: CDM genes overexpressed in CC were identified from the TCGA-COAD dataset and CDM platform. A CDM signature and prognostic nomogram were constructed by Lasso Cox regression and multivariate Cox analyses. A weighted correlation network analysis (WGCNA) and consensus clustering were used to define coexpressed genes with CDM risk scores and to determine two new immune subtypes. A comprehensive investigation was performed between the two subtypes and immune regulation, the immune microenvironment and the impact of immunotherapy.
Results: First, 1304 overexpressed CDM genes were identified. Then, a CDM signature with five cancer-dependent genes (*MMS19, NOP14, POLRMT, SNAPC5* and *TIGD1*) and a prognostic nomogram were constructed, and they demonstrated robust predictive performance and a close relationship with clinical characteristics in different CC datasets. Patients with high CDM risk scores showed worse survival outcome and weaker response to chemotherapy. Additionally, *TIGD1* genes were oncogenes that affected the CC cell cycle, according to cell functional experiments that involved the suppression of the *TIGD1* gene. Furthermore, WGCNA and consensus clustering were used to define coexpressed genes with CDM risk scores and to determine two new immune subtypes. Finally, systematic investigations were conducted with the relationship between the CDM subtypes and immune regulation.
Conclusions: This study constructed a CDM signature consisting of five risk genes that predict survival in CC patients. In addition, the immune subtypes provided valuable insights into immunotherapy for CC patients. *TIGD1*, as an oncogene, is independent prognostic factors for CC, and contributes to CC progression.

INTRODUCTION

Recent studies have suggested that tumorigenesis is closely related to the alteration of genetic material and the deterioration of the environment of the organism [1, 2]. Moreover, environmental degradation can cause somatic cell biological dysfunction, such as impaired DNA damage repair [3], abnormal activation or inhibition of signalling pathways [4], and dysregulation of protein production and degradation [5], which can lead to the deterioration of normal epithelial cells into tumour cells. The most common gastrointestinal cancer is colon

cancer (CC), and the number of new cases is increasing every year [6]. However, the pathogenesis of CC has yet to be fully elucidated. Normal intestinal mucosal epithelium-proliferating microadenoma-early adenoma-intermediate adenoma-advanced adenoma-cancer-cancer metastasis, seems to be the growth history chain of most CC. How to interrupt this growth chain and identify oncogenes associated with tumour growth have become the keys to preventing and treating CC.

The Cancer Dependency Map (CDM) project [7], initiated by scientists at the Broad Institute and Dana-Farber Cancer Institute, aims to use RNA interference (RNAi) and CRISPR–Cas9, two gene-silencing technologies, to uncover genes that drive cancer cell proliferation. The CDM project's dataset is updated quarterly, and data published in 2017 revealed that the screening of hundreds of cancer cell lines was completed and that 769 strongly associated tumour-dependent genes were identified [8]. Data released in April 2022 show that the project has screened 1,393 cell lines for 33 cancers, and identified 19,177 genes. Due to its excellent non-off-target effects, CRISPR–Cas9 has been employed extensively in gene editing in recent years [9]. Therefore, the basis of our study was built on CRISPR–Cas9 technology and the CDM project.

Tumour cells of different origins, and even cells of the same tumour, have different genomic mechanisms, which determines the diversity and complexity of tumour-dependent genes. This poses a great challenge to screening for tumour-dependent genes and exploring the relationship between genes. Methyl-methanesulfonate sensitivity 19 (*MMS19*) has been found in a variety of tumours [10–12] and is involved in DNA metabolism and genomic stability through the regulation of nucleotide excision repair and Fe-S protein expression, thereby affecting tumor progression [13]. Recently, studies regarding the influence of nucleolar complex protein 14 (*NOP14*) on cancers have been conducted, and this protein has been shown to play an important role in cancer migration and invasion [14, 15]. In addition, Yin et al. [16] discovered that Trigger transposable element-derived 1 (*TIGD1*) is coupled with a malignancy survival and associated with regulating cell-cycle progression in human cancer. Mitochondrial RNA polymerase (*POLRMT*) is often reported as a target protein for the treatment of malignant tumours [17], however snRNA-activating protein complex 5 (*SNAPC5*), also named *SNAP19* [18], has rarely been reported in tumours.

We constructed a valid and robust CDM signature and predictive nomogram by organically combining RNA sequencing data from The Cancer Genome Atlas–colon adenocarcinoma (TCGA-COAD) tumour samples with data on the tumour-dependent gene interference efficiency of CC cells in the CDM; the signature and nomogram were validated with both internal and external data. Furthermore, two novel immune subtypes were established by weighted correlation network analysis (WGCNA) and consensus clustering, that can be used to precisely determine whether patients will benefit from immunotherapy. Finally, functional experiments revealed thatTIGD1 genes was oncogenes that impacted the CC cell cycle.

MATERIALS AND METHODS

Identifying CDM genes from COAD cells

Interference efficiency files identified by CRISPR–Cas9 for 55 CC cell lines with dependent genes were downloaded from the CDM (https://depmap.org/portal/) [19]. As of April 2022, the CDM has validated 19,177 genes covering 1,393 cell lines from 33 malignancies, making it the most comprehensive research database of tumour cells available. A negative value of interference efficiency indicated that knockdown of a gene inhibited the growth of tumour cells, while a positive value promoted cell growth. When the interference efficiency of a gene was less than -0.5, knockdown of the gene was considered to significantly inhibit the growth of tumour cells, and it was referred to as a cancer dependent gene. A total of 1,613 genes were considered CC dependent genes, as their interference efficiency averaged less than -0.5 in 55 types of CC cells.

Downloading and preprocessing of TCGA and GEO data

TCGA-COAD RNA-Seq and clinical information data, including 452 samples, were downloaded from TCGA (https://portal.gdc.cancer.gov) in March 2022. The format of RNA-Seq expression data was converted from FPKM to TPM, which is more in consistent with the GEO (Gene Expression Omnibus) (https://www.ncbi.nlm.nih.gov/geo/) chip data and provides more convenience for joint analysis of TCGA and GEO data. The GSE17536 microarray dataset contained 177 CC sample RNA-Seq expression and clinical data (Supplementary Table 1). Subsequently, CC datasets from different data sources were log2 (value+1) normalized using the "limma" and "sva" bioconductor packages in the 4.10 R software.

Intersecting genes among the upregulated and CDM genes

The "limma" package was used to compare the differences in gene expression between normal tissue and tumour tissue for the corrected TCGA-COAD data and a p value less than 0.05 and log2-fold change

(logFC) greater than 1 indicated upregulated COAD genes. The "VennDiagram" package was used to identify 1,034 intersecting genes among the upregulated and CDM genes.

Functional enrichment analysis of intersecting genes

Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) analyses were performed using the "clusterProfiler", "org.Hs.eg.db", "enrichplot" and "ggplot2" packages.

Constructing and validating the CDM signature and nomogram

First, univariate Cox hazard analysis was used to screen for prognostic genes via the "limma" and "survival" packages. Then, the TCGA-COAD data were randomly separated into a training group and a test group at a 1:1 ratio. Next, Lasso Cox regression analysis was used to reduce the effect of multicollinearity among multiple genetic variables. Multivariate Cox analysis was performed to construct the CDM risk model based on the genes identified in the Lasso analysis. The validity of the CDM signature containing the TCGA-COAD training cohort, TCGA-COAD training cohort, TCGA-COAD all cohort and GSE17536 cohort was evaluated using a time-dependent receiver operating characteristic (ROC) curve with the R package "timeROC". Based on the signature, the risk score of each sample in the TCGA-COAD and GSE17536 cohorts was determined, and risk curves and patient survival status corresponding to each sample were plotted along with the expression heatmap of the five genes in the signature. Subsequently, the Kaplan–Meier plot was used to investigate the overall survival heterogeneity between the high- and low- risk groups. Finally, a nomogram evaluating the survival time of CC patients was created and validated by R package "rms", combined with patient clinical characteristics. Decision curve analysis (DCA) and ROC curves were calculated to detect the accuracy of the nomogram using the R packages "ggDCA" and "timeROC".

Signature gene interference effects from CRISPR–Cas9 screens

The interference efficiency of each gene in CC cells was collated from the CDM, and the interference efficiency of the *SNAPC5, TIGD1, NOP14, MMS19* and *POLRMT* genes in the signature was extracted and visualized using GraphPad 7.0 software. We also further calculated the efficiency of the risk model in determining cell growth interference in common CC cells, including DLD1, HCT116, HT29 and SW620 cells.

Immunohistochemical (IHC), immunofluorescence (IF) and Western blotting (WB)

The Human Protein Atlas (HPA) [20] (https://www.proteinatlas.org/) has been developing of antibodies to human-encoded proteins, and the database includes a large number of immunohistochemical sections of normal and pathological human tissues and partial immunofluorescence sections of cells. The protein expression and intracellular localization of five genes were exhibited on the HPA platform. Western blotting was performed according to a previous study [21]. The following antibodies were used in this study: GAPDH (No. 60004-1-Ig, 1:10000 dilution, Proteintech, China), MMS19 (No: 66049-1-Ig, 1:1000 dilution, Proteintech, China), NOP14 (No:26854-1-AP, 1:500 dilution, Proteintech, China), POLRMT (No: A15605, 1:500 dilution, ABclonal, China), SNAPC5 (No:17272-1-AP, 1:500 dilution, Proteintech, China), and TIGD1 (No:13833-1-AP, 1:500 dilution, Proteintech, China).

Cell culture, infection, siRNAs, and cell function assays

The COAD cells DLD1, HT29, HCT116, and SW620 and the normal-derived colon mucosal cells (NCM460) were maintained in Dulbecco's modified Eagle's medium (DMEM) with 10% foetal bovine serum (HyClone GE Healthcare Life Sciences, Logan, UT, USA) at 37° C in a 5% CO_2 atmosphere. Small interfering RNA (siRNA) targeting the TIGD1 genes and a negative control were purchased from GenePharma Company (Shanghai, China). These siRNAs and negative control RNA were transfected into HCT116 and SW620 cells. The sequences of siRNAs that disturbed *TIGD1* gene are shown in Table 1.

Colony formation: The experimental procedure can be found in our previous study [21]. Cell proliferation assay: A total of 1000 cells were inoculated in 96-well culture plates, and 10 µl Cell Counting Kit (CCK8) reagent (Beyotime Biotechnology Company, Shanghai, China) was added to each well at hours 24, 48, 72, and 96. The cells were cultured for 2 hours, the absorbance was detected at 450 nm, and the growth curve was plotted. Cell cycle assay: The transfected cells were inoculated into 6-well plates, and when the cell abundance reached 80%, the cells were digested and fixed in 70% alcohol overnight. The next day, the fixed cells were stained with a cell cycle kit (Dojindo Laboratories, Japan) and the cell cycle was detected by flow cytometry.

Table 1. *TIGD1* gene siRNA sequence.

Gene name	Sequences	
	Sense (5'-3')	**Antisense (5'-3')**
TIGD1-988	CCAGAACAAAGCCCUAACUTT	AGUUAGGGCUUUGUUCUGGTT
TIGD1-1170	CAGAAGCUUUAGCUAAGAUTT	AUCUUAGCUAAAGCUUCUGTT
Negative control	UUCUCCGAACGUGUCACGUTT	ACGUGACACGUUCGGAGAATT

Drug IC50 analysis based on the CDM signature

The half maximal inhibitory concentration (IC50) values for dozens of drugs were evaluated in both the high- and low- risk groups with the R package "pRRophetic" [22].

Screening hub oncogene in CDM risk genes

Sangerbox 3.0 (http://www.sangerbox.com/) [23] was used to analysis *TIGD1* expression, the relationship with clinical characteristic and overall survival in TCGA-COAD dataset. The PrognoScan (http://www.prognoscan.org/) [24] database was used to analyse the prognostic value of gene from the GSE17536 dataset.

Weighted gene coexpression network analysis (WGCNA) and CDM immune subtypes construction

WGCNA is a method for clustering genes with similar expression patterns and further investigating the association between the modules of various clusters and clinical phenotypes, and is therefore widely used in the analysis of the correlation between clinical phenotypes and gene expression in tumours. WGCNA, as an algorithm for gene coexpression networks, involves three main principles: 1) assuming that gene expression obeys a scale-free distribution, 2) defining the expression matrix of genes and the adjacency function formed by the gene network, and 3) calculating the dissimilarity coefficients of different nodes to construct hierarchical clustering trees. Following the risk scoring of the TCGA-COAD samples, the risk score was incorporated into the WGCNA as a significant clinical phenotype with the following relevant parameters: minModuleSize = 30, MEDissThres = 0.3 and R package "WGCNA". After obtaining the black module genes associated with the risk model, we acquired the differentially expressed black module genes by intersecting the black module genes with the differentially expressed genes, and then divided the TCGA samples into two types by the R package "ConsensusClusterPlus". The two types were termed CDM subtypes.

Immunity analysis of CDM-associated immune subtypes

TCGA-COAD expression data were used to investigate the relationship between the two CDM-associated subtypes and immunity. First, we calculated the immune microenvironment score of the samples by the R package "estimate"; second, we used the Microenvironment Cell Populations counter (MCPcounter) algorithm to calculate the immune cell score of the samples by the R package "MCPcounter" [25]; finally, we estimated the difference between subtypes by the R package "limma" and drew a violin map. Based on the strong association between CDM subtypes and immune cell infiltration, we downloaded TCGA-COAD immunotherapy data from The Cancer Imaging Archive [26] (TCIA: https://tcia.at/home) and performed subtype correlation analysis with programmed cell death protein 1 (PD1) and cytotoxic T-lymphocyte-associated protein 4 (CTLA4) immunotherapy using the R package "ggpubr". Sankey diagrams were used to explore the association between the CDM subtypes and immune subtypes commonly found in clinical tumours: namely, wound healing (C1), IFN-dominant (C2), inflammatory (C3), lymphocyte depleted (C4), immunologically quiet (C5) and TGF-dominant (C6), by the R packages "ggplot2", "ggalluvial" and "dplyr".

Data availability statement

The inquiries of original data can be directed to the corresponding authors.

Code availability

Bioinformatics analysis was based on R 4.1.0, and the involved R packages were described in detail in the methods section. All codes were written and reviewed by GY.Z and ZF.F, respectively. GY.Z and ZF.F had the ultimate power of interpretation and ownership of the code.

Consent for publication

All authors agreed to be published.

RESULTS

CDM analysis with CC flow chart

A series of procedures was designed to investigate the prognosis of CC based on the characteristics of the CDM genes (Figure 1).

Expression of upregulated CDM genes and functional enrichment analysis

After normalization of the TCGA-COAD data, 8,179 genes with upregulated expression in CC were screened by the selection criteria of p<0.05 and logFC>1. Additionally, 1,613 CC-dependent genes were discovered by collating interference efficiency with in 55 type of CC cells. A total of 1,304 CDM genes were differentially expressed in CC (Supplementary Figure 1A), and the heatmap in Supplementary Figure 2 depicted their differential expression in colon normal tissue and CC tissue. To explore the signalling pathways and functional analysis of the different CDM genes in CC in depth, we performed KEGG and GO analyses. As shown in Supplementary Figure 1B, five significant pathways cell cycle, nucleocytoplasmic transport, ribosome, ribosome biogenesis in eukaryotes and spliceosome were enriched. Biological process analysis revealed that different CDM genes were involved in ncRNA metabolic processes, ncRNA processing ribonucleoprotein complex biogenesis and ribosome biogenesis (Supplementary Figure 1C).

Construction and validation of the CDM signature

First, 446 TCGA-COAD samples, excluding eight samples with missing clinical data, were randomly assigned to the training and testing cohorts in a 1:1 ratio, with 224 cases in the training cohort and 222 cases in the testing cohort. Next, univariate Cox hazard analysis was used to regress the differentially expressed CDM genes in the TCGA-COAD data according to a p<0.05 criterion, and 29 genes were discovered to be substantially associated with the survival prognosis of the TCGA-COAD patients (Supplementary Table 1). Twenty-nine genes are not conducive to medical detection and promotion. As a result, Lasso Cox regression analysis was performed to filter for precise and valid prognosis-related genes, and the confidence interval under each lambda is shown in Supplementary Figure 3A. The prognosis-related genes were narrowed down to seven genes (Supplementary Figure 3B). Multivariate Cox analysis was employed to model the prognostic assessment of the five CDM genes. The five CDM-gene model formula was as follows: CDM risk score = $0.0538052 \times \exp^{SNAPC5} + 0.925782 \times \exp^{TIGD1} + -1.54469 \times \exp^{NOP14} + 1.559607 \times \exp^{MMS19} + 0.592523 \times \exp^{POLRMT}$.

We calculated the risk scores for the training cohort samples based on the expression level of the five CDM genes, and then determined the patient survival status distribution according to the risk scores. It was discovered that the high-risk CC population had shorter
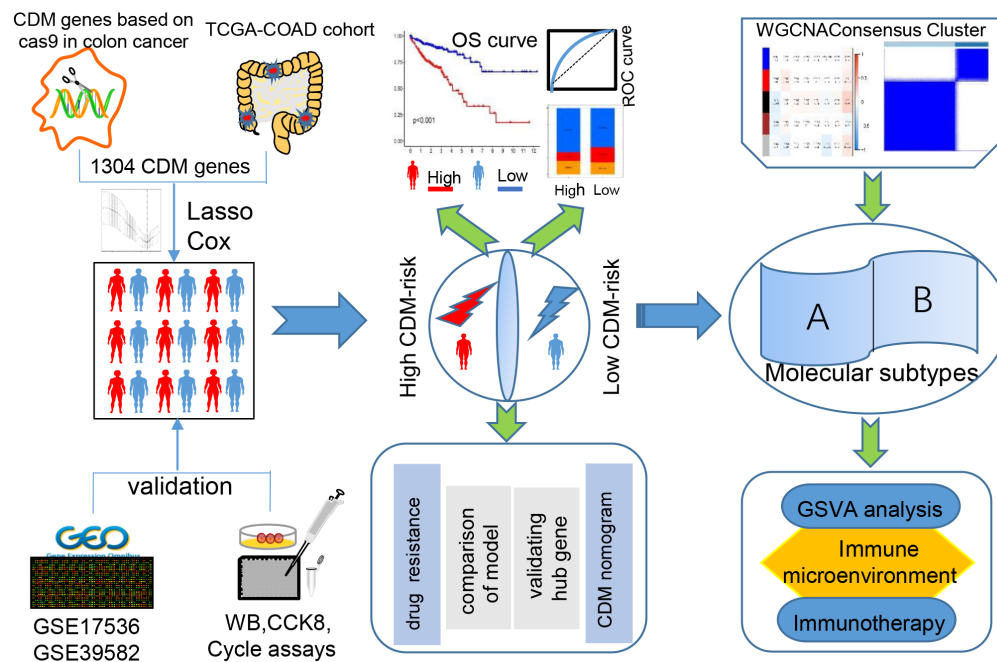

Figure 1. Cancer Dependency Map (CDM) with colon cancer (CC) analysis flow chart.

survival times in the training cohort than the low-risk group, and the AUC predictive values in the time-dependent ROC curves for 1, 3, and five years were 0.801, 0.789, and 0.853, respectively, indicating that the CDM risk signature has good survival predictive power for CC. Subsequently, the heatmap revealed that SNAPC5, TIGD1, MMS19 and POLRMT were highly expressed in the high-risk group, while NOP14 was expressed at low levels in the low-risk group (Figure 2A–2E). To determine the robustness of CDM signature, we used the same coefficients obtained from the training cohort in the internal testing cohort, all TCGA-COAD cohorts, and GSE17536 external validation cohort. The risk score evaluation, patient survival status distribution, heatmap of the five CDM genes, K-M survival curve and time-dependent ROC curves with GSE17536 external validation cohort are shown in Figure 2F–2J. Supplementary Figure 4A–4J correspond to the all TCGA-COAD cohort and partial testing TCGA-COAD cohort, respectively. The above findings showed that the CDM signature is robust, and play a consistent prognostic predictive role in various CC datasets.

Good predictive ability of the CDM signature compared with other models

Based on a literature review, four prognostic signatures were chosen for comparison with the CDM signature, namely, the Li [27], Nie [28], Ren [29], and Xu [30] signature, were chosen for comparison with the CDM signature to TCGA-COAD cohort. To make the results more informative, we calculated risk scores for each TCGA-COAD sample based on the risk genes modelled for each individual, divided the samples into high and low risk groups, and performed survival analysis and ROC curve validation.

When comparing Figure 3A–3D, 3I, the AUC of the 1-year, 3-year, and 5-year time-dependent ROC curves for the CDM risk models was clearly higher than that of the other four signature. Similarly, the survival curves of the CDM signature outperformed those of the others as shown in Figure 3E–3H, 3J. The consistency index (c-index) and RMS values were introduced to further compare the prognostic prediction
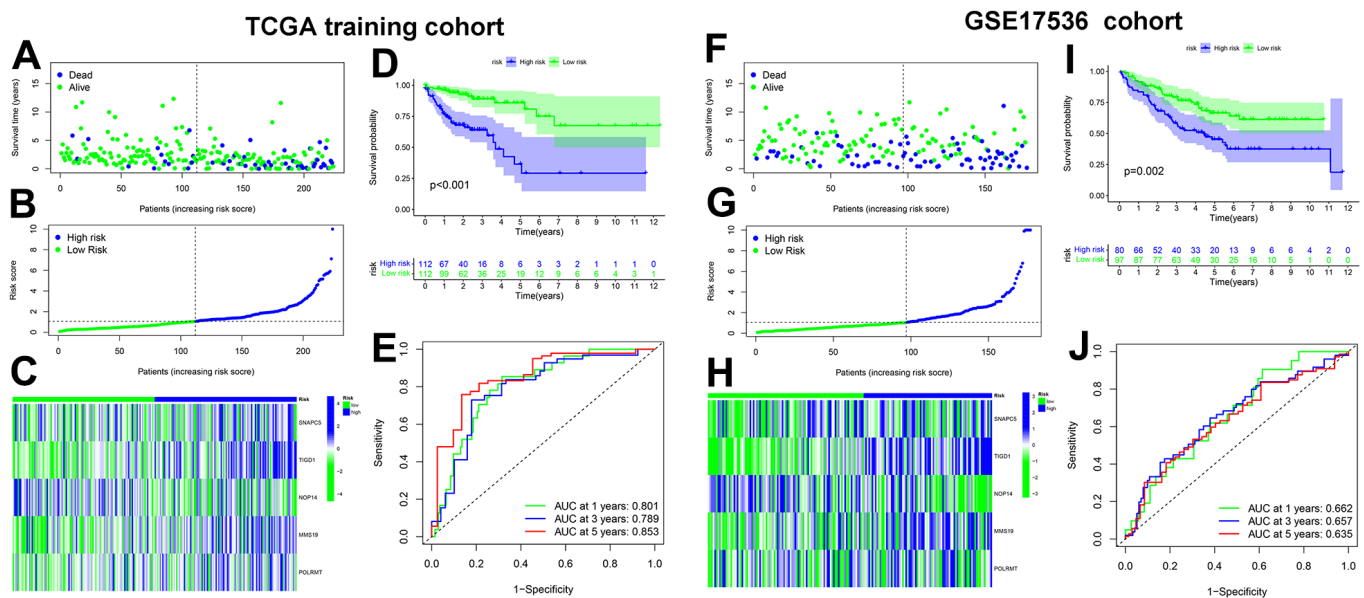

Figure 2. Cancer Dependency Maps (CDMs) signature construction and validation in TCGA training cohort and GSE17536 cohort. (**A**) Survival scatter plot in TCGA training cohort. A dot represents a CC patient (blue represents Dead, Green represents alive). Dotted lines show the median of risk score that dichotomize patients into high and low groups. (**B**) Risk score plot in TCGA training cohort. The patients of CC are ordered by the risk score of the CDM signature. Dotted lines show the median of risk score that dichotomize patients into high (blue) and low (green) groups. (**C**) Heatmap of five CDM signature genes (*SNAPC5*, *TIGD1*, *NOP14*, *MMS19* and *POLRMT*) in TCGA training cohort. Blue represents high-CDM risk group and green represents low-CDM risk group. (**D**) Kaplan-Meier plot in TCGA training cohort (*p*<0.001). Survival curves for high (blue) and low (green) risk-CDM groups dichotomized at the median of risk-CDM score are plotted. 95% confidence intervals for each group are also indicated by shadow area. (**E**) The ROC curve of CDM signature in TCGA training cohort. Green show one year AUC value (AUC value=0.801), blue show three years AUC value (AUC value=0.789) and red show five years AUC value (AUC value=0.853). (**F**) Survival scatter plot in GSE17536 cohort. (**G**) Risk score plot in GSE17536 cohort. (**H**) Heatmap of five CDM signature genes (*SNAPC5*, *TIGD1*, *NOP14*, *MMS19* and *POLRMT*) in GSE17536 cohort. (**I**) Kaplan-Meier plot in GSE17536 cohort (*p*=0.002). (**J**) The ROC curve of CDM signature in GSE17536 cohort. One year AUC value (AUC value=0.662), three years AUC value (AUC value=0.657) and five years AUC value (AUC value=0.635).

performance of these models for TCGA-COAD cohort. As shown in Figure 3K, 3L, the CDM signature seemed to have the highest c-index of all of the models, and the RMS values were on par with them. A series of verifications and comparisons have proven that the performance of the CDM signature overall was better than that of the other four. Furthermore, and perhaps most importantly, the CDM signature employs fewer genes to produce more accurate prediction results with good tractability for clinical replication.

Clinical correlation analysis and the nomogram improves CDM signature predictive ability in CC

To illustrate the clinical utility of the CDM signature more fully and effectively, we explored the association between the signature and clinical features, and further constructed and evaluated nomogram by integrating the signature with clinical features. In terms of age and gender, based on an analysis of the CDM genes in the TCGA-COAD and GSE17536 datasets, there was no
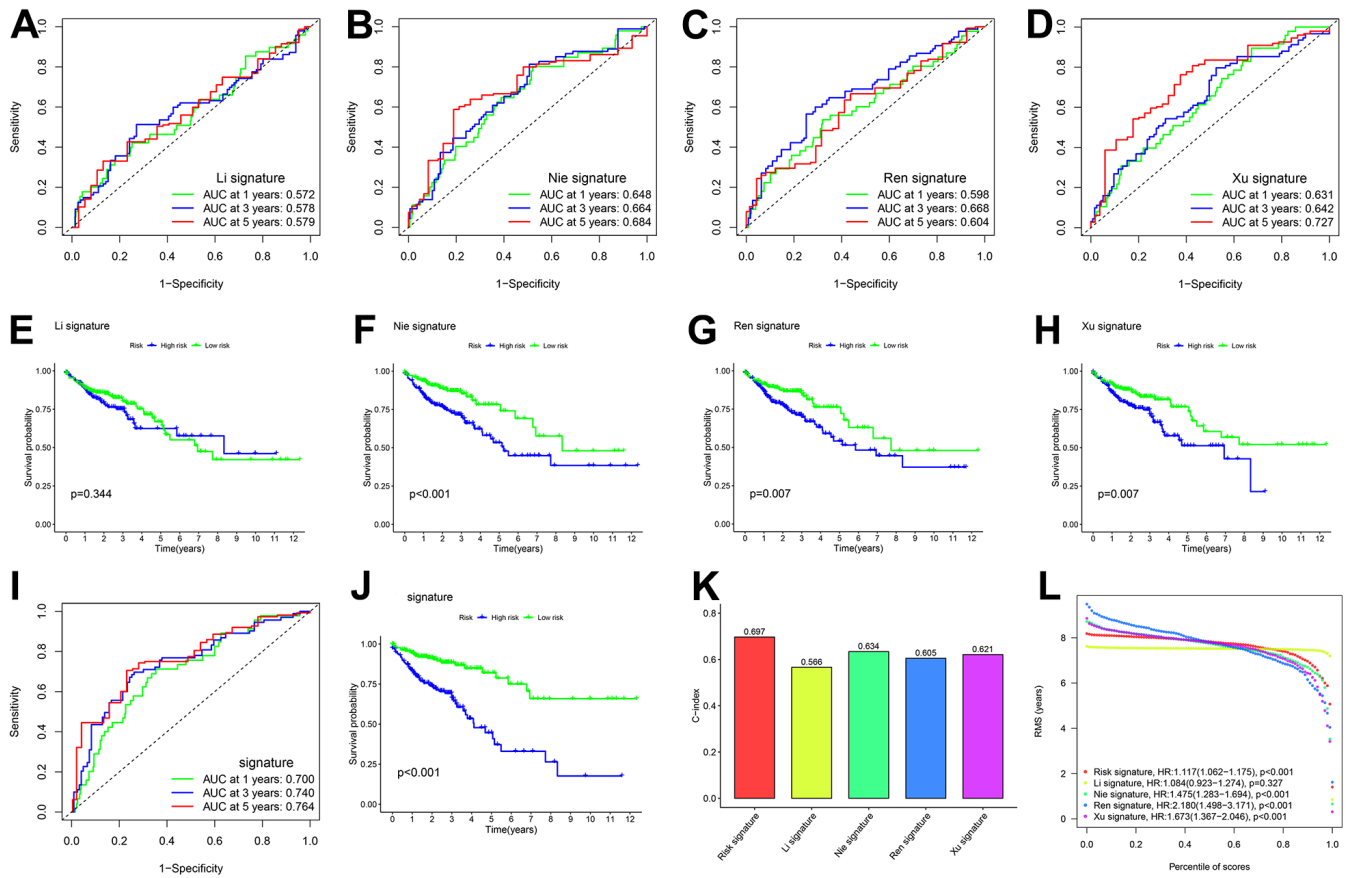

Figure 3. Comparison of five CDM genes model with other models. (**A**, **B**) The ROC curve of signatures in TCGA-COAD cohort. (**A**) Li signature (ten genes). Green show one year AUC value (AUC value=0.572), blue show three years AUC value (AUC value=0.578) and red show five years AUC value (AUC value=0.579). (**B**) Nie signature (five genes). Green show one year AUC value (AUC value=0.648), blue show three years AUC value (AUC value=0.664) and red show five years AUC value (AUC value=0.684). (**C**) Ren signature (eight genes). Green show one year AUC value (AUC value=0.598), blue show three years AUC value (AUC value=0.668) and red show five years AUC value (AUC value=0.604). (**D**) Xu signature (eight gene). Green show one year AUC value (AUC value=0.631), blue show three years AUC value (AUC value=0.642) and red show five years AUC value (AUC value=0.727). (**E**–**H**) Kaplan-Meier plot in all TCGA –COAD cohort (p<0.001). Survival curves for high (blue) and low (green) risk groups dichotomized at the median of risk score are plotted. (**E**) shows Li signature (p=0.344), (**F**) shows Nie signature (p<0.001), (**G**) shows Ren signature (p=0.007) and (**H**) shows Xu signature (p=0.007). (**I**) The ROC curve of CDM signature in all TCGA-COAD cohort. Green show one year AUC value (AUC value=0.700), blue show three years AUC value (AUC value=0.740) and red show five years AUC value (AUC value=0.764). (**J**) Kaplan-Meier plot in all TCGA –COAD cohort (p<0.001). Survival curves for high (blue) and low (green) risk-CDM groups dichotomized at the median of risk-CDM score are plotted. (**K**) C-index of the five prognostic risk models. The higher the C-index, and the more reliable the signature. Red shows CDM risk signature (C-index=0.697). Yellow shows Li signature (C-index=0.566). Cyan shows Nie signature (C-index=0.634). Blue shows Ren signature (C-index=0.605), Dull red shows Xu signature (C-index=0.621). (**L**) The RMS analysis of the five prognostic risk signatures. Red shows CDM risk signature. Yellow shows Li signature. Cyan shows Nie signature. Blue shows Ren signature, Dull red shows Xu signature.

significant difference between the high- and low- risk groups (Supplementary Figure 3E–3H). However, we discovered that the worse the tumour stage was, the higher the patient CDM risk scores (Figure 4A, 4D). According to the TCGA-COAD dataset, the depth of tumour infiltration (T) and the number of lymph node metastases (N) were related to the CMD risk scores (Figure 4B, 4C). However, there was no significant difference between the CDM risk scores and tumour grades based on the GES17536 dataset (Figure 4E). Microsatellite instability (MSI) is a clinically significant marker that results from a functional defect in DNA mismatch repair in tumour tissue for CC. There is substantial evidence that microsatellite instability high (MSI-H) is a prognostic marker in patients with stage II colorectal cancer [31, 32], but there is also evidence that MSI-H patients do not benefit from adjuvant chemotherapy with 5-fluorouracil (5-Fu), and MSI-H can be used as a predictive marker of the

ineffectiveness of adjuvant 5-FU treatment for CC [33]. Microsatellite stability (MSS) patients accounted for 67 percent of the low-risk CDM group, microsatellite instability low (MSI-L) for 13 percent, and MSI-H for 20 percent; the composition ratios of the three in the high-risk CDM group were significantly different from those in the low-risk group, with MSS patients accounting for 59 percent, MSI-L for 23 percent, and MSI-H for 18 percent (Figure 4F). Figure 4G also shows that MSI-L patients had the highest CDM risk score, followed by MSI-H patients, and finally MSS patients, with a statistically significant difference.

To improve the accuracy of CDM signature prediction, the nomogram was constructed by combining clinical features from two datasets, TCGA-COAD and GSE17536. The nomogram in Figure 5A is more intuitive and convenient for predicting patients' clinical
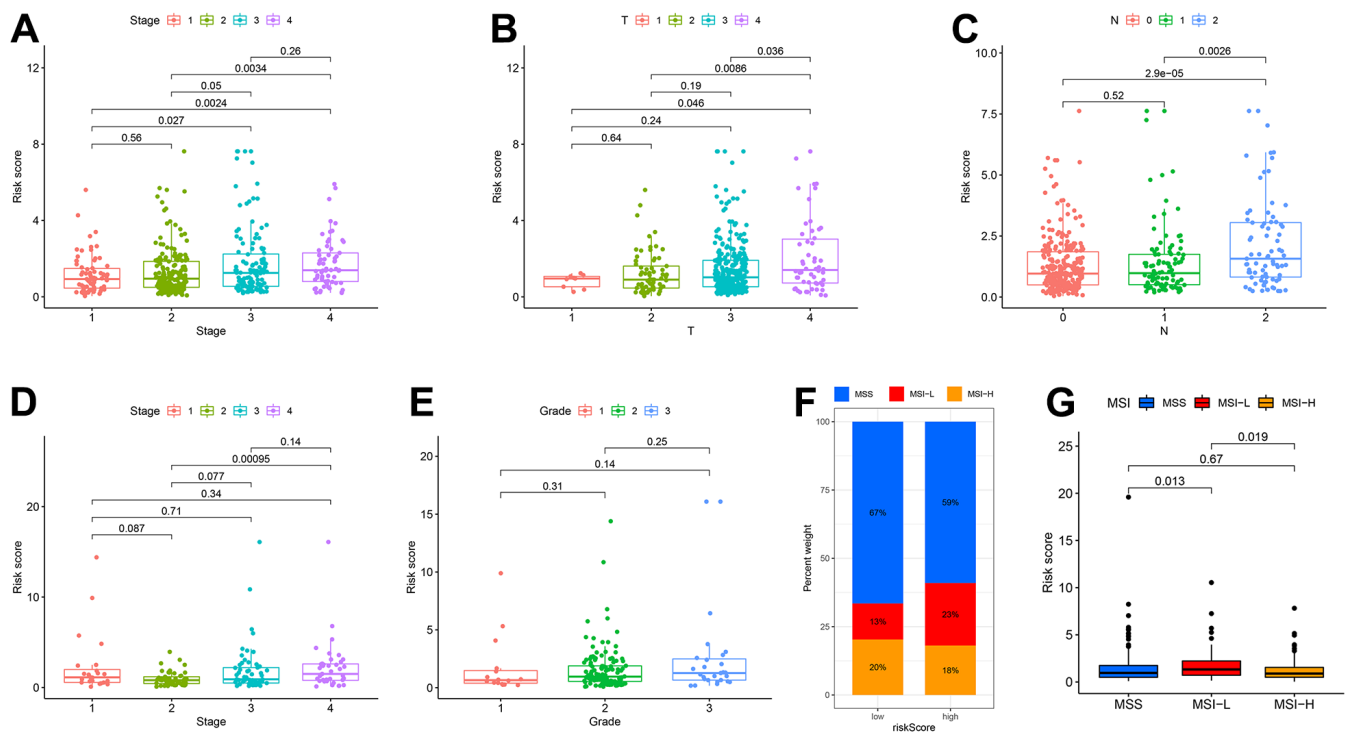

Figure 4. Correlation among clinicopathological classifications and CDM signature risk scores. (**A**) Boxplot. Correlation between tumour stage and risk-CDM score in TCGA-COAD cohort. Pink represents I stage, green represents II stage, Cyan represents III stage and violet represents IV stage. (**B**) Boxplot. Correlation between depth of tumor invasion (T) and risk-CDM score in TCGA-COAD cohort. Pink represents T1, green represents T2, cyan represents T3 and violet represents T4. (**C**) Boxplot. Correlation between lymph node metastasis (N) and risk-CDM score in TCGA-COAD cohort. Pink represents N0 (not metastasis), green represents N1 (1-3 regional lymph node metastases) and blue represents N2 (4 or more regional lymph node metastases). (**D**) Boxplot. Correlation between tumour stage and risk-CDM score in GSE17536 cohort. Pink represents I stage, green represents II stage, Cyan represents III stage and violet represents IV stage. (**E**) Boxplot. Correlation between grade and risk-CDM score in GSE17536 cohort. Pink represents G1 (good), green represents G2 (moderate) and blue represents G3 (poor). (**F**) Microsatellite stability histogram. MSS (blue), MSI-L (red) and MSI-H (orange) of High/Low risk group samples in TCGA-COAD cohort. (**G**) Boxplot. Correlation between microsatellite stability and risk-CDM score in TCGA-COAD samples. MSS (blue), MSI-L (red) and MSI-H (orange).

prognoses, with a high operability. In addition, because the calibration plot is close to the 45° line, the column line plot in Supplementary Figure 3C, 3D performed well. Similarly, the results of the DCA considered the CDM signature and nomogram to be highly credible (Figure 5B). When compared to the AUC value of the CDM signature, the AUC of the nomogram was significantly improved (Figure 5C). Supplementary Figure 5A–5C exhibited the result of nomogram, DCA and ROC in GSE17536 cohort.

CDM risk score improves the prediction of targeted therapy resistance

To make the CDM signature more useful for guiding the treatment of CC patients, we comprehensively explored drug resistance relationships based on the model and found that many drugs, such as molecular targeting inhibitors, immunosuppressive drugs and even some Chinese herbal medicines, had significantly different IC50 values in both the high- and low- risk groups. The drugs shown in Figure 6A–6O had higher IC50 values in the low-risk group than in the high-risk group, implying that these drugs may benefit low-risk patients to some extent.

Expression of CDM risk genes in CC tissues and cells

To determine the protein expression characteristics of these CDM signature genes, we investigated IHC of CC tissues, IF with U2-OC cell images from the HPA database and WB of CC cells. The WB results shown in Figure 7A, demonstrated that the protein expression of MMS19, NOP14, POLRMT, SNAPC5 and TIGD1 in NCM460 cells was far lower than that in CC cells. Additionally, the IHC results demonstrated that MMS19, NOP14, POLRMT, SNAPC5 and TIGD1 were moderately to strongly expressed in CC tissues (Figure 7B–7F). Cellular immunofluorescence localization revealed that NOP14 and TIGD1 were mainly localized in the nucleus, POLRMT was mostly found on microtubules in the cytoplasm, and MMS19 and SNAPC5 were expressed in both the nucleus and cytoplasm (Figure 7G–7K).

Gene effect scores based on CRISPR–Cas9 and silencing TIGD1 genes inhibited CC cell proliferation

To further screen for oncogenes, the survival performance of five risk genes was analyzed in both
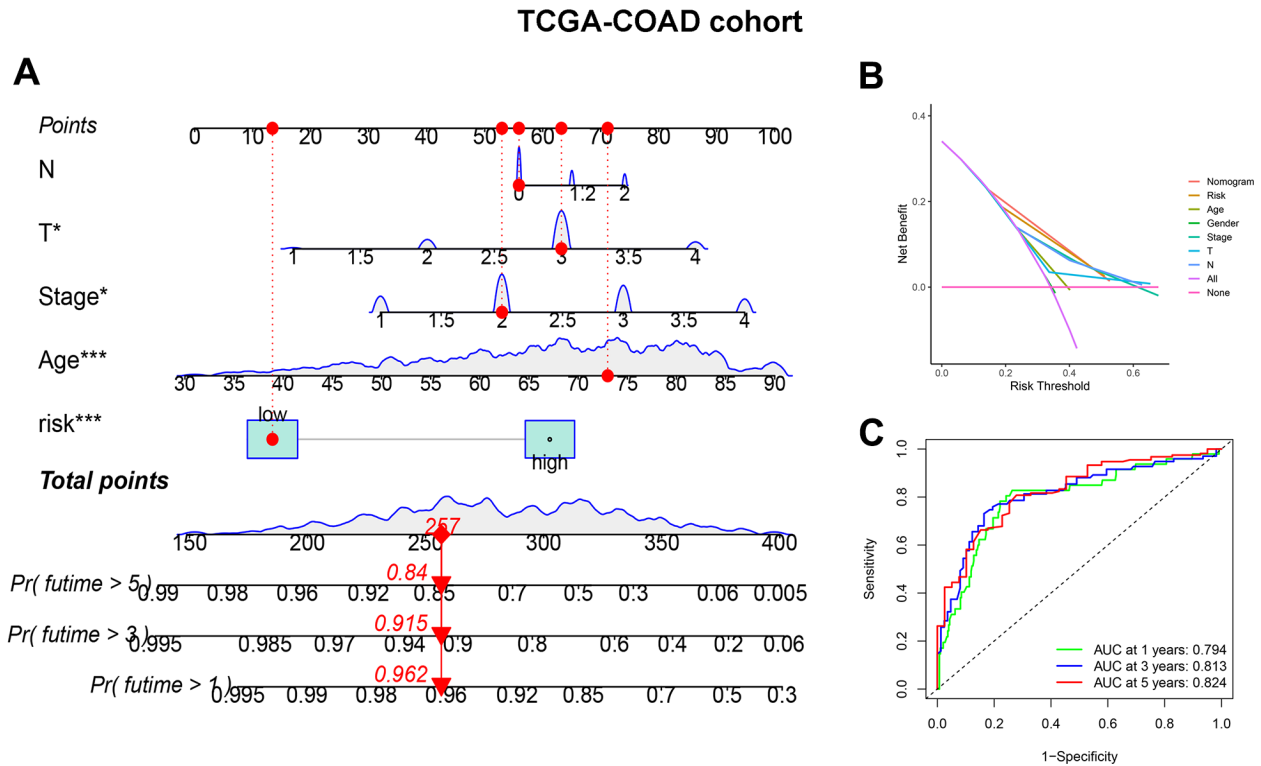

Figure 5. CDM-associated nomogram construction in TCGA-COAD cohort. (**A**) CDM-associated Nomogram for predicting 1-, 3-, and 5-year OS of patients with TCGA-COAD cohort. (**B**) The decision curve (DCA) analysis for CDM-associated nomogram in TCGA-COAD cohort. (**C**) The ROC curve of CDM-associated Nomogram in all TCGA-COAD cohort. Green show one year AUC value (AUC value=0.794), blue show three years AUC value (AUC value=0.813) and red show five years AUC value (AUC value=0.824).
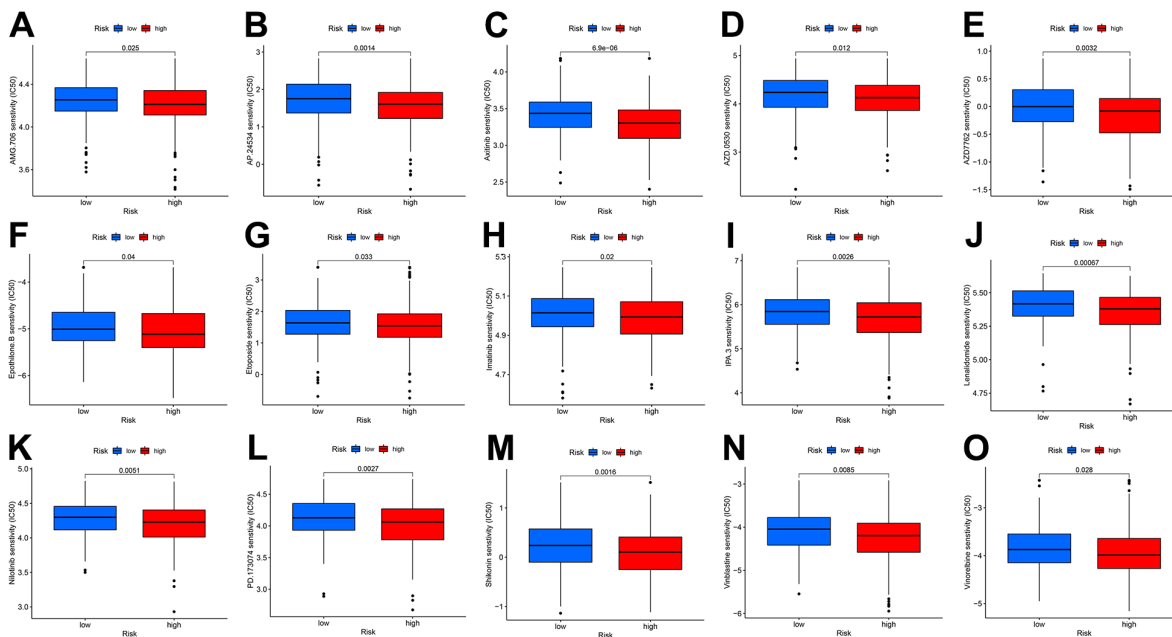
Figure 6. Drug IC50 analysis in CDM risk signature (blue represents low CDM risk group, red represents high group). (**A**) AMG.706 (*p*=0.025); (**B**) AP.24534 (*p*=0.0014); (**C**) Axitinib (*p*=6.9e^-6); (**D**) AZD.0530 (*p*=0.012); (**E**) AZD7762 (*p*=0.0032); (**F**) Epothilone (*p*=0.04); (**G**) Etoposide (*p*=0.033); (**H**) Imatinib (*p*=0.02); (**I**) IPA.3 (*p*=0.0026); (**J**) Lenalidomide (*p*=0.00067); (**K**) Nilotinib (*p*=0.0051); (**L**) PD.173074 (*p*=0.0027); (**M**) Shikonin (*p*=0.0016); (**N**) Vinblastine (*p*=0.0085); (**O**) Vinorelbine (*p*=0.028).
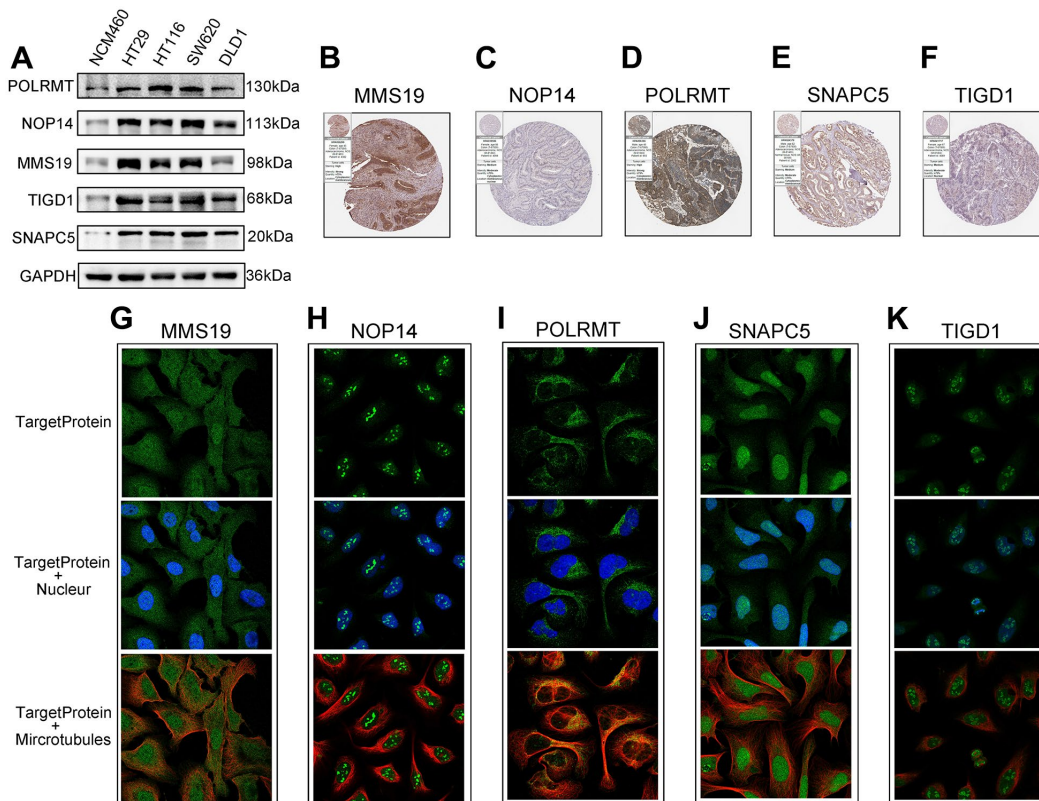
Figure 7. The expression of the identified five CDM genes from the human protein atlas database (HPA). (**A**) Western blotting. The protein expression of five CDM genes in the normal derived colon mucosal (NCM460) and colon cancer cells (HT29, HCT116, SW620 and

DLD1). (**B–F**) Protein expressions of five CDM genes in colon cancer tissue specimens by immunohistochemical; (**B**) for MMS19; (**C**) for NOP14; (**D**) for POLRMT; (**E**) for SNAPC5; (**F**) for TIGD1. (**G–K**) Protein expressions and localization of five CDM genes in U2-OC cell by immunofluorescence (green represents five CDM-associated protein, blue represents nucleus and red represents microtubule); (**G**) for MMS19; (**H**) for NOP14; (**I**) for POLRMT; (**J**) for SNAPC5; (**K**) for TIGD1.

TCGA-COAD and GSE17536 datasets (Supplementary Figure 6A–6J). Taking into account the expression differences, only *TIGD1* has excellent predictive power in both datasets (Figure 8A–8C). The clinicopathological characteristics and the expression of TIGD1 are closely correlated. For instance, the higher the expression of

TIGD1, the more advanced the tumour stage and the more likely to lymph node metastasis and distant metastasis; while, tumour infiltration depth not significant (Figure 8D–8G). Subsequently, the prognostic value of *TIGD1* gene was further verified by the ProgScan in the GSE17536 dataset (Figure 9).
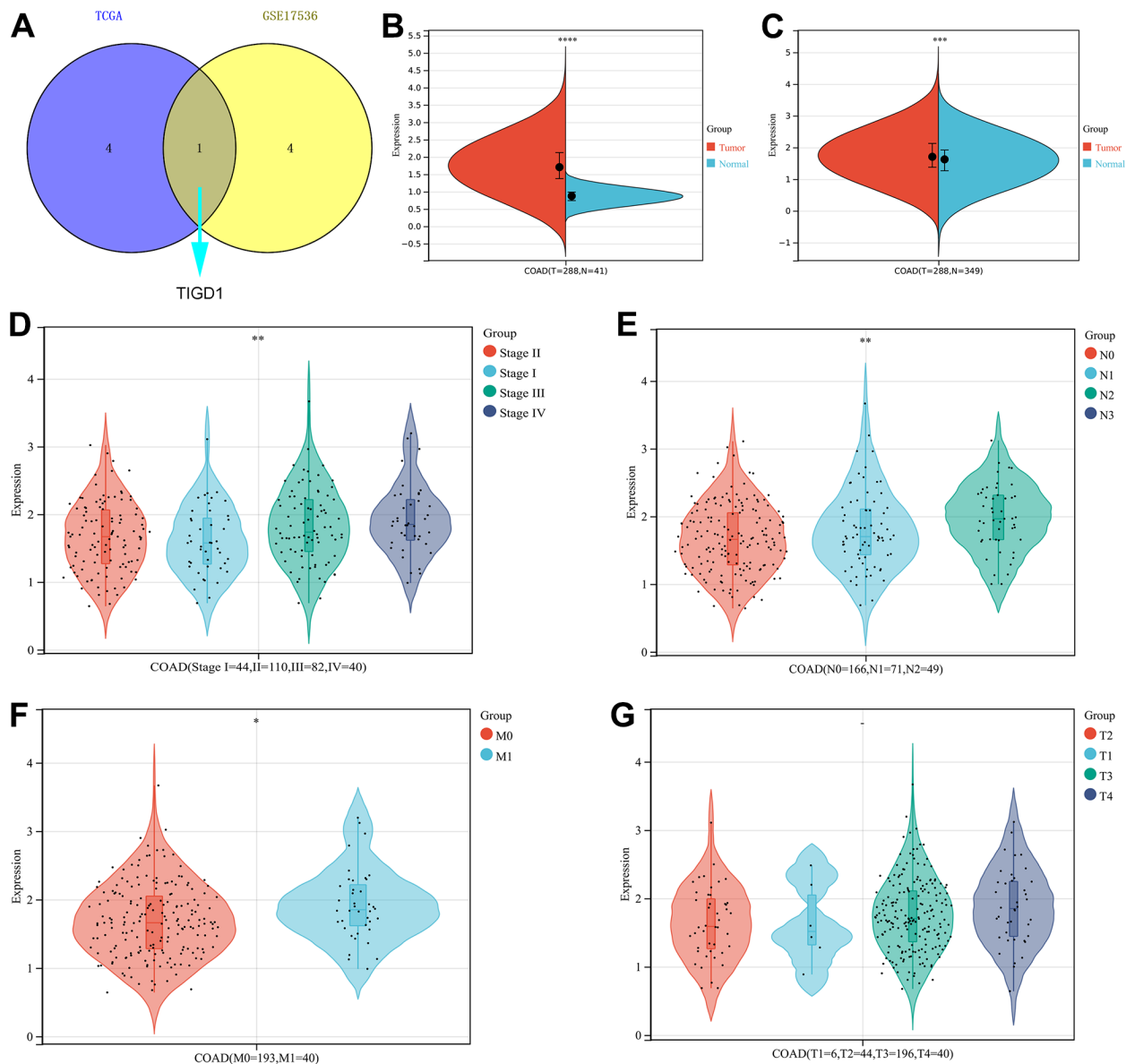

Figure 8. The expression and clinicopathological classifications of *TIGD1* in TCGA-COAD cohort. (**A**) Venn plot. Kaplan-Meier survival analysis for five CDM-associated genes in in TCGA–COAD and GSE17536 cohort. Blue represents TCGA-COAD and yellow represents GSE17536 cohort. (**B, C**) Violin plot. The mRNA expression of TIGD1. Red represents tumour and blue represents normal. (**B**) for TCGA-COAD dataset (p<0.0001). (**C**) for TCGA-COAD and GTEx dataset (p<0.001). (**D**) Violin plot. Correlation between tumour stage and TIGD1 expression

in TCGA-COAD cohort. Pink represents I stage, green represents II stage, Cyan represents III stage and violet represents IV stage ($p<0.01$). (**E**) Violin plot. Correlation between lymph node metastasis (N) and TIGD1 expression in TCGA-COAD cohort. Pink represents N0 (not metastasis), blue represents N1 (1-3 regional lymph node metastases) and cyan represents N2 (4 or more regional lymph node metastases) ($p<0.01$). (**F**) Violin plot. Correlation between metastasis (M) and TIGD1 expression in TCGA-COAD cohort. Pink represents M0 (not metastasis), blue represents M1 (metastases) ($p<0.05$). (**G**) Violin plot. Correlation between depth of tumor invasion (T) and TIGD1 expression in TCGA-COAD cohort. Pink represents T1, blue represents T2, Cyan represents T3 and violet represents T4 ($p>0.05$).
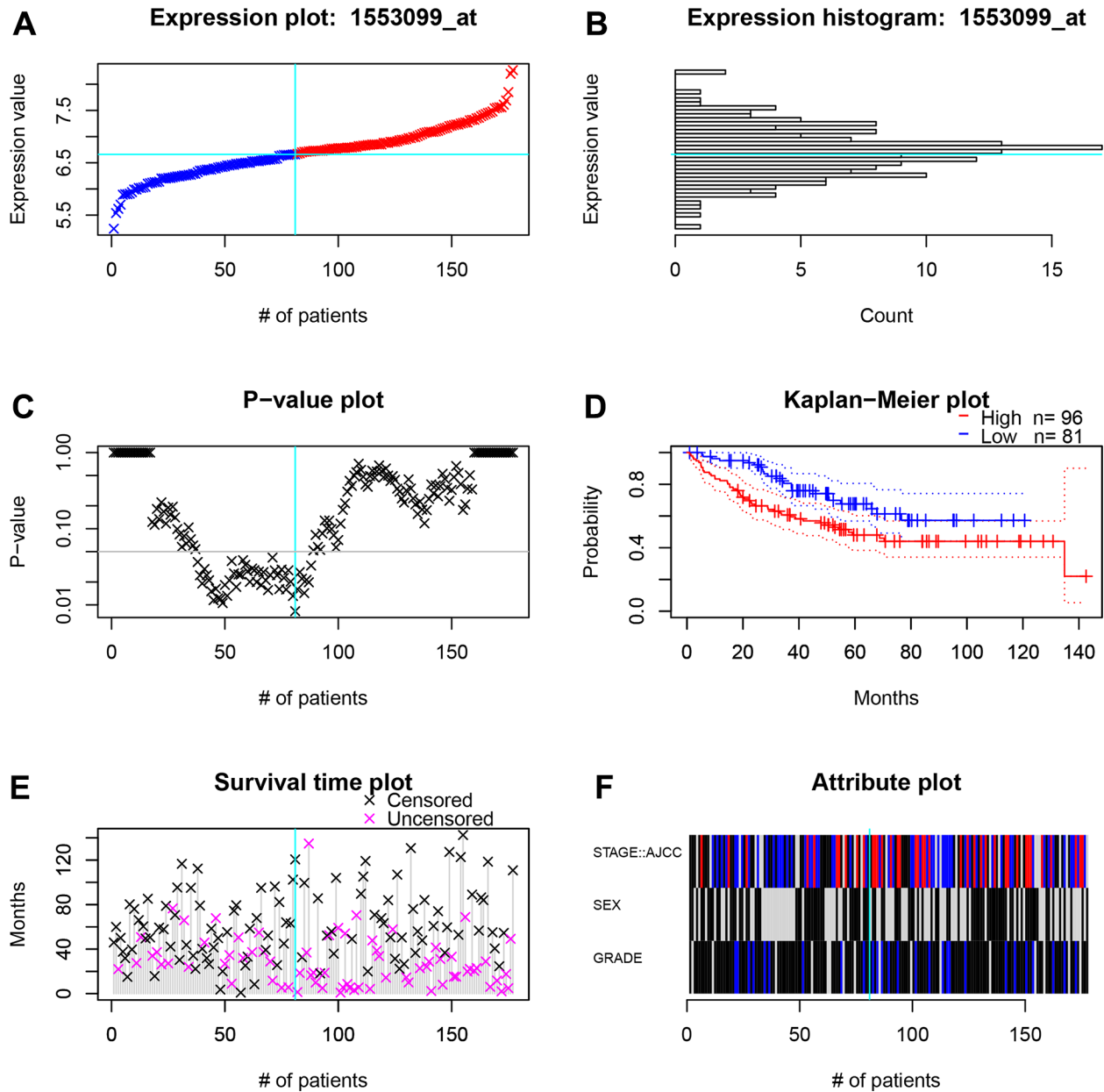

Figure 9. The survival and clinicopathological classifications of *TIGD1* in GSE17536 cohort. (**A**) Expression plot. CC patients are ordered by the TIGD1 expression. Cyan lines show the optimal cutpoints that dichotomize patients into high (red) and low (blue) expression groups. (**B**) TIGD1 expression histogram. The line of the optimal cutpoint is also shown (cyan). (**C**) P-value plot of *TIGD1*. For each potential cutpoint of expression measurement, patients are dichotomized and survival difference between high and low expression groups is calculated by log-rank test. The cutpoint to minimize the P-value is determined and indicated by the cyan line. (**D**) Kaplan-Meier plot of TIGD1. Survival curves for high (red) and low (blue) expression groups dichotomized at the optimal cutpoint are plotted. 95% confidence intervals for each group are also indicated by dotted lines. (**E**) Survival scatter plot of TIGD1. A dot represents a CC patient (pink represents uncensored, black represents censored). (**F**) Heatmap of *TIGD1* expression among stage, sex and grader. Black represents low expression and red represents high expression.

CRISPR–Cas9 is a relatively mature gene editing technology that is widely used in basic and clinical settings. In this study, the interference effects of genes associated with the CDM signature were deeply mined on the CDM platform by using CRISPR–Cas9 technology. As Figure 10A shows, the interference values of all five risk genes were less than -0.5, with the best interference value associated with the knockout SNAPC5 gene and the worst being with POLRMT, both

of which are important factors affecting CC cell growth. After further statistical analysis, it was found that the growth of two types of CC cells, DLD1 and HCT116, was most affected by this model, followed by SW620, and finally HT29 (Figure 10B).

To determine whether the *TIGD1* genes is hub genes in CC, we designed siRNAs for the interference sites of *TIGD1*. As shown in Figure 10C, siRNAs interfering
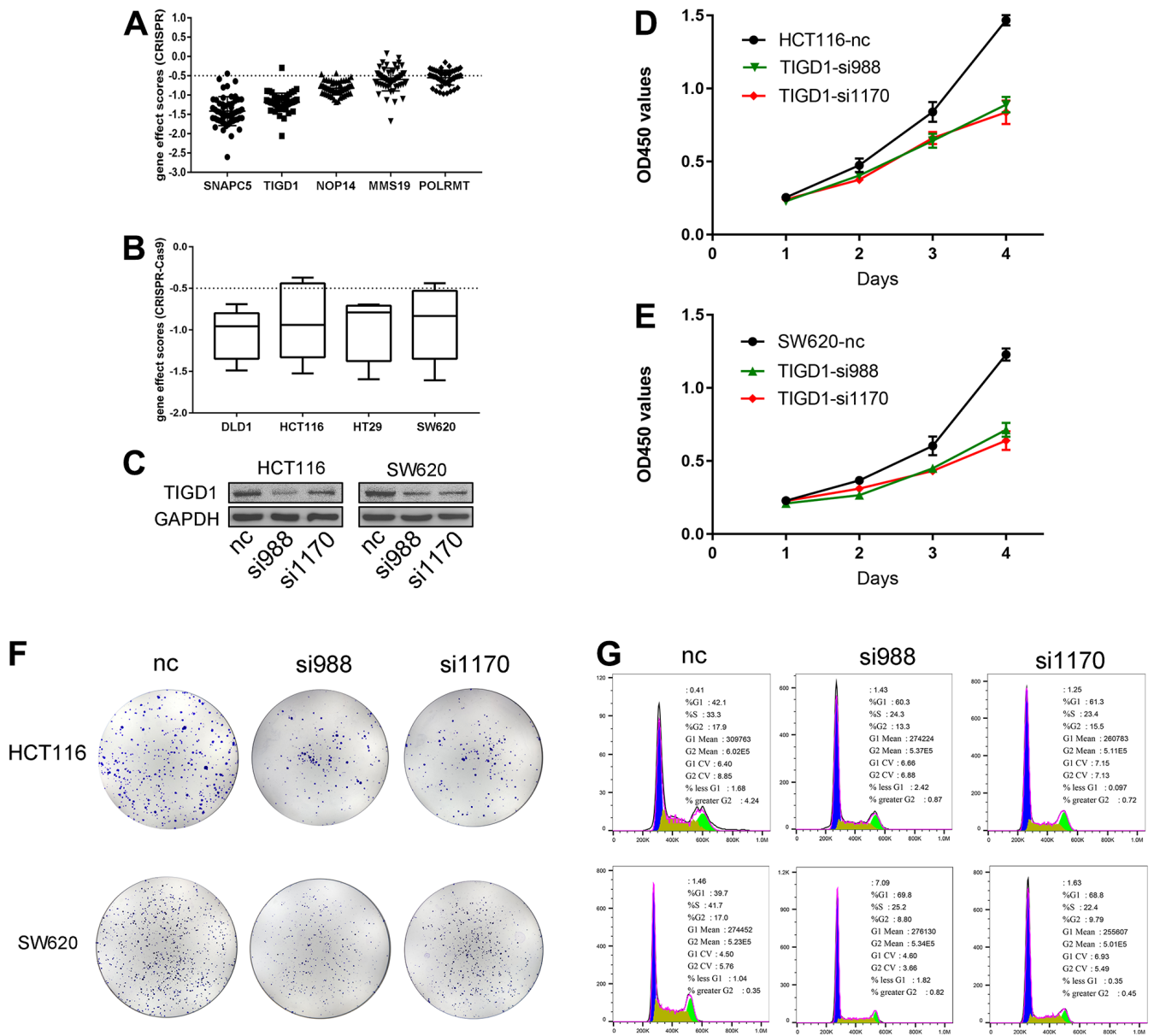

Figure 10. Gene effect scores based on CRISPR-Cas9 and Silencing TIGD1 gene inhibited CC cells proliferation and cycle. (**A**) The interference values of five CDM risk genes. (**B**) The interference values of colon cancer cell lines (DLD1, HCT116, HT29, SW620). (**C**) siRNAs interfered TIGD1 expression in HCT116 and SW620 (three group: nc, si988 and si1170). (**D, E**) HCT116 and SW620 cells were significantly inhibited in growth with knocking down the *TIGD1* gene expression by cck8 assays. (**D**) for HCT116, (**E**) for SW620. (**F**) The cloning ability of HCT116 and SW620 cell lines were significantly inhibited with knocking down the *TIGD1* gene expression by clone formation assays. (**G**) Silencing the expression of *TIGD1* gene retarded the proliferation cycle of HCT116 and SW620 in the G1 phase.

with *TIGD1* had a good inhibitory effect in SW620 and HCT116 cells. Cell proliferation was detected by clone formation and CCK8 assays. The growth of both HCT116 and SW620 cells was significantly inhibited by knocking down the *TIGD1* (Figure 10D, 10E). In addition, the cloning ability of the cells was also clearly affected by silencing the gene as shown in Figure 10F. The results of cell cycle experiments revealed that silencing the gene retarded the proliferation cycle of HCT116 and SW620 cells more or less in the G1 phase, thus retarding the cell growth rate (Figure 10G). This result is also consistent with our previous KEGG findings regarding the impact of CDM-related genes on the cell cycle (Supplementary Figure 1B).

WGCNA and CDM immune subtype construction

WGCNA was used to analyse coexpressed gene clusters associated with CDM signature scores based on the expression profiles in the TCGA-COAD data. The power value was chosen to be three to ensure a scale-free network (R^2=0.9, Supplementary Figure 7A, 7B). Five coexpression models were clustered in the hierarchical clustering tree (Supplementary Figure 7C), with the black module having the strongest positive correlation with the CDM risk scores (Cor=0.2, P = 4e[-05], Figure 11A). The black module contained 5,662 coexpressed genes, there were 8,179 TCGA-COAD differentially upregulated genes, and the two together included 1,220 differentially coexpressed upregulated genes (Figure 11B).
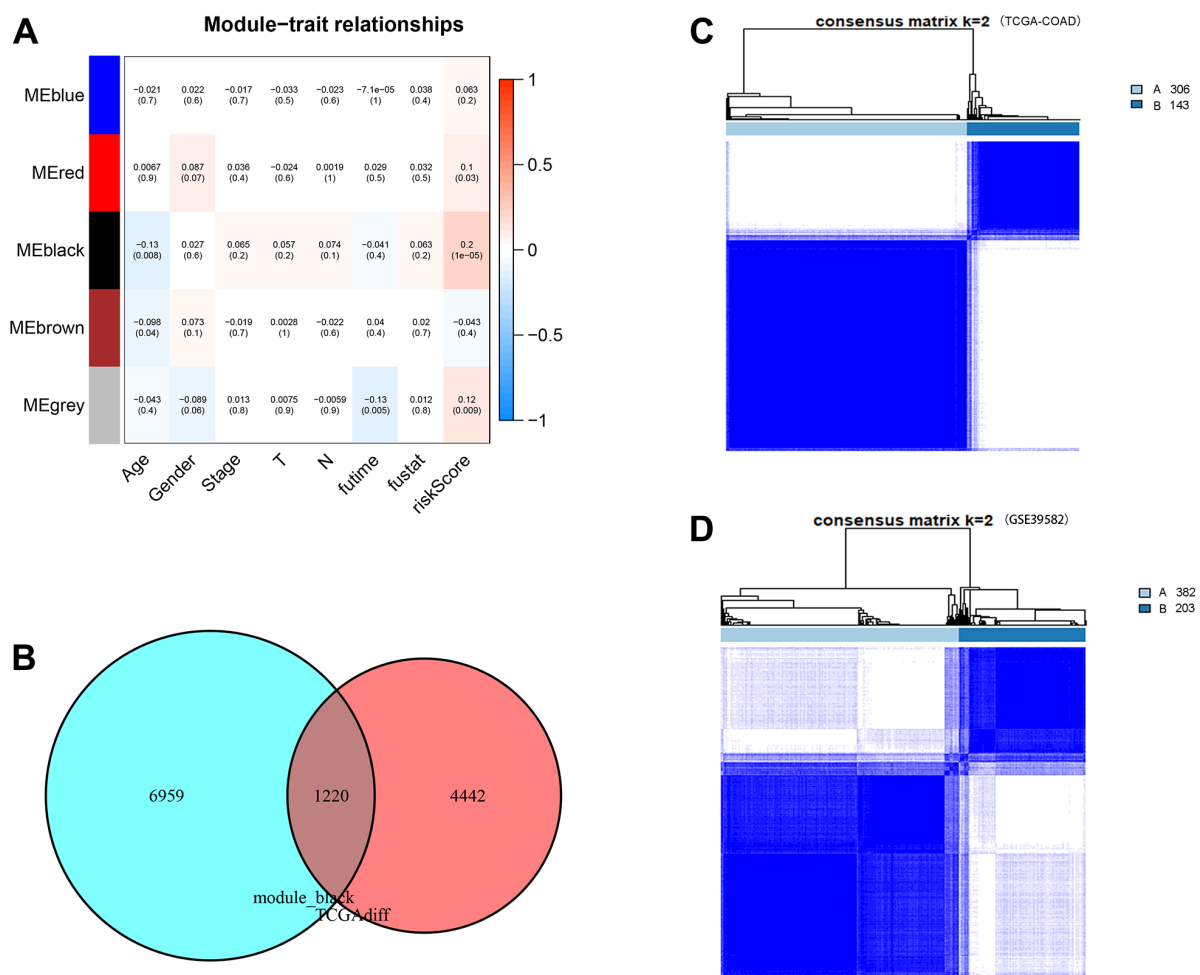

Figure 11. WGCNA analysis, consensus clustering and immune subtype construction. (**A**) Module–Clinical Trait Relationships of consensus module eigengene and different clinical features of COAD. Each row in the table corresponds to a module eigengene (ME), and each column to a clinical parameter (five module: blue, red, black and grey; eight clinical features: Age, Gender, Stage, T, N, futime, fustat and CDM riskScore). Each cell contained the correlation coefficients and p value (red, positively correlated; green, negative correlated). (**B**) Venn plot of CDM-associated co-expressed genes (5,662 genes) and upregulated genes (8,179 genes). Red represents coexpressed genes and cyan represents TCGA-COAD upregulated genes. (**C**) The consensus score matrix of TCGA-COAD samples when k = 2. Subtype A contains 306 samples and subtype B contains 143 samples. (**D**) The consensus score matrix of GSE39582 samples when k = 2. Subtype A contains 382 samples and subtype B contains 203 samples.

To define new immune subtypes of CC based on the coexpressed upregulated genes in the CDM signature, the TCGA-COAD and GSE39582 samples were clustered using the consensus clustering method [34]. The optimal k value was determined using cophenetic correlation coefficients, and the CDF curve had the flattest middle segment when k = 2 (Supplementary Figure 7D, 7E). In addition, the consensus matrix heatmap also maintained clear boundaries at k = 2. Therefore, the optimal number of subtypes was determined to be k = 2, and two new immune subtypes named "subtype A" (n = 306) and "subtype B" (n = 143) in TCGA-COAD samples, and "subtype A" (n = 382) and "subtype B" (n = 203) in GSE39582samples were

identified (Figure 11C, 11D). Subsequently, the heatmap of Supplementary Figure 7F shows the expression of 1,220 coexpressed upregulated genes in immune subtype A and B TCGA-COAD samples. It is clear that the CDM-associated gene expression in subtype A is significantly higher than that in subtype B samples.

New immune subtypes A and B distinguish immune benefit patients

Tumour immune microenvironment analysis revealed that the sample of subtype B exhibited stronger immune cell infiltration than subtype A, in particular GSE39582 cohort (Figure 12A). Furthermore, Figure 12B–12G were
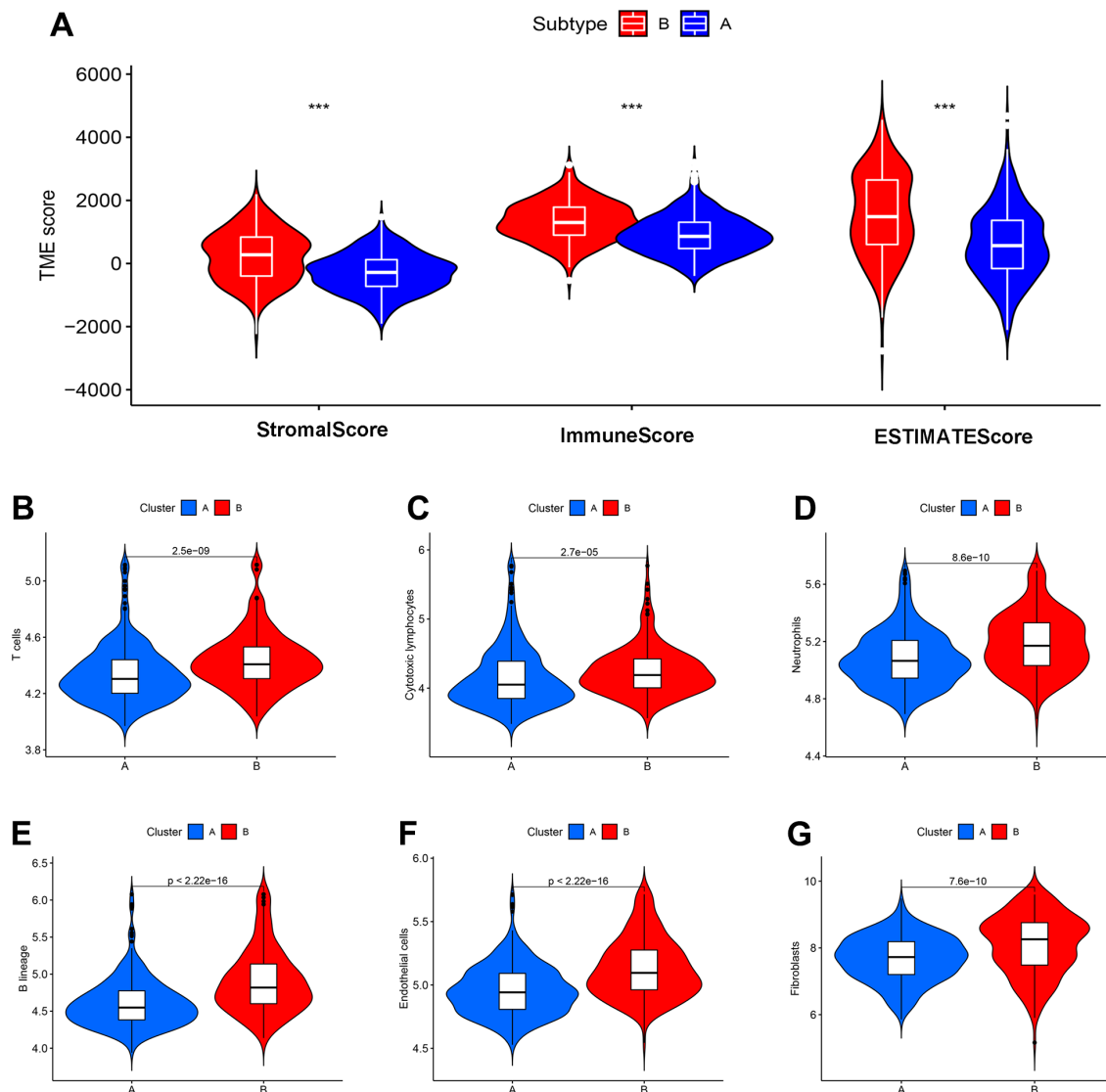

Figure 12. CDM subtypes associated with immunotherapeutic response predictors in GSE39582 datasets. (**A**) Tumor immune microenvironment analysis between CDM immune subtype A and subtype B by Microenvironment Cell Populations counter algorithm. (**B–G**) The association between CDM subtypes and immune cell infiltration; (**B**) for T cells; (**C**) for cytotoxic lymphocytes; (**D**) for neutrophils; (**E**) for B lineages; (**F**) for endothelial cells; (**G**) for fibroblasts.

analyzed the abundance of immune cells from subtypes A and B in GSE39582 samples. In TCGA-COAD samples, tumour immune microenvironment analysis revealed a higher ImmuneScore in subtype B than in subtype A samples, while there was no difference in ImmuneStroma and ESTIMATEScore, indicating that there is a definite link between the two new CDM molecular subtypes and immune cell infiltration (Figure 13A). Based on this, we found that cells with tumour-killing properties, T cells, cytotoxic lymphocytes, neutrophils, and B cells were significantly more abundant in subtype B than in subtype A samples by the MCP-counter immune algorithm; however, the opposite result was found for immunosuppressive cells, endothelial cells and fibroblasts (Figure 13B–13G). According to the article "The Immune Landscape of Cancer" published in the journal Immunity in 2018, an immunogenicity analysis of 33 cancers in TCGA was performed, and all tumours were classified into six immune subtypes, namely, wound healing (Immune C1), IFN-dominant (Immune C2), inflammatory (Immune C3), lymphocyte depleted (Immune C4), immunologically quiet (Immune C5) and TGF-dominant (Immune C6) [35]. As shown in Figure 13H, a Sankey diagram illustrated the relationship between the CDM immune subtypes and the six immune subtypes. Further statistics showed that the proportion of the C2 population with immune benefit in the subtype B samples was significantly higher than that in the subtype A samples, which was consistent with the previous results (Supplementary Figure 7G–7H). With the above analysis of the immune correlation between the two subtypes, patients with subtype B may have a better prognostic benefit from an immunotherapy perspective. Using TCGA-COAD immunotherapy data from TCIA, we analysed the relationship between the two subtypes and the immunotherapy data of PD1 and CTLA4 in the TCGA-COAD samples. The subtype B patients had better treatment outcomes (Figure 13I–13L).

DISCUSSION

The morbidity of CC ranks among the top three in the world representing a very high proportion of all malignant tumours, and posing a serious threat to human life, health, and property security. Although the treatment concept and technical means of CC have been greatly improved, poor patient prognosis management, the occurrence of multidrug resistance and the high recurrence rate of the tumour remain major issues. Personalized patient management and optional multilocus targeted therapy can minimize tumour recurrence and prolong survival time, creating a promising prognosis for patient survival. CRISPR–Cas9 technology, a cornerstone of research for screening functional genes, is an effective method for discovering tumour-dependent genes. CDMs have been developed

to detect tens of thousands of tumour-dependent genes in thousands of tumour cell lines via using CRISPR–Cas9 and are being updated continuously, thus offering an abundant resource and a solid foundation for this study. Based on the basic framework of CDM, Kenichi et al. developed a ShinyDepMap web server (https://labsyspharm.shinyapps.io/depmap) [36] that predicts and evaluates the efficacy of gene-specific drugs, while also retrieving the most sensitive cancer cell lines for the specified drugs. Moreover, Chiu and his colleagues [37] introduced the use of deep learning in CDM, which more quickly and more effectively predicts cancer dependencies. Apart from that, Neekesh V [38] established a first-generation paediatric CDM, that included 13 different forms of paediatric solid and brain tumours.

First, CC dependent gene screening results from the CDM combined with TCGA-COAD differentially expressed genes identified 1,304 overexpressed and CC-dependent oncogenes. Then, the CDM signature with five cancer-dependent genes and prognostic nomogram were constructed by Lasso Cox regression and multivariate Cox analyses, and they demonstrated robust predictive performance and close relationship with clinical characteristics in different datasets. Furthermore, WGCNA and consensus clustering were used to define coexpressed genes with the CDM risk scores and identify two new immune subtypes. Finally, we conducted a comprehensive investigation of the relationship between the two subtypes and immune regulation and the immune microenvironment and the impact of immunotherapy based on the different subtypes.

The CDM signature comprised the *MMS19, NOP14, POLRMT, SNAPC5* and *TIGD1* genes. Statistical analysis revealed that these genes were highly expressed in CC tissues and were tumour-dependent genes from CDMs. Experiments revealed that the TIGD1 genes were oncogenes that impacted the CC cell cycle. Li et al. also [16] discovered that *TIGD1*, an unidentified gene, is substantially expressed in colorectal, gastric, liver, lung, and pancreatic cancers; may be involved in tumour cell cycle regulation; and is negatively correlated with prognosis. Zhu [14] discovered that NOP14 regulates the NRIP1/GSK-3β/β-catenin signalling pathway to promote colorectal cancer proliferation and migration. In pancreatic cancer [39], NOP14 is highly expressed and promotes tumour invasion and metastasis by targeting p53 mutation. In contrast, some studies have reported that NOP14 inhibits melanoma by regulating the Wnt/β-catenin pathway [40] and cancer stemness [41] and inhibits breast cancer by inhibiting the NRIP1/Wnt/β-catenin pathway [42]. In combination with Zhu's study, we found that *NOP14* mRNA is abundantly expressed in
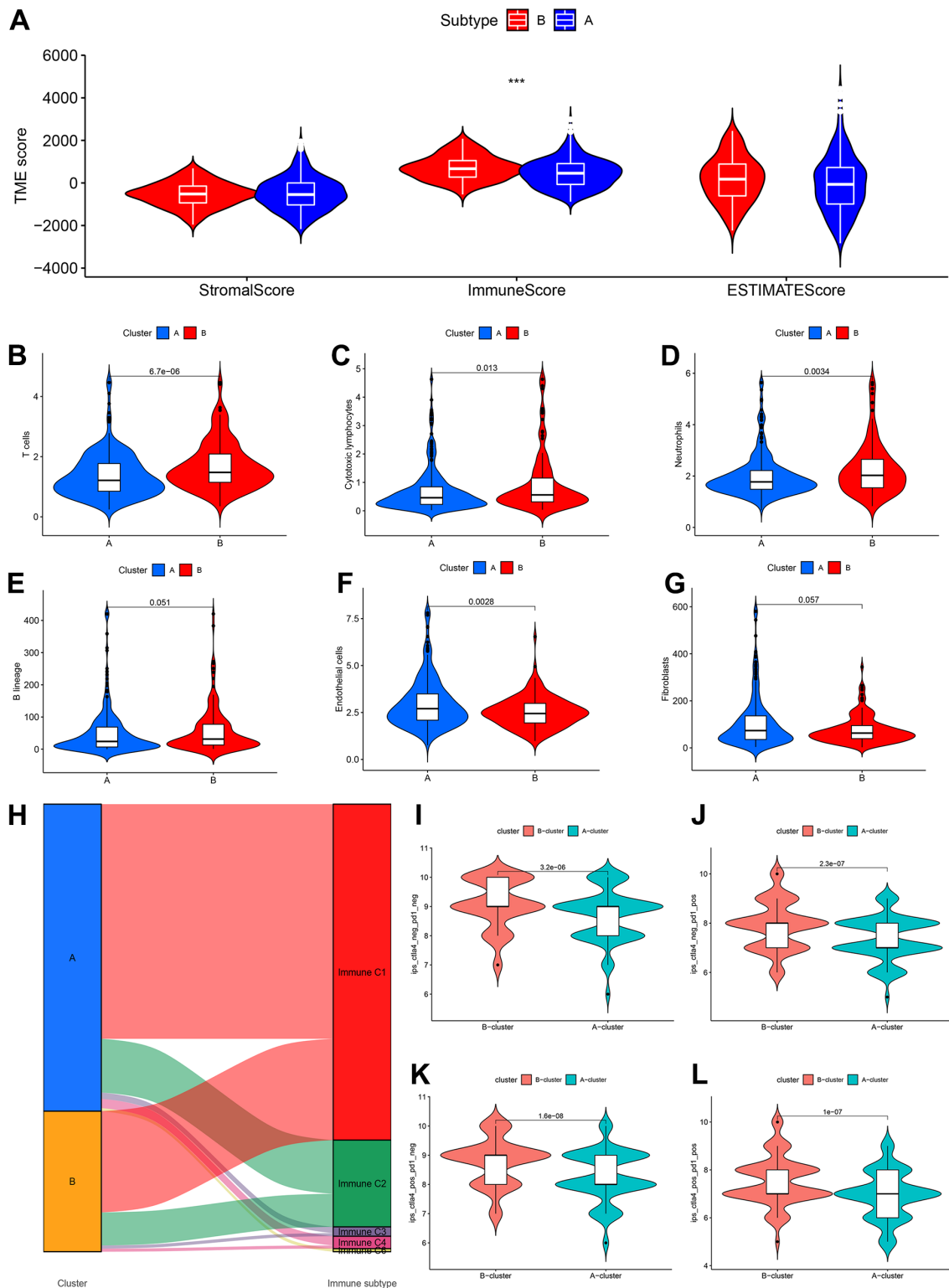
Figure 13. CDM subtypes associated with immunotherapeutic response predictors in TCGA-COAD datasets. (**A**) Tumor immune microenvironment analysis between CDM immune subtype A and subtype B by Microenvironment Cell Populations counter algorithm. (**B**–**G**) The association between CDM subtypes and immune cell infiltration; (**B**) for T cells; (**C**) for cytotoxic lymphocytes; (**D**) for neutrophils; (**E**) for B lineages; (**F**) for endothelial cells; (**G**) for fibroblasts. (**H**) Sankey diagram between the CDM subtypes and immune subtypes (C1-C6). (**I**–**L**) Relationship between the CDM subtypes and the immunotherapy data of PD1 and CTLA4 (**I**) for PD1 and CTLA4 negative expression; (**J**) for PD1 positive expression and CTLA4 negative expression; (**K**) for CTLA4 positive expression and PD1 negative expression; (**L**) for PD1 and CTLA4 negative expression.

CC tissues; however, it is a protective prognostic factor for CC patients, which seems to be a contradiction. Of course, it is unfortunate that the relationship between the expression of NOP14 and CC patient prognosis was not explored in Zhu's study. The existing research on NOP14 in CC is just the tip of the iceberg, and the specific mechanism deserves to be further explored.

We further explored the relationship between the CDM signature and the IC50 value of targeted drugs when elucidating the association between this signature and the prognostic role and clinical characteristics of CC. These targeted drugs are mainly composed of tyrosine kinase inhibitors (AMG.706: motesanib, AP.24534: ponatinib, axitinib, AZD0530, AZD7762, imatinib and nilotinib), a p21-activated kinase inhibitor (IPA3), and a traditional Chinese medicine (shikonin). Drug sensitivity analysis found that patients with low CDM risk scores were sensitive to tyrosine kinase inhibitors. In addition, the previous KEGG and GO analyses revealed that the highly expressed CDM-related genes were mainly involved in the cell proliferation cycle and regulated the proliferation of CC cells. Ho et al. [43] found that AMG.706 exerts an anticancer effect by regulating cell cycle progression and apoptotic cell death. IPA3, a small-molecule p21-activated kinase inhibitor, has excellent antitumour effects by enforcing cell cycle arrest and inducing apoptosis [44, 45]. Shikonin, a naphthoquinone compound extracted from Lithospermum, has been extensively studied for its antitumour activity. In CC, shikonin can arrest the cell cycle by inhibiting the hypoxia-inducible factor signalling pathway [46], promote cell apoptosis by downregulating the PI3K/Akt signalling pathway [47], and inhibit growth by regulating the STAT3 signalling pathway [48]. Therefore, we hypothesized that the tumour cells of patients with high CDM risk have self-regulated cell cycle activity to resist the tumour-killing effects of these drugs, especially tyrosine kinase inhibitors.

As Volker reported, vioprolide A, a natural product, has an outstanding therapeutic effect for acute lymphoblastic leukaemia by targeting NOP14 expression to affect ribosome biogenesis [49]. The results of the KEGG pathway and GO functional analyses are consistent with the conclusions drawn from Volker's study. *POLRMT*, a core mitochondrial gene, is responsible for mitochondrial metabolism, and inhibition of POLRMT expression can be utilized to induce an energy crisis from multidrug resistant cells and thus resensitize them to chemotherapeutic agents [50]. In a tumour cell reconstituted recombinant system, the inhibitor of mitochondrial transcription (IMTs), which specifically targets *POLRMT*, significantly hampers mitochondrial

DNA (mtDNA) transcription and triggers the dose-dependent inhibition of mtDNA expression and the oxidative phosphorylation system [17]. According to Eliseo's report [51] based on supervised support vector machine learning, the high expression of TIGD1, to a certain extent, affects the sensitivity of tumour cells to cisplatin and improves patient outcomes after chemotherapy. The drugs shown in Figure 6 have good clinical potential for patients at high risk according to the CDM signature, especially patients with advanced multidrug resistant disease, according to the CDM signature for drug resistance analysis. A recent study found that knockdown of TIGD1 was effective in accelerating copper death in colorectal cancer cells induced by the copper death inducer elesclomol, and thus TIGD1 could serve as a key protein in regulating copper-dependent cell death [52]. Qiao et al. [53] found that TIGD1, which was significantly highly expressed in oral squamous cell carcinoma, effectively promoted the proliferation and invasive metastasis of cancer cells and was also involved in immune cell infiltration in tumor tissues, thus promoting the progression of oral squamous cell carcinoma.

Based on the CDM-related coexpression of differential genes, the consensus clustering divided the TCGA-COAD and GSE39582 samples into two new subtypes A and B. In the overall analysis, immune cell scores were significantly higher in the B molecular subtype than the A subtype, and tumour-killing immune cells, such as T cells, B cells, cytotoxic lymphocytes, and neutrophils, were also highly infiltrated in the B molecular subtype samples. From another perspective, all samples with Immune C6 were subtype A. In addition, in terms of PD1 and CTLA4 immune therapy benefit, subtype A did not perform as well as subtype B. These results also reflect the fact that subtype A is associated with a poorer immune benefit than the subtype B.

Although this study is founded on a significant amount of TCGA-COAD, GSE17536 and GSE39582 data, there are still some limitations. The present study is based on a retrospective analysis of a public database of CC samples, which lacks prospective clinical application and validation. Thus, more research is necessary to comprehensively guide CC clinical diagnosis and treatment.

AUTHOR CONTRIBUTIONS

Guoyang Zhang and Zongfeng Feng drafted the manuscript. Zongfeng Feng and Qingwen Zeng performed the experiments. Qingwen Zeng analyzed data. Ping Huang designed the study and revised the manuscript.

REFERENCES

1. Sethi NS, Kikuchi O, Duronio GN, Stachler MD, McFarland JM, Ferrer-Luna R, Zhang Y, Bao C, Bronson R, Patil D, Sanchez-Vega F, Liu JB, Sicinska E, et al. Early TP53 alterations engage environmental exposures to promote gastric premalignancy in an integrative mouse model. Nat Genet. 2020; 52:219–30.
https://doi.org/10.1038/s41588-019-0574-9
PMID:32025000

2. Barua S, Junaid MA. Lifestyle, pregnancy and epigenetic effects. Epigenomics. 2015; 7:85–102.
https://doi.org/10.2217/epi.14.71 PMID:25687469

3. Sakthivel KM, Hariharan S. Regulatory players of DNA damage repair mechanisms: Role in Cancer Chemoresistance. Biomed Pharmacother. 2017; 93:1238–45.
https://doi.org/10.1016/j.biopha.2017.07.035
PMID:28738540

4. Martínez-Rodríguez OP, Thompson-Bonilla MDR, Jaramillo-Flores ME. Association between obesity and breast cancer: Molecular bases and the effect of flavonoids in signaling pathways. Crit Rev Food Sci Nutr. 2020; 60:3770–92.
https://doi.org/10.1080/10408398.2019.1708262
PMID:31899947

5. Rodina A, Wang T, Yan P, Gomes ED, Dunphy MP, Pillarsetty N, Koren J, Gerecitano JF, Taldone T, Zong H, Caldas-Lopes E, Alpaugh M, Corben A, et al. The epichaperome is an integrated chaperome network that facilitates tumour survival. Nature. 2016; 538:397–401.
https://doi.org/10.1038/nature19807
PMID:27706135

6. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, Cercek A, Smith RA, Jemal A. Colorectal cancer statistics, 2020. CA Cancer J Clin. 2020; 70:145–64.
https://doi.org/10.3322/caac.21601 PMID:32133645

7. Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, Santos R, Rao Y, Sassi F, Pinnelli M, Ansari R, Harper S, Jackson DA, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature. 2019; 568:511–6.
https://doi.org/10.1038/s41586-019-1103-9
PMID:30971826

8. Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, Gill S, Harrington WF, Pantel S, Krill-Burger JM, Meyers RM, Ali L, Goodale A, et al. Defining a Cancer Dependency Map. Cell. 2017; 170:564–76.e16.
https://doi.org/10.1016/j.cell.2017.06.010
PMID:28753430

9. Smith I, Greenside PG, Natoli T, Lahr DL, Wadden D, Tirosh I, Narayan R, Root DE, Golub TR, Subramanian A, Doench JG. Evaluation of RNAi and CRISPR technologies by large-scale gene expression profiling in the Connectivity Map. PLoS Biol. 2017; 15:e2003213.
https://doi.org/10.1371/journal.pbio.2003213
PMID:29190685

10. Zhang JL, Wang HY, Yang Q, Lin SY, Luo GY, Zhang R, Xu GL. Methyl-methanesulfonate sensitivity 19 expression is associated with metastasis and chemoradiotherapy response in esophageal cancer. World J Gastroenterol. 2015; 21:4240–7.
https://doi.org/10.3748/wjg.v21.i14.4240
PMID:25892874

11. An M, Zang X, Wang J, Kang J, Tan X, Fu B. Comprehensive analysis of differentially expressed long noncoding RNAs, miRNAs and mRNAs in breast cancer brain metastasis. Epigenomics. 2021; 13:1113–28.
https://doi.org/10.2217/epi-2021-0152
PMID:34148372

12. Stehling O, Vashisht AA, Mascarenhas J, Jonsson ZO, Sharma T, Netz DJ, Pierik AJ, Wohlschlegel JA, Lill R. MMS19 assembles iron-sulfur proteins required for DNA metabolism and genomic integrity. Science. 2012; 337:195–9.
https://doi.org/10.1126/science.1219723
PMID:22678362

13. Papatriantafyllou M. DNA Metabolism: MMS19: CIA agent for DNA-linked affairs. Nat Rev Mol Cell Biol. 2012; 13:538.
https://doi.org/10.1038/nrm3411

PMID:22828929

14. Zhu X, Jia W, Yan Y, Huang Y, Wang B. NOP14 regulates the growth, migration, and invasion of colorectal cancer cells by modulating the NRIP1/GSK-3β/β-catenin signaling pathway. Eur J Histochem. 2021; 65:3246.
https://doi.org/10.4081/ejh.2021.3246
PMID:34218653

15. Ying Y, Li J, Xie H, Yan H, Jin K, He L, Ma X, Wu J, Xu X, Fang J, Wang X, Zheng X, Liu B, Xie L. CCND1, NOP14 and DNMT3B are involved in miR-502-5p-mediated inhibition of cell migration and proliferation in bladder cancer. Cell Prolif. 2020; 53:e12751.
https://doi.org/10.1111/cpr.12751
PMID:31971654

16. Yin L, Yan J, Wang Y, Sun Q. TIGD1, a gene of unknown function, involves cell-cycle progression and correlates with poor prognosis in human cancer. J Cell Biochem. 2019; 120:9758–67.
https://doi.org/10.1002/jcb.28256
PMID:30548305

17. Bonekamp NA, Peter B, Hillen HS, Felser A, Bergbrede T, Choidas A, Horn M, Unger A, Di Lucrezia R, Atanassov I, Li X, Koch U, Menninger S, et al. Small-molecule inhibitors of human mitochondrial DNA transcription. Nature. 2020; 588:712–6.
https://doi.org/10.1038/s41586-020-03048-z
PMID:33328633

18. Henry RW, Mittal V, Ma B, Kobayashi R, Hernandez N. SNAP19 mediates the assembly of a functional core promoter complex (SNAPc) shared by RNA polymerases II and III. Genes Dev. 1998; 12:2664–72.
https://doi.org/10.1101/gad.12.17.2664
PMID:9732265

19. Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, Dharia NV, Montgomery PG, Cowley GS, Pantel S, Goodale A, Lee Y, Ali LD, et al. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nat Genet. 2017; 49:1779–84.
https://doi.org/10.1038/ng.3984
PMID:29083409

20. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015; 347:1260419.
https://doi.org/10.1126/science.1260419
PMID:25613900

21. Feng Z, Li L, Zeng Q, Zhang Y, Tu Y, Chen W, Shu X, Wu A, Xiong J, Cao Y, Li Z. *RNF114* Silencing Inhibits the Proliferation and Metastasis of Gastric Cancer. J

Cancer. 2022; 13:565–78.
https://doi.org/10.7150/jca.62033 PMID:35069903

22. Geeleher 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

23. Shen W, Song Z, Zhong X, Huang M, Shen D, Gao P, Qian X, Wang M, He X, Wang T, Li S, Song X. Sangerbox: A comprehensive, interaction-friendly clinical bioinformatics analysis platform. iMeta. 2022; 1:e36.
https://doi.org/10.1002/imt2.36

24. Mizuno H, Kitada K, Nakai K, Sarai A. PrognoScan: a new database for meta-analysis of the prognostic value of genes. BMC Med Genomics. 2009; 2:18.
https://doi.org/10.1186/1755-8794-2-18
PMID:19393097

25. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautès-Fridman C, Fridman WH, de Reyniès A. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. 2016; 17:218.
https://doi.org/10.1186/s13059-016-1070-5
PMID:27765066

26. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep. 2017; 18:248–62.
https://doi.org/10.1016/j.celrep.2016.12.019
PMID:28052254

27. Li X, Wen D, Li X, Yao C, Chong W, Chen H. Identification of an Immune Signature Predicting Prognosis Risk and Lymphocyte Infiltration in Colon Cancer. Front Immunol. 2020; 11:1678.
https://doi.org/10.3389/fimmu.2020.01678
PMID:33013820

28. Nie J, Shan D, Li S, Zhang S, Zi X, Xing F, Shi J, Liu C, Wang T, Sun X, Zhang Q, Zhou M, Luo S, et al. A Novel Ferroptosis Related Gene Signature for Prognosis Prediction in Patients With Colon Cancer. Front Oncol. 2021; 11:654076.
https://doi.org/10.3389/fonc.2021.654076
PMID:34046350

29. Ren J, Feng J, Song W, Wang C, Ge Y, Fu T. Development and validation of a metabolic gene signature for predicting overall survival in patients with colon cancer. Clin Exp Med. 2020; 20:535–44.
https://doi.org/10.1007/s10238-020-00652-1

PMID:32772211

30. Xu J, Dai S, Yuan Y, Xiao Q, Ding K. A Prognostic Model for Colon Cancer Patients Based on Eight Signature Autophagy Genes. Front Cell Dev Biol. 2020; 8:602174.
https://doi.org/10.3389/fcell.2020.602174
PMID:33324651

31. Akce M, Zakka K, Jiang R, Williamson S, Alese OB, Shaib WL, Wu C, Behera M, El-Rayes BF. Impact of Tumor Side on Clinical Outcomes in Stage II and III Colon Cancer With Known Microsatellite Instability Status. Front Oncol. 2021; 11:592351.
https://doi.org/10.3389/fonc.2021.592351
PMID:33859934

32. Oh BY, Huh JW, Park YA, Cho YB, Yun SH, Kim HC, Lee WY, Chun HK. Prognostic factors in sporadic colon cancer with high-level microsatellite instability. Surgery. 2016; 159:1372–81.
https://doi.org/10.1016/j.surg.2015.11.028
PMID:26775578

33. Ribic CM, Sargent DJ, Moore MJ, Thibodeau SN, French AJ, Goldberg RM, Hamilton SR, Laurent-Puig P, Gryfe R, Shepherd LE, Tu D, Redston M, Gallinger S. Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. N Engl J Med. 2003; 349:247–57.
https://doi.org/10.1056/NEJMoa022289
PMID:12867608

34. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010; 26:1572–3.
https://doi.org/10.1093/bioinformatics/btq170
PMID:20427518

35. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA, Ziv E, Culhane AC, Paull EO, et al, and Cancer Genome Atlas Research Network. The Immune Landscape of Cancer. Immunity. 2018; 48:812–30.e14.
https://doi.org/10.1016/j.immuni.2018.03.023
PMID:29628290

36. Shimada K, Bachman JA, Muhlich JL, Mitchison TJ. shinyDepMap, a tool to identify targetable cancer genes and their functional connections from Cancer Dependency Map data. Elife. 2021; 10:e57116.
https://doi.org/10.7554/eLife.57116
PMID:33554860

37. Chiu YC, Zheng S, Wang LJ, Iskra BS, Rao MK, Houghton PJ, Huang Y, Chen Y. Predicting and characterizing a cancer dependency map of tumors with deep learning. Sci Adv. 2021; 7:eabh1275.
https://doi.org/10.1126/sciadv.abh1275
PMID:34417181

38. Dharia NV, Kugener G, Guenther LM, Malone CF,

Durbin AD, Hong AL, Howard TP, Bandopadhayay P, Wechsler CS, Fung I, Warren AC, Dempster JM, Krill-Burger JM, et al. A first-generation pediatric cancer dependency map. Nat Genet. 2021; 53:529–38.
https://doi.org/10.1038/s41588-021-00819-w
PMID:33753930

39. Du Y, Liu Z, You L, Hou P, Ren X, Jiao T, Zhao W, Li Z, Shu H, Liu C, Zhao Y. Pancreatic Cancer Progression Relies upon Mutant p53-Induced Oncogenic Signaling Mediated by NOP14. Cancer Res. 2017; 77:2661–73.
https://doi.org/10.1158/0008-5472.CAN-16-2339
PMID:28280038

40. Li J, Zhao R, Guo Y, Fang R. EMG1 interacts with NOP14 to regulate the growth, migration, and invasion of melanoma cells via the Wnt/β-catenin pathway. Transl Cancer Res. 2020; 9:3669–79.
https://doi.org/10.21037/tcr.2020.03.79
PMID:35117729

41. Li J, Fang R, Wu J, Si Y, Bai J, Wang Q. The NOP14 nucleolar protein suppresses the function and stemness of melanoma stem-like cells through Wnt/beta-catenin signaling inactivation. Bioengineered. 2022; 13:7648–58.
https://doi.org/10.1080/21655979.2022.2050491
PMID:35282769

42. Lei JJ, Peng RJ, Kuang BH, Yuan ZY, Qin T, Liu WS, Guo YM, Han HQ, Lian YF, Deng CC, Zhang HJ, Chen LZ, Feng QS, et al. NOP14 suppresses breast cancer progression by inhibiting NRIP1/Wnt/β-catenin pathway. Oncotarget. 2015; 6:25701–14.
https://doi.org/10.18632/oncotarget.4573
PMID:26213846

43. Ho JN, Byun SS, Lee SE, Youn JI, Lee S. Multikinase inhibitor motesanib enhances the antitumor effect of cisplatin in cisplatin-resistant human bladder cancer cells via apoptosis and the PI3K/Akt pathway. Oncol Rep. 2019; 41:2482–90.
https://doi.org/10.3892/or.2019.7005 PMID:30816494

44. Dukel M, Fiskin K. Combination of PAKs inhibitors IPA-3 and PF-3758309 effectively suppresses colon carcinoma cell growth by perturbing DNA damage response. Int J Radiat Biol. 2023; 99:340–54.
https://doi.org/10.1080/09553002.2022.2110326
PMID:35939342

45. Wang Y, Li W, Zhang Q, Gu X, He X, Men Y, Zhang C. Targeting Phosphorylation of p21-activated Kinase 1 at Thr423 Induces Cell Cycle Arrest and Apoptosis in Cutaneous T-cell Lymphoma Cells. Acta Derm Venereol. 2019; 99:1022–8.
https://doi.org/10.2340/00015555-3263
PMID:31304555

46. Li MY, Mi C, Wang KS, Wang Z, Zuo HX, Piao LX, Xu GH,

Li X, Ma J, Jin X. Shikonin suppresses proliferation and induces cell cycle arrest through the inhibition of hypoxia-inducible factor-1α signaling. Chem Biol Interact. 2017; 274:58–67.
https://doi.org/10.1016/j.cbi.2017.06.029
PMID:28684144

47. Zhu Y, Zhong Y, Long X, Zhu Z, Zhou Y, Ye H, Zeng X, Zheng X. Deoxyshikonin isolated from *Arnebia euchroma* inhibits colorectal cancer by down-regulating the PI3K/Akt/mTOR pathway. Pharm Biol. 2019; 57:412–23.
https://doi.org/10.1080/13880209.2019.1626447
PMID:31230505

48. Shi W, Men L, Pi X, Jiang T, Peng D, Huo S, Luo P, Wang M, Guo J, Jiang Y, Peng L, Lin L, Li S, Lv J. Shikonin suppresses colon cancer cell growth and exerts synergistic effects by regulating ADAM17 and the IL-6/STAT3 signaling pathway. Int J Oncol. 2021; 59:99.
https://doi.org/10.3892/ijo.2021.5279 PMID:34726248

49. Kirsch VC, Orgler C, Braig S, Jeremias I, Auerbach D, Müller R, Vollmar AM, Sieber SA. The Cytotoxic Natural Product Vioprolide A Targets Nucleolar Protein 14, Which Is Essential for Ribosome Biogenesis. Angew Chem Int Ed Engl. 2020; 59:1595–600.
https://doi.org/10.1002/anie.201911158
PMID:31658409

50. Yu HJ, Xiao GL, Zhao YY, Wang XX, Lan R. Targeting Mitochondrial Metabolism and RNA Polymerase POLRMT to Overcome Multidrug Resistance in Cancer. Front Chem. 2021; 9:775226.
https://doi.org/10.3389/fchem.2021.775226
PMID:34976949

51. Mucaki EJ, Zhao JZL, Lizotte DJ, Rogan PK. Predicting responses to platin chemotherapy agents with biochemically-inspired machine learning. Signal Transduct Target Ther. 2019; 4:1.
https://doi.org/10.1038/s41392-018-0034-5
PMID:30652029

52. Wu Z, Lin C, Zhang F, Lu Z, Wang Y, Liu Y, Zhou Z, Li L, Song L. TIGD1 Function as a Potential Cuproptosis Regulator Following a Novel Cuproptosis-Related Gene Risk Signature in Colorectal Cancer. Cancers (Basel). 2023; 15:2286.
https://doi.org/10.3390/cancers15082286
PMID:37190215

53. Qiao X, Zhu L, Song R, Shang C, Guo Y. A novel oncogene trigger transposable element derived-1 promotes oral squamous cell carcinoma progression via evoking immune inhibition. Mol Carcinog. 2023. [Epub ahead of print].
https://doi.org/10.1002/mc.23557
PMID:37144838
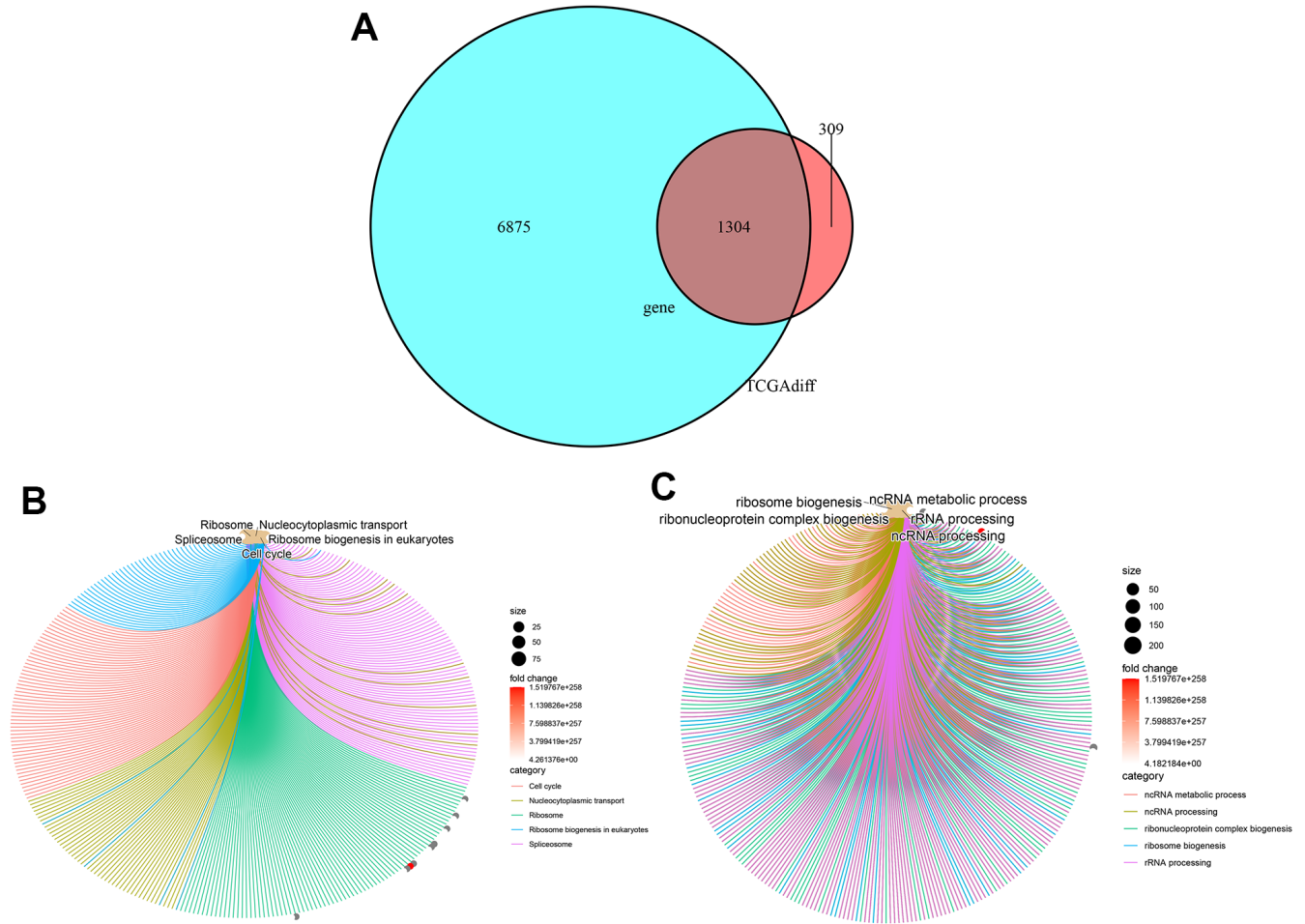
Supplementary Figure 1. Venn and circle graph of KEGG and GO in differentially expressed CDM genes. (**A**) Venn plot. 1,304 differentially expressed CDM genes between differentially expressed genes of TCGA-COAD and CDM genes. Cyan shows 8,179 differentially expressed genes in TCGA-COAD, dull red shows 1,613 cancer-dependent CDM-associated genes and overlapping area shows the 1,304 differentially expressed CDM genes. (**B**) Circle plot. KEGG pathway enrichment of 1304 differential expression of CDM genes. (**C**) Circle plot. GO Biological process analysis of 1304 differential expression of CDM genes.
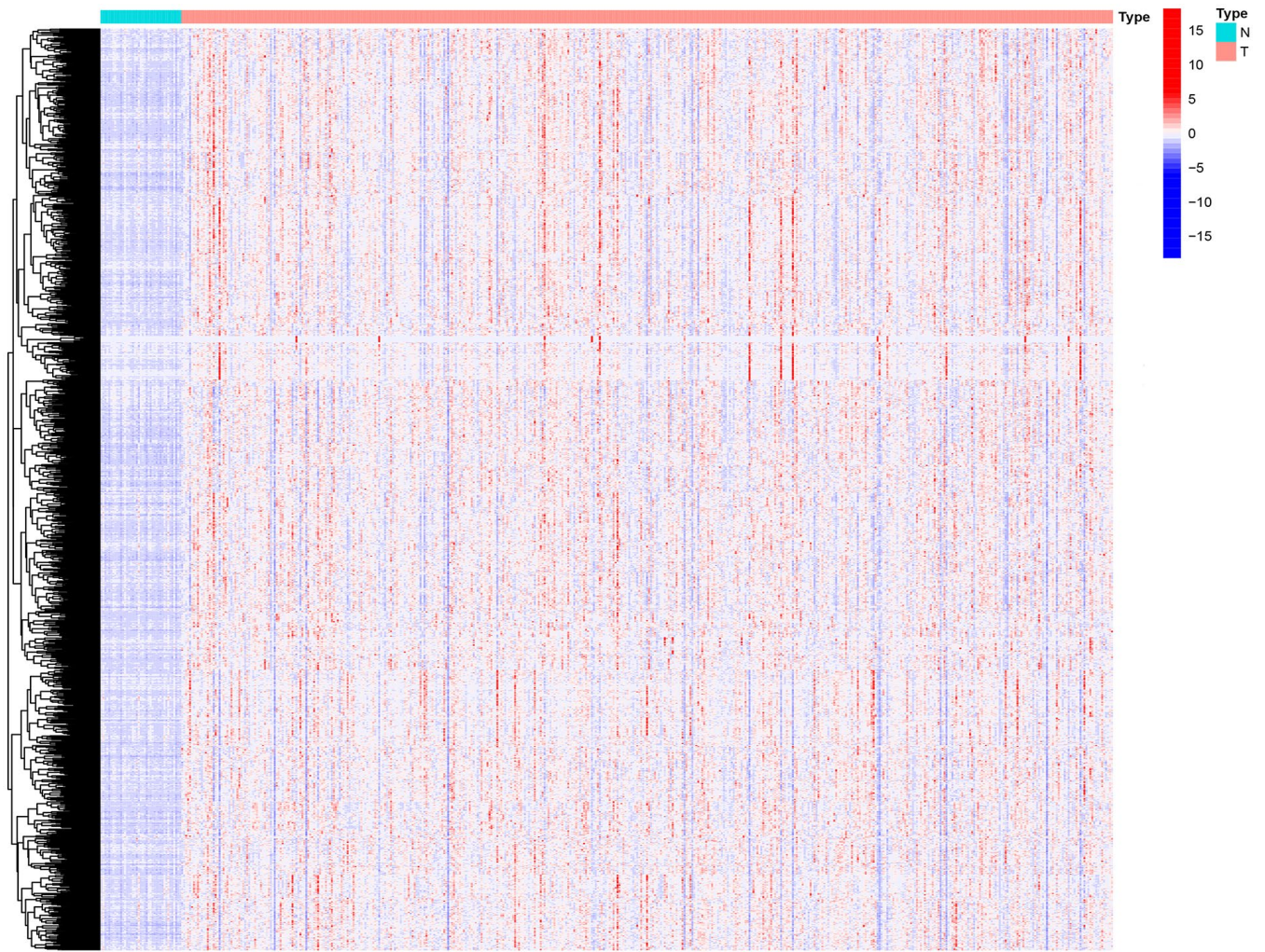
Supplementary Figure 2. Heatmap 1,304 differentially expressed CDM genes between differentially expressed genes of TCGA-COAD.
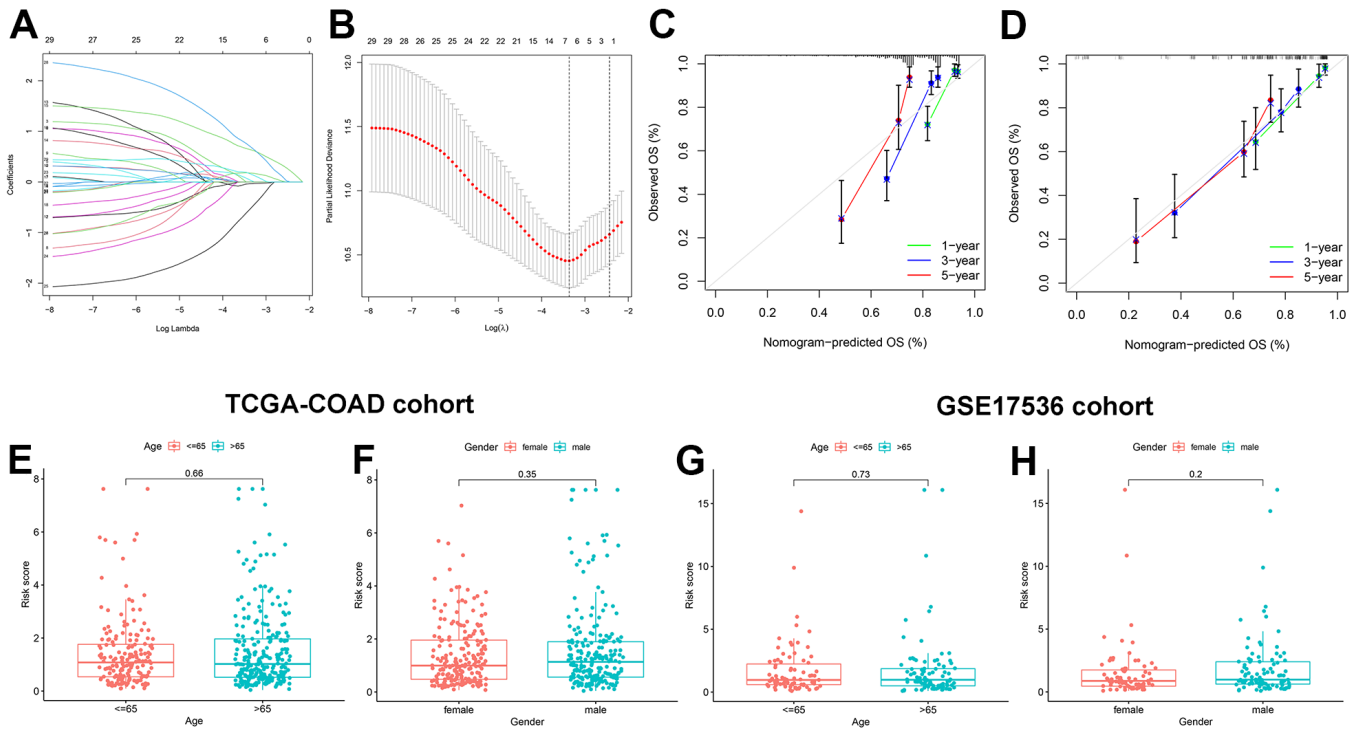
Supplementary Figure 3. Least absolute shrinkage and selection operator (LASSO) Cox regression analysis and nomogram calibration. (**A**, **B**) LASSO Cox regression analysis in TCGA-COAD cohort. (**A**) LASSO coefficient profiles of the 29 prognosis-correlated CDM genes. (**B**) Genes from univariate Cox regression analysis were narrowed down by the lasso algorithm, The left vertical line represents the 5 CDM genes finally identified. (**C**, **D**) The calibration plot of the CDM nomogram for predicting 1-year(green), 3-year(blue) and 5-year(red) survival rates; (**C**) for TCGA-COAD cohort, (**D**) for GSE17536. (**E**) Boxplot. Correlation between age and risk-CDM score in TCGA-COAD cohort. Pink represents age less than or equal to 65 years old and cyan represents age greater than 65 years old (*p*=0.66). (**F**) Boxplot. Correlation between gender and risk-CDM score in TCGA-COAD cohort. Pink represents female and cyan represents male (*p*=0.35). (**G**) Boxplot. Correlation between age and risk-CDM score in GSE17536 cohort. Pink represents age less than or equal to 65 years old and cyan represents age greater than 65 years old (*p*=0.73). (**H**) Boxplot. Correlation between gender and risk-CDM score in GSE17536 cohort. Pink represents female and cyan represents male (*p*=0.2).
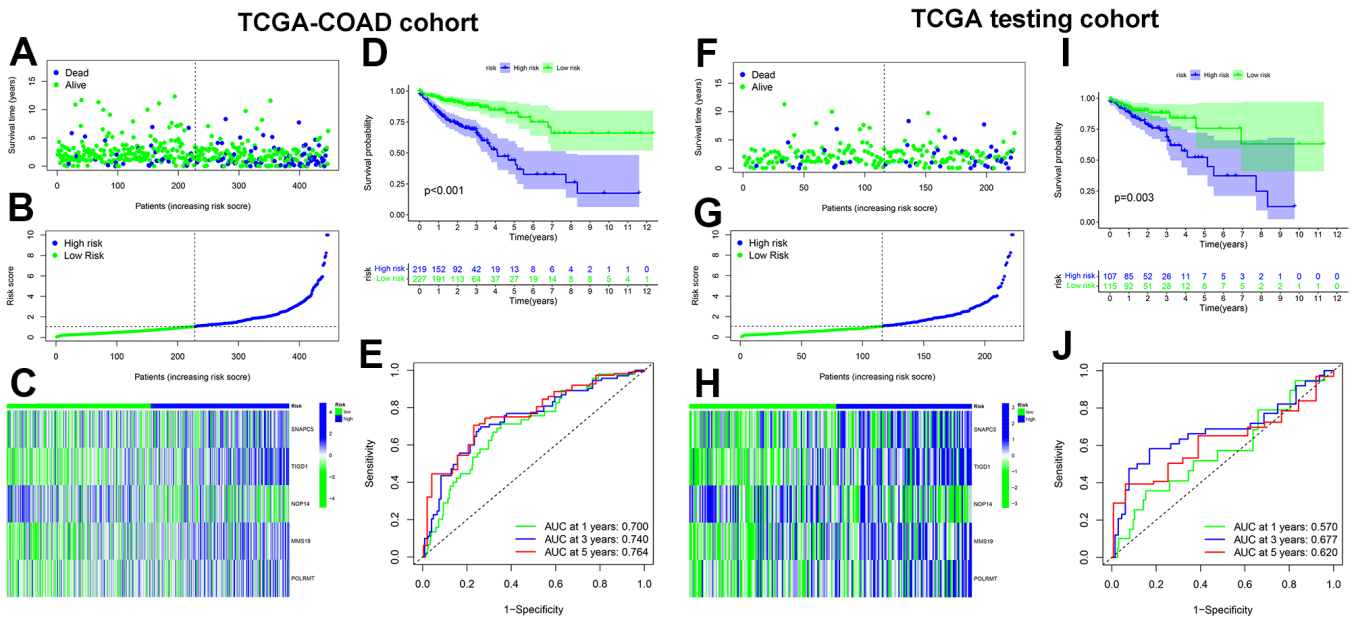
Supplementary Figure 4. Cancer Dependency Maps (CDMs) signature construction and validation in all TCGA-COAD cohort and TCGA testing cohort. (**A**) Survival scatter plot in all TCGA-COAD cohort. A dot represents a CC patient (blue represents Dead, Green represents alive). Dotted lines show the median of risk score that dichotomize patients into high and low groups. (**B**) Risk score plot in all TCGA-COAD cohort. The CC patients are ordered by the risk score of the CDM signature. Dotted lines show the median of risk score that dichotomize patients into high (blue) and low (green) groups. (**C**) Heatmap of five CDM signature genes (*SNAPC5*, *TIGD1*, *NOP14*, *MMS19* and *POLRMT*) in all TCGA-COAD cohort. Blue represents high-CDM risk group and green represents low-CDM risk group. (**D**) Kaplan-Meier plot in all TCGA –COAD cohort (p<0.001). Survival curves for high (blue) and low (green) risk-CDM groups dichotomized at the median of risk-CDM score are plotted. 95% confidence intervals for each group are also indicated by shadow area. (**E**) The ROC curve of CDM signature in all TCGA-COAD cohort. Green show one year AUC value (AUC value=0.700), blue show three years AUC value (AUC value=0.740) and red show five years AUC value (AUC value=0.764). (**F**) Survival scatter plot in TCGA testing cohort. (**G**) Risk score plot in TCGA testing cohort. (**H**) Heatmap of five CDM signature genes (*SNAPC5*, *TIGD1*, *NOP14*, *MMS19* and *POLRMT*) in TCGA testing cohort. (**I**) Kaplan-Meier plot in TCGA testing cohort (*p*=0.003). (**J**) The ROC curve of CDM signature in TCGA testing cohort. Green show one year AUC value (AUC value=0.570), blue show three years AUC value (AUC value=0.670) and red show five years AUC value (AUC value=0.620).
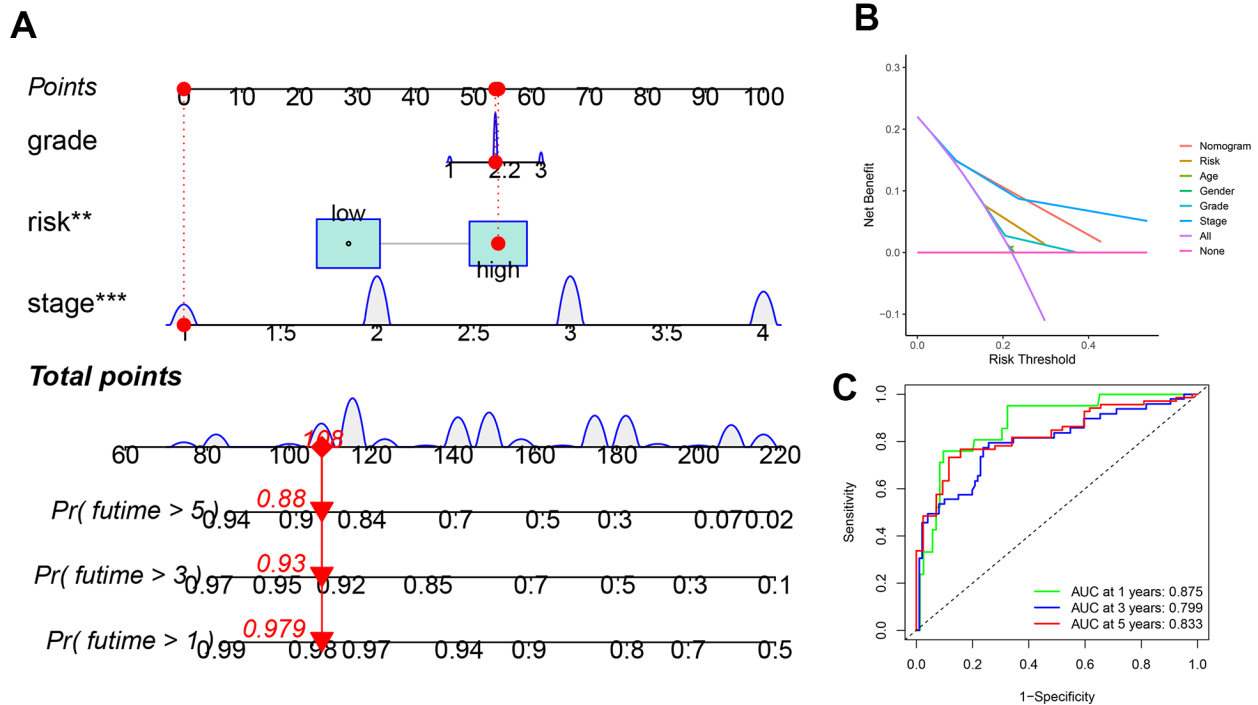
Supplementary Figure 5. CDM-associated nomogram validation in GSE17536 cohort. (**A**) CDM-associated Nomogram for predicting 1-, 3-, and 5-year OS of patients in GSE17536 cohort. (**B**) DCA analysis for CDM-associated nomogram in GSE17536 cohort. (**C**) The ROC curve of CDM-associated Nomogram in GSE17536 cohort. Green show one year AUC value (AUC value=0.875), blue show three years AUC value (AUC value=0.799) and red show five years AUC value (AUC value=0.833).
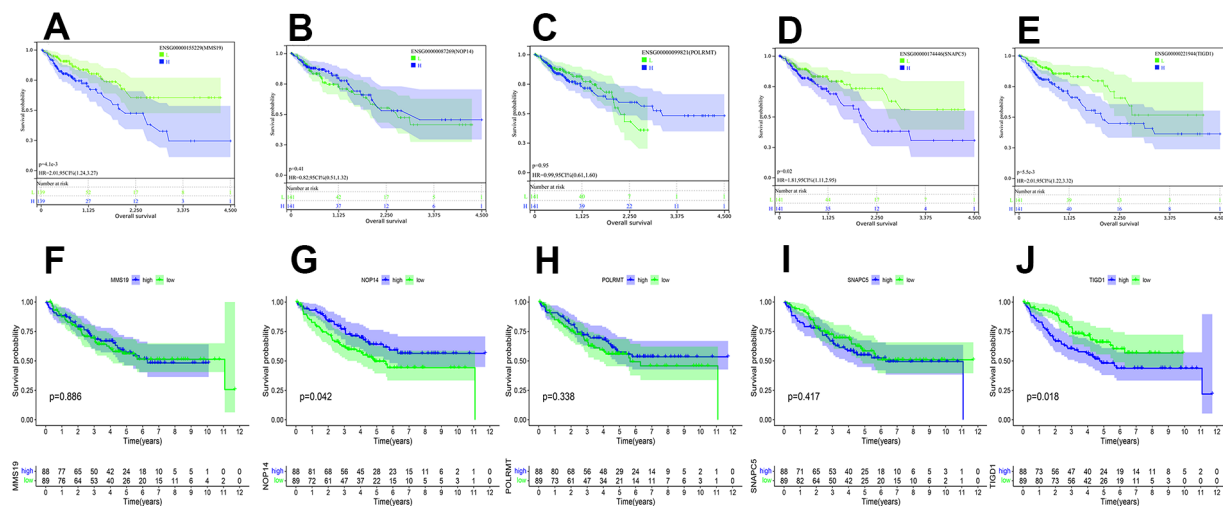
Supplementary Figure 6. Kaplan-Meier plot for five CDM signature genes (MMS19, NOP14, POLRMT, SNAPC5 and TIGD1) in all TCGA–COAD and GSE17536 cohort (p<0.001). Survival curves for high (blue) and low (green) expression groups dichotomized at the median expression are plotted. 95% confidence intervals for each group are also indicated by shadow area. (**A**) MMS19 in TCGA–COAD (p=4.1e3). (**B**) NOP14 in TCGA–COAD (p=0.41). (**C**) POLRMT in TCGA–COAD (p=0.95). (**D**) SNAPC5 in TCGA–COAD (p=0.02). (**E**) TIGD1 in TCGA–COAD (p=5.5e3). (**F**) MMS19 in GSE17536 cohort (p=0.886). (**G**) NOP14 in GSE17536 cohort (p=0.042). (**H**) POLRMT in GSE17536 cohort (p=0.338). (**I**) SNAPC5 in GSE17536 cohort (p=0.417). (**J**) TIGD1 in GSE17536 cohort (p=0.018).
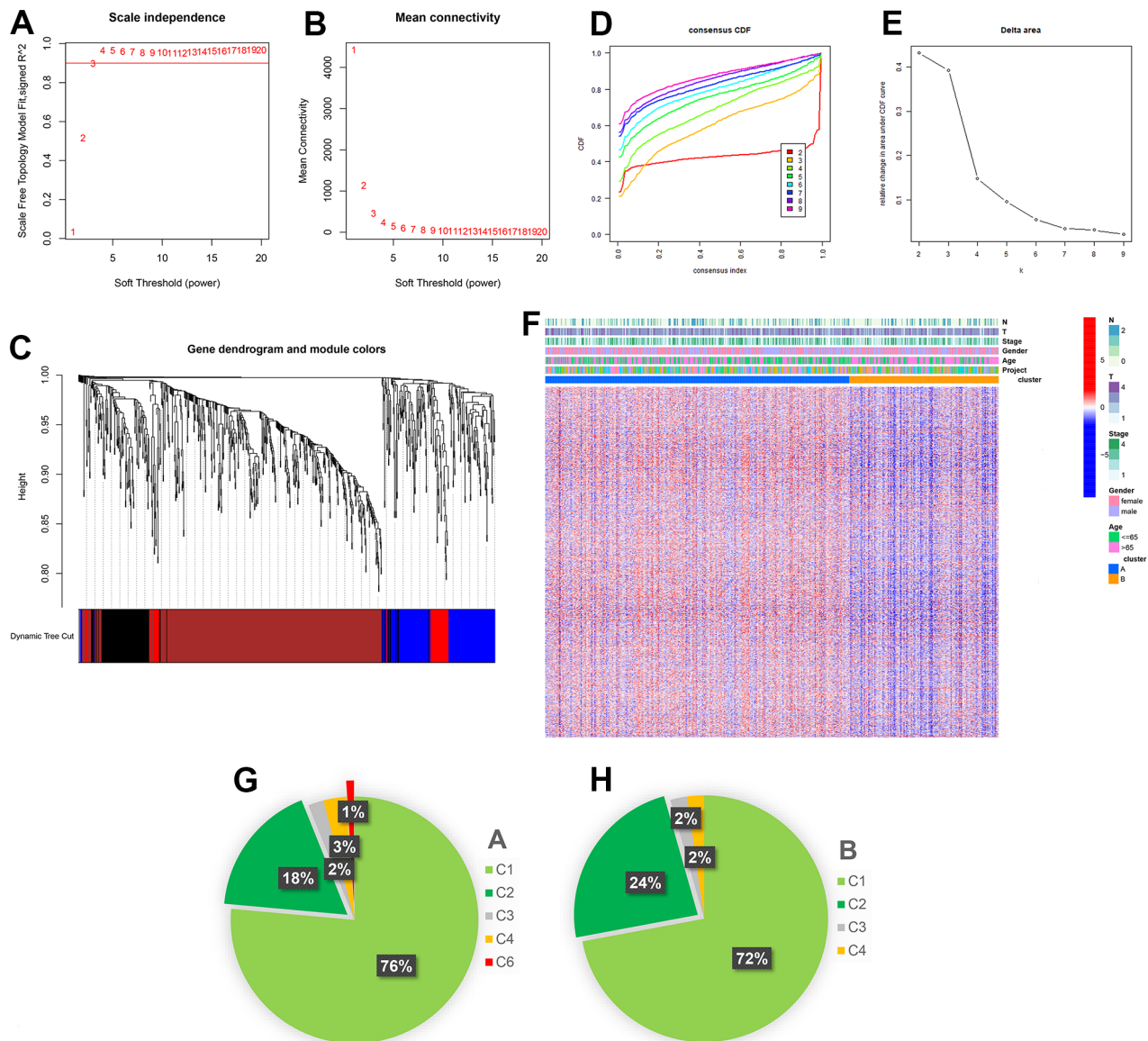
Supplementary Figure 7. WGCNA analysis, consensus clustering and heatmap of two immune subtypes. (**A**, **B**) The scale-free fit index for soft-thresholding powers. (**A**) The relationship between the soft-threshold and scale-free R2. (**B**) The relationship between the soft-threshold and mean connectivity. (**C**) Different modules are labeled in different colors and dendrogram of all genes clustered in TCGA-COAD. (**D**, **E**) Identification of CDM molecular subtypes; (**D**) The cumulative distribution function curves of consensus scores based on different subtype numbers (k = 2~9); (**E**) The comparison of the relative changes of the area under the CDF curve between k and k−1. (**F**) The heatmap of 1220 differential expression of CDM co-expressed genes in TCGA-COAD Subtype A and Subtype B. (**G**, **H**) The proportion of C1-C6 six immune subtypes in subtype A and subtype B; (**G**) for subtype A; (**H**) for subtype B.

Supplementary Table

Supplementary Table 1. 29 prognostic related genes were discovered by univariate Cox hazard analysis in COAD.

id	HR	HR.95L	HR.95H	pvalue
PSMA5	0.589785	0.372309	0.934294	0.024479
SF3A2	1.555346	1.004943	2.407203	0.047469
USP5	1.724162	1.004683	2.958878	0.048051
POLR2H	1.908472	1.137331	3.202469	0.014396
SNAPC4	1.809706	1.171914	2.794604	0.007461
ELL	1.855009	1.063099	3.236817	0.029601
PSMD12	0.640136	0.41484	0.987789	0.043855
SUPT5H	1.73605	1.033636	2.915795	0.037068
EIF2B5	2.217337	1.131043	4.346947	0.020423
BRF1	1.851907	1.008635	3.400201	0.046845
SNAPC5	1.773781	1.111222	2.831388	0.016308
CDK9	1.944925	1.150271	3.288557	0.013051
SART1	1.661597	1.052816	2.622399	0.029181
PNN	1.463655	1.000863	2.14044	0.049483
TIGD1	1.679921	1.228964	2.296353	0.001143
GTF2E2	0.671251	0.473699	0.95119	0.025004
TAF1C	1.866283	1.237083	2.815504	0.002939
INTS1	1.467568	1.019766	2.112012	0.03889
ERCC3	1.992833	1.071324	3.706989	0.029441
CIAO1	2.052983	1.029583	4.093638	0.041076
PELP1	1.499056	1.030459	2.180745	0.034272
SNAPC3	1.629411	1.053366	2.520475	0.028267
DDX11	1.637622	1.121073	2.392178	0.01074
RAD9A	1.596121	1.015991	2.507504	0.042479
NOP14	0.619611	0.401996	0.955029	0.030127
DHX38	1.60973	1.003591	2.581959	0.048287
SFSWAP	1.931741	1.078866	3.458841	0.026734
MMS19	2.041935	1.171788	3.558236	0.011754
POLRMT	1.546128	1.034675	2.3104	0.033479