



OPEN The anoikis-related gene signature predicts survival accurately in colon adenocarcinoma

Gunchu Hu¹, Jian Li¹, Yi Zeng^{2,3}, Lixin Liu¹, Zhuowen Yu^{2,4}, Xiaoyan Qi¹, Kuijie Liu^{1✉} & Hongliang Yao^{1✉}

Colon adenocarcinoma (COAD) is a serious public health problem, the third most common cancer and the second most deadly cancer in the world. About 9.4% of cancer-related deaths in 2020 were due to COAD. Anoikis is a specialized form of programmed cell death that plays an important role in tumor invasion and metastasis. The presence of anti-anoikis factors is associated with tumor aggressiveness and drug resistance. Various bioinformatic methods, such as differential expression analysis, and functional annotation analysis, machine learning, were used in this study. RNA-sequencing and clinical data from COAD patients were obtained from the Gene expression omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases. Construction of a prognostic nomogram for predicting overall survival (OS) using multivariate analysis and Lasso-Cox regression. Immunohistochemistry (IHC) was our method of validating the expression of seven genes that are linked to anoikis in COAD. We identified seven anoikis-related genes as predictors of COAD survival and prognosis, and confirmed their accuracy in predicting colon adenocarcinoma prognosis by KM survival curves and ROC curves. A seven-gene risk score consisting of NAT1, CDC25C, ATP2A3, MMP3, EEF1A2, PBK, and TIMP1 showed strong prognostic value. Meanwhile, we made a nomogram to predict the survival rate of COAD patients. The immune infiltration assay showed T cells. CD4 memory. Rest and macrophages. M0 has a higher proportion in COAD, and 11 genes related to tumor immunity are important. GDSC2-based drug susceptibility analysis showed that 6 out of 198 drugs were significant in COAD. Anoikis-related genes have potential value in predicting the prognosis of COAD and provide clues for developing new therapeutic strategies for COAD. Immune infiltration and drug susceptibility results provide important clues for finding new personalized treatment options for COAD. These findings also suggest possible mechanisms that may affect prognosis. These results are the starting point for planning individualized treatment and managing patient outcomes.

COAD is the fourth most common malignancy with an estimated 247,563 deaths in China in 2018¹. COAD is a disease that occurs only in the colon and is caused by excessive proliferation of epithelial cells in the colonic gland. There are three main types of COAD: sporadic, hereditary, and colitis-associated. If COAD is detected early enough, it can be completely cured with surgery and treatment. However, high recurrence rates and anti-cancer drug resistance increase the rate of treatment failure rates². In recent years, with the development of the economy, the incidence and mortality of COAD have risen rapidly³. Molecular studies are playing an increasingly important role in predicting prognosis and determining optimal treatment in poor-prognosis colon cancer⁴. Emphasizing the status, development, risk factors and controls of COAD is important to raise public awareness.

Anoikis, a form of programmed cell death resulting from loss of interaction with the cells outer membrane, is known as a physiological metastatic disease^{5–7}. Anoikis occurs when cells detach from their normal extracellular matrix (ECM), which provides them with signals for survival and growth. By inducing anoikis, the body prevents cells from growing and implanting in inappropriate locations, such as other organs, where they could cause harm. However, tumor cell metastasis is the process by which cancer cells spread from their primary site to distant sites in the body, forming secondary tumors. To achieve metastasis, tumor cells must overcome anoikis and survive in ECM-deprived conditions. Cancer cells must develop anoikis resistance to survive in the bloodstream before

¹Department of General Surgery, The Second Xiangya Hospital of Central South University, 139 Renmin Road, Changsha 410011, Hunan, China. ²Department of Orthopedics, the Second Xiangya Hospital of Central South University, Changsha 410011, China. ³Orthopedic Biomedical Materials Engineering Laboratory of Hunan Province, Changsha 410011, China. ⁴Hunan Key Laboratory of Tumor Models and Individualized Medicine of Hunan Province, Changsha 410011, China. ✉email: liukuijie@csu.edu.cn; yaohl0326@csu.edu.cn

metastasizing to distant organs^{8,9}. Acquired anaerobic resistance allows cancer cell survival in circulation and is important for metastatic progression. However, the molecular mechanism of anoikis resistance in tumor cells remains unknown. Jin Lingtao confirmed that the increased expression of GDH1 mediated by the transcription factor PLAG1 and its downstream substrate AMPK can provide anti-anoikis and pro-metastasis signals for lung cancer cells¹⁰. Lindsay J shows overexpression of CBX2 in high-grade primary serous ovarian cancer tumors and cell lines to prevent anoikis in suspension culture, showing high-grade serous ovarian cancer cells under different culture conditions. Lindsay J show that loss of CBX2 is relevant. It is associated with decreased proliferation, increased blood sensitivity and decreased stem cell numbers¹¹. Anoikis-related genes are genes that control cell-ECM communication, which avoid cell proliferation and implantation in inappropriate locations during tissue formation and maintenance. Moreover, they also involve cancer cell anoikis tolerance, invasive ability and TME attributes, which are associated with tumor clinical characteristics, prognosis and efficacy. Therefore, they are candidate biomarkers or targets for tumors.

The corresponding information of the COAD samples was obtained by mining the TCGA database. We then used Genecards to isolate anoikis-related genes. A prediction model of anoikis-related genes in COAD was then constructed and validated using the GEO database. Simultaneously, immune infiltration and drug susceptibility tests were performed. The purpose of this study is to construct prognostic markers related to COAD and investigate their potential mechanism of action and clinical value in COAD and make recommendations to find new treatment options.

Materials and methods

Acquisition and processing of data and DEGs identification. Expression sets and clinical information [gender, age, OS, PFS, TNM stage] (expression was measured as raw read counts using STAR, then converted to transcripts per million (TPM). Table 1 listed the baseline clinical characteristics of TCGA-COAD. Only COAD patients with survival information were included in this study. Anoikis-related genes were obtained from Genecards (<https://www.genecards.org/>).

Using the Limma package¹², a linear model of the microarray data was used to distinguish DEGs from COAD patients and healthy controls in an integrated microarray expression matrix. The false discovery rate (FDR) was examined using the Benjamini-Hochberg method. A P-value < 0.05, and $|\log_2(\text{fold change})|(|\log_{10}(\text{FC})|) \geq 1$ indicates a significant difference. Heatmaps and volcano maps were created using R. ComplexHeatmap^{13,14} and ggplot2 package.

Gene enrichment analysis. To investigate possible related pathways of COAD, we performed Gene Enrichment Analysis (GSEA) using the “GOSemSim” package^{15,16}, “clusterProfiler” package¹⁷, “biomaRt” package^{18,19} and “string” package in R. GO Enrichment and KEGG pathway analysis. The “ggplot2” package displays the results. Based on the associated anoikis, we used the Genecards online resource to exclude genes that might play a role in anoikis. The screening criteria were as follows: category = “encoded protein” score cut-off > 0.15.

Colon adenocarcinoma CNVs analysis. Whole exome/genome sequence (WXS/WGS) somatic mutation COAD data were downloaded from the GDC-TCGA-COAD project to the UCSC-Xena server. The Uctect2 algorithm assigns high confidence to somatic variants and detects other germline mutations. Tumors were mapped in descending order of mutation using the R package “maftools”²⁰.

Construction of risk score model and validation. A total of 426 samples will be selected from the TCGA-COAD dataset as the training cohort. In the training cohort, 189 anoikis-related genes were screened based on univariate Cox analysis. Thirteen potential prognostic genes were then screened using the Least Absolute Shrinkage Regression and Selection Operator (LASSO). Finally, only 7 genes were included in the risk signature according to the results of the multivariate cox regression. The risk score was calculated according to the gene expression level and the regression coefficient with “glmnet” package²¹, and the formula was risk score = coefficient Σ (Genei) $\times$ expression(Genei). According to the principle of median risk score, the OS of this group was divided into high-risk group and low-risk group, and the “Survivor” R package was run for analysis. We then used KM survival curves and ROC curves to evaluate their value in predicting OS survival prognosis with “pROC” package²². In addition, risk score histograms, survival status scatterplots, and DEGs distribution heatmaps were generated for both groups of patients. Survival curves from the validation set (GSE39582, 1 $\leq$ OS.time $\leq$ 120, months) were used to assess the reliability of the validated group risk assessment model. The main clinical features of GSE39582 (1 $\leq$ OS.time $\leq$ 120, months) are shown in Table 2. The 1-, 3-, and 5-year survival rates of COAD patients were also predicted from the training and validation sets.

Immune cell infiltration analysis. We extracted expression data from 22 tumor-infiltrating immune cells using the Cybersort platform. We then analyzed this using sample OS data and implemented algorithms using CyberSort to estimate the proportion of different immune cells among OS cells. The algorithm is based on Monte Carlo sampling, providing scattered p-values for each sample.

Drug sensitivity analysis. The “OncoPredict” R package²³ was used to search susceptibility data in the GDSC2 database and to predict drug COAD responses in the GDSC2 database. Spearman correlation analysis was then performed to obtain drugs associated with risk outcomes and plot the correlation scatterplot with “ggstatsplot” package.

Basic characteristics	Variables	Overall (N = 426)
Age (years)	Median (1st Qu, 3rd Qu)	68.00 (58.00, 77.00)
Sex	Male	229 (53.8%)
	Female	197 (46.2%)
History of colon polyps	Yes	127 (29.8%)
	No	234 (54.9%)
	Unknown	65 (15.3%)
History of neoadjuvant treatment	Yes	3 (0.7%)
	No	423 (99.7%)
Histological type	Colon adenocarcinoma	363 (85.2%)
	Colon mucinous adenocarcinoma	59 (13.9%)
	Unknown	4 (0.9%)
TNM stage	Stage I	75 (17.6%)
	Stage II	167 (39.2%)
	Stage III	124 (29.1%)
	Stage IV	60 (14.1%)
T stage	T1	11 (2.6%)
	T2	74 (17.4%)
	T3	292 (68.5%)
	T4	49 (11.5%)
N stage	N0	251 (58.9%)
	N1	101 (23.7%)
	N2	74 (17.4%)
M stage	M0	365 (85.7%)
	M1	61 (14.3%)
Longest dimension(cm)	≤ 1 cm	125 (29.3%)
	> 1 cm	124 (29.1%)
	Unknown	177 (41.6%)
Residual tumor	R0	310 (72.8%)
	R1	4 (0.9%)
	R2	21 (4.9%)
	Rx	91 (21.4%)
Survival status	Alive	334 (78.4%)
	Dead	92 (21.6%)
OS.time (months)	Median (1st Qu, 3rd Qu)	13.23 (23.07,36.50)
Progression-free-survival status	No progressive disease	311 (73.0%)
	Progressive disease	115 (27.0%)
PFS.time (months)	Median (1st Qu, 3rd Qu)	20.28 (12.22,33.52)

Table 1. Clinical characteristics of COAD samples from the TCGA database.

Immunohistochemistry for model genes. We performed IHC using clinical samples of COAD to confirm the expression differences of model genes in COAD clinically. For IHC experiments, we collected 70 tissue sections (five samples per COAD and tumor-adjacent normal tissue). We rehydrated the sections with ethanol and dewaxed them with xylene. Then we treated them with 3% H₂O₂ for 20 min to block endogenous peroxidase and with 1 mM EDTA to retrieve the antigen. We incubated the sections with 1:50 diluted model genes antibody (From: SANTA CRUZ BIOTECHNOLOGY) at 4 °C overnight. After incubation with a 2-step plus PolyHRP Anti-Mouse/Rabbit IgG Detection System (PV-9000, Zhongshan Jinqiao Biotechnology Company, Beijing, China), we visualized the sections with diaminobenzidine (DAB; Zhongshan Jinqiao Biotechnology Company, Beijing, China), counterstained them with hematoxylin, and dehydrated them. We scanned the sections using a Zeiss microscope. We measured and quantified staining intensity using the ImageJ plug-in “IHC Toolbox” and GraphPad Prism version 7 software, respectively. We considered p-value < 0.05 as significant.

Statistical analysis. In this experiment, all statistical analyzes were performed using R 4.2.2. The Wilcoxon test was used to compare nonparametric data from two independent samples. Parametric data were analyzed by *t* test and one-way ANOVA. A P value < 0.05 was considered statistically significant (*p-value < 0.05; **p-value < 0.01; ***p-value < 0.001). Related R packages include ggplot2, ggpubr, survival, survminer, and others taken from the Bioconductor package or the R package. P < 0.05 was considered statistically significant for each analysis.

Basic characteristics	Variables	Overall (N = 491)
Age (years)	Median (1st Qu, 3rd Qu)	68.10 (58.00, 76.00)
Sex	Male	268 (54.6%)
	Female	223 (45.4%)
Cin.status	+	311 (63.3%)
	–	102 (20.8%)
	Unknown	78 (15.9%)
Cimp.status	+	83 (16.9%)
	–	354 (72.1%)
	Unknown	54 (11.0%)
TNM stage	Stage I	30 (6.1%)
	Stage II	219 (44.6%)
	Stage III	183 (37.3%)
	Stage IV	59 (12.0%)
T stage	T1	9 (1.8%)
	T2	43 (8.8%)
	T3	332 (67.6%)
	T4	107 (21.8%)
N stage	N0	264 (53.8%)
	N1	124 (25.2%)
	N2	103 (21.0%)
M stage	M0	431 (87.8%)
	M1	60 (12.2%)
Kras.mutation	WT	282 (57.4%)
	M	188 (38.3%)
	Unknown	21 (4.3%)
Tp53.mutation	WT	137 (27.9%)
	M	162 (33.0%)
	Unknown	192 (39.1%)
Braf.mutation	WT	398 (81.1%)
	M	47 (9.6%)
	Unknown	46 (9.3%)
MMR.status	pMMR	389 (79.2%)
	dMMR	69 (14.1%)
	Unknown	33 (6.7%)
Chemotherapy.adjuvant	+	205 (41.8%)
	–	270 (55.0%)
	Unknown	16 (3.2%)
Survival status	Alive	320 (65.2%)
	Dead	171 (34.8%)
OS.time (months)	Median (1st Qu, 3rd Qu)	50.00 (26.00, 75.00)

Table 2. Clinical characteristics of COAD samples from the GSE39582 ($1 \leq \text{OS.time} \leq 120$, months).

Results

Identification of DEGs and GSEA analysis. Figure 1 is the flow chart of our study. Based on the Limma package and previously defined thresholds, a total of 3335 DEGs that could be further enriched were identified, of which 1547 were downregulated and 1788 were upregulated (Supplementary Table S1). Figure 2a shows the DEG volcano map and Fig. 2b shows the heatmap.

First, GSEA analysis is performed to examine the most relevant genetic ontologies. As shown in Fig. 2c,d (Supplementary Table S2), “adaptive immune response”, “antigen binding”, “immunoglobulin complex” and “immunoglobulin receptor binding” were the most abundant terms in Gene Ontology. GSEA pathway accumulation analysis showed that DEGs were enriched in pathways such as “colon cancer”, “mineral absorption”, “hippo pathway” and “IL-17 pathway” (Fig. 2e,f) (Supplementary Table S3).

GO and KEGG analysis as well as CNVs analysis. GO analysis demonstrated that these 3335 DEGs were primarily enriched in complement activation, B cell receptor signaling pathway, and humoral immune response mediated by circulating immunoglobulin in the biological process, immunoglobulin complex, external

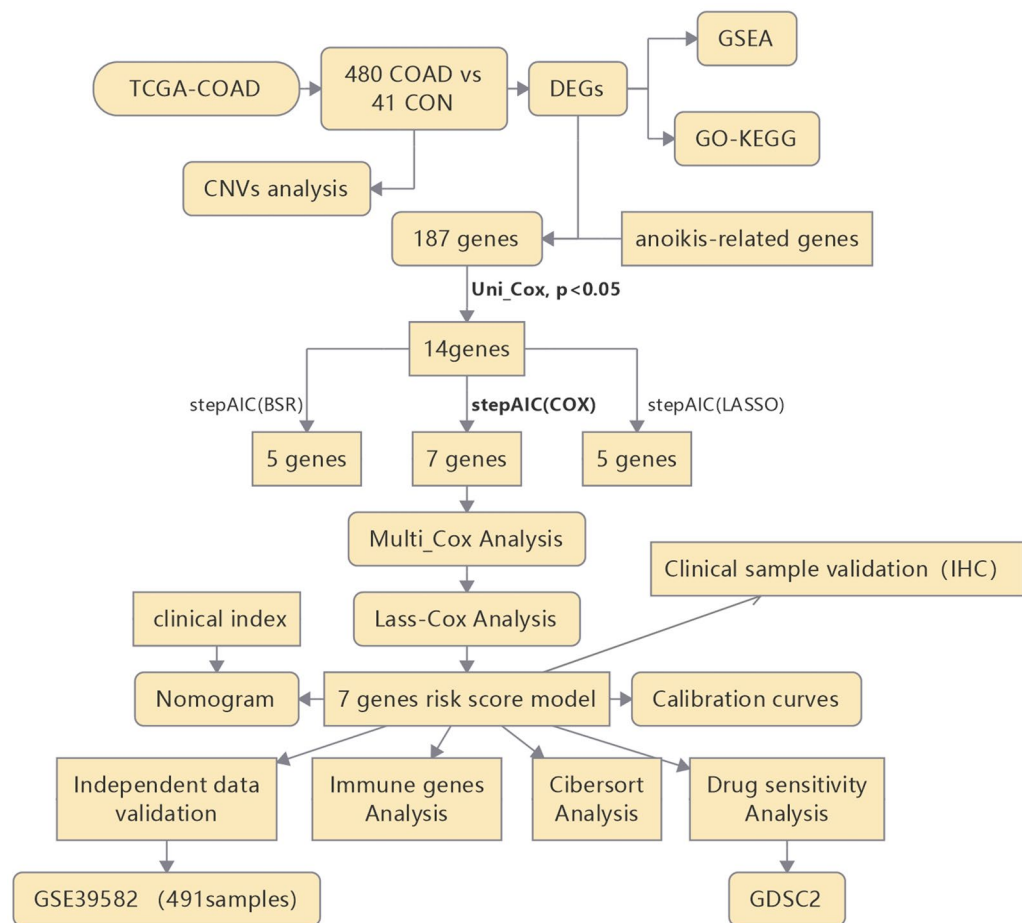


Figure 1. Flow chart of this study.

side of plasma membrane, and collagen-containing extracellular matrix in the cellular components, antigen binding, glycosaminoglycan binding, and extracellular matrix structural constituent in the molecular functions (Fig. 3a,b) (Supplementary Table S4). The KEGG pathway enrichment analysis indicated that these DEGs were enriched in the pathways such as PI3K-Akt signaling pathway, Cytokine-cytokine receptor interaction and Viral protein interaction with cytokine and cytokine receptor (Fig. 3c,d) (Supplementary Table S5).

According to the classification of mutations, we found that missense mutations were the most common (Fig. 3e). SNPs were the most common type of variation, with C > T (112,505) ranking first in the single nucleotide variation (SNV) category. We found a higher tumor mutational burden in COAD (Fig. 3f). We also found that APC (72%) and TP53 (55%) were more frequently mutated than other genes (Fig. 3g). We found the variant allele frequencies for the top 10 genes in COAD (Fig. 3h).

Filter anois-related genes. We excluded genes that might play a role in anois using the online resource Genecards. Screening criteria: category = “encoding protein”, score cutoff > 0.15 (Supplementary Table S6).

Then, we take the intersection of DEGs and the anois-related genes (Supplementary Fig. S1, Table S7). Next, we performed univariate COX hazard analysis on 187 genes, with OS as the outcome measure (Supplementary Table S8), with PFS as the outcome measure (Supplementary Table S9). We selected 25 genes to draw the forest plot of Univariate Cox regression (Fig. 4a,b). Afterwards, we performed stepwise backward regression on the results separately. The variables of the three different models are variables screened by Univariate Cox ($p < 0.05$), variables screened by All-Subsets Regression (BSR) (Supplementary Fig. S2), and variables screened by LASSO regression. After comparing the three methods (Supplementary Fig. S3), 7 genes were finally selected to draw the forest plot of multicox regression (Fig. 4c,d).

Establishment of the seven-gene risk signature. We performed Lasso algorithm on these 7 DEGs and regarded them as candidate genes (Fig. 5a,b). Generation of risk scores using seven anois-related genes to predict survival and prognosis in COAD patients (Fig. 5c). The risk score = $-0.52021 \times \text{NAT1} + 0.18218 \times \text{EF1A2} + (-0.39270 \times \text{CDC25C}) + 0.28190 \times \text{TIMP1} + (-0.11126 \times \text{MMP3}) + 0.21511 \times \text{PBK} + (-0.19454 \times \text{ATP2A3})$. KM analysis showed that the training set has a high risk score. Corresponded with poorer overall survival (HR = 2.78 (1.97–3.93), log-rank $p = 6.45 \times 10^{-9}$) and poorer progress free survival (HR = 2.13 (1.58–2.88),

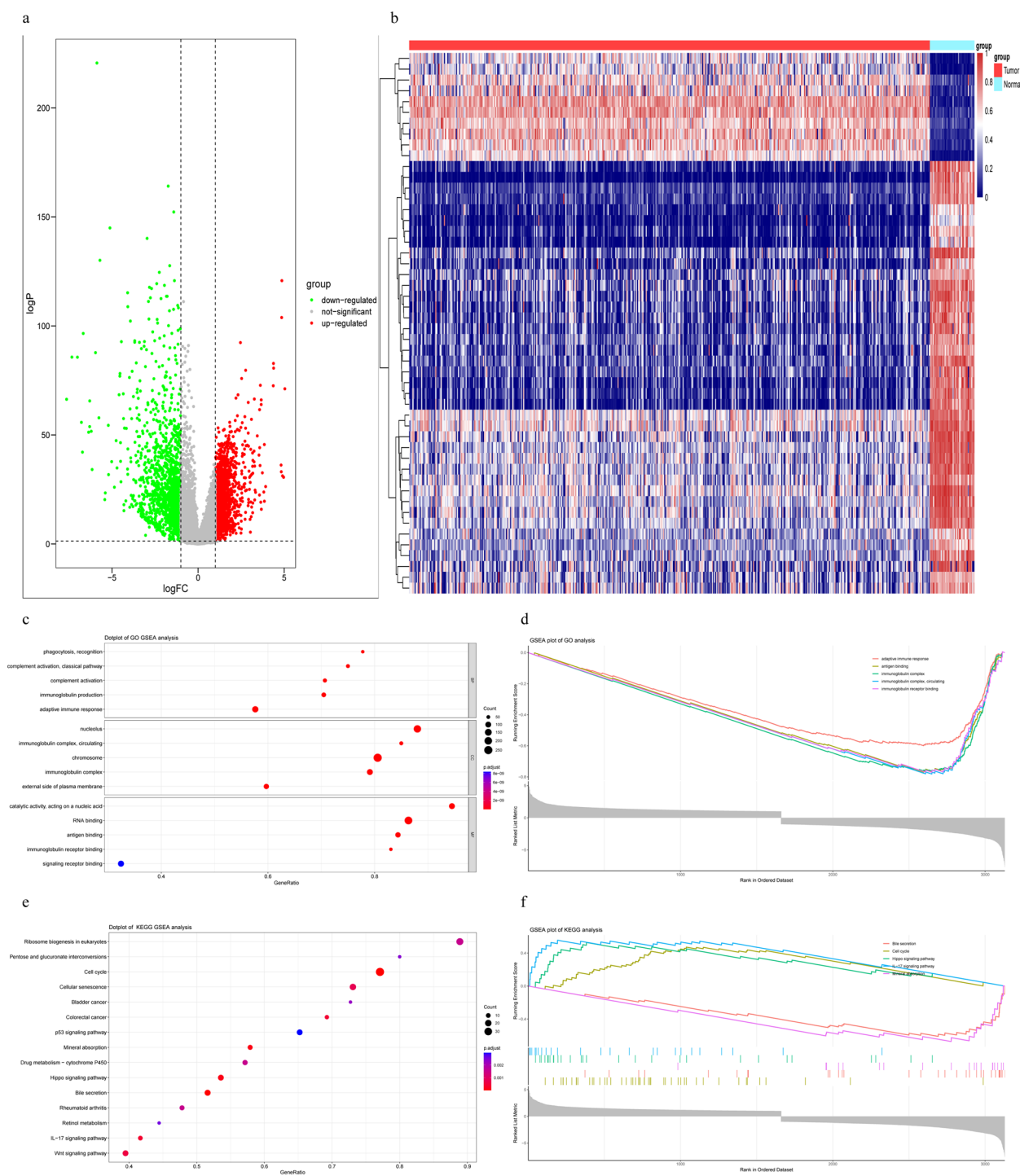


Figure 2. DEGs screening in integrated TCGA database and GSEA analysis. **(a)** Volcano plot of COAD samples from the TCGA database, Data points in red represent up-regulated, and green represent down-regulated genes, **(b)** Heatmap of DEGs identified in integrated microarray. Legend on the top right indicates the change of the genes, **(c,d)** GO results of GSEA analysis, **(e,f)** KEGG results of GSEA analysis.

log-rank $p=7.20 \times 10^{-7}$) (Figs. 5d,e, Supplementary Table S10). Seven genetic risk signatures could effectively divide patients into a high-risk group and a low-risk group with an intermediate risk score. Therefore, we can intuitively conclude that the patients who died basically happened first were higher ranking (higher risk score) (Fig. 5d–g). We draw the distribution of gene expression in the 426 COAD samples (Fig. 6a). We also applied a risk score covering all relevant genes to estimate OS and PFS after 1, 3, and 5 years (Fig. 6b,c). The prediction accuracy evaluated by AUCs was reported to be 0.69, 0.66 and 0.66 in the 1-year, 3-year and 5-year ROC curves, respectively. The same analyses were conducted for the PFS outcome.

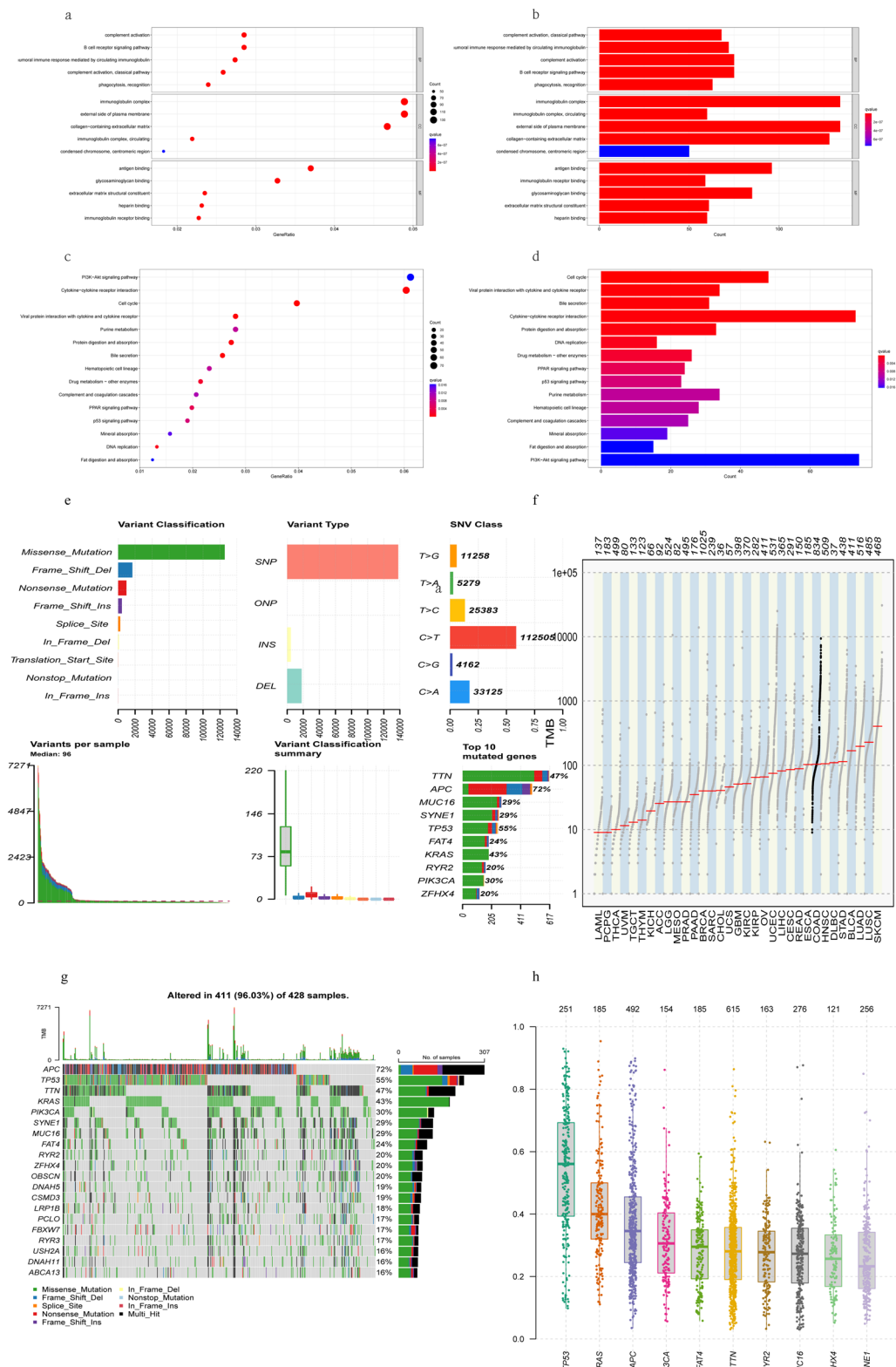


Figure 3. Enrichment Analyses of the DEGs from exprset and CNVs analysis in COAD. **(a,b)** Results of GO analysis with DEGs, **(c,d)** Results of KEGG analysis with DEGs, **(e–h)** the CNV and mutation frequency in COAD. * $p < 0.01$, *** $p < 0.001$; COAD colon adenocarcinoma, SNP single nucleotide polymorphism, INS insertion, DEL deletion.

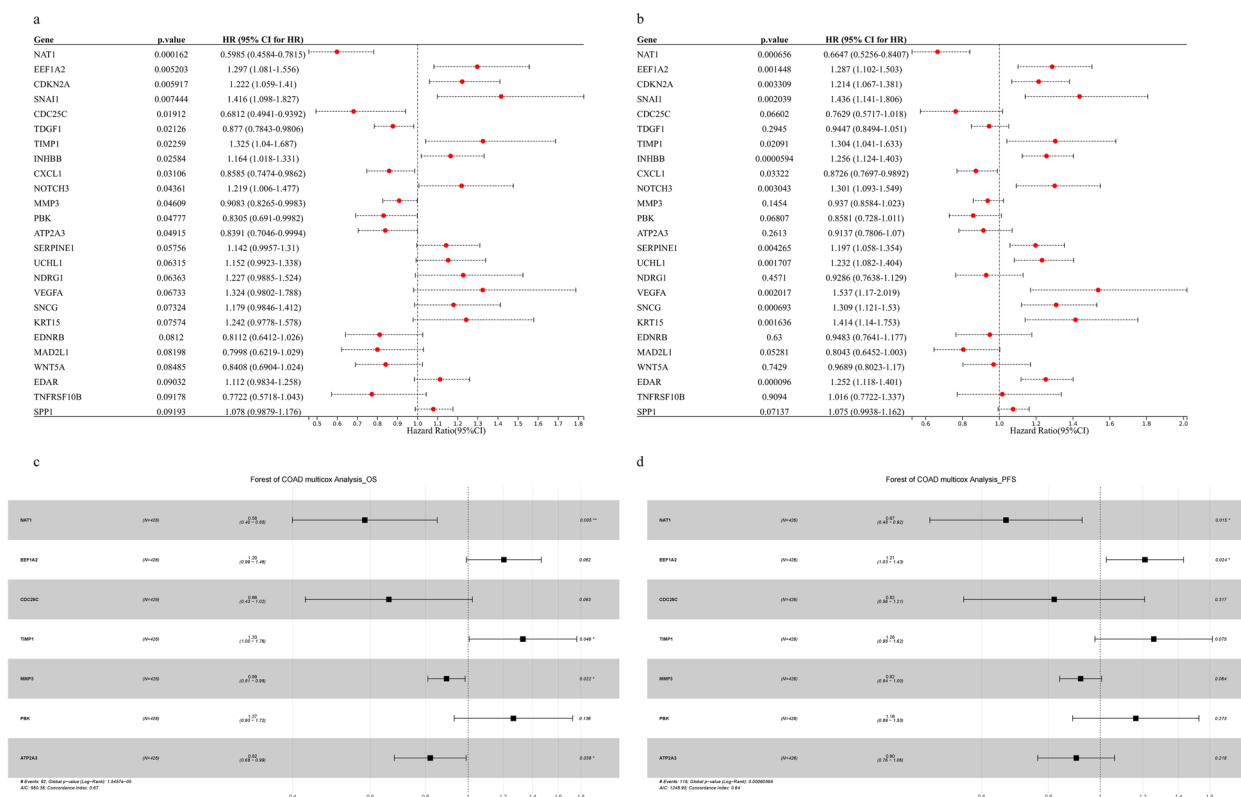


Figure 4. Forest plots for anoikis-related genes from univariate and multivariate cox proportional hazards mode. **(a)** The effect of 25 anoikis-related genes on the overall survival of TCGA datasets, **(b)** The effect of 25 anoikis-related genes on the progress free survival of TCGA datasets, For **(c)** OS and **(d)** PFS, hazard ratios and p-value of the constituents involved in multivariate Cox regression in COAD.

Nomogram development and validation for COAD. To facilitate the clinical application of the predictive model, we integrated the clinical information and genetic characteristics of TCGA patients, and constructed a nomogram using a multivariate Cox regression model (Fig. 6d,e). Use the discrimination and calibration method to identify and calibrate OS and PFS results. The c-index for OS was 0.775 and the c-index for PFS was 0.742 (Supplementary Table S11), reflecting the superior predictive power of the nomogram. Furthermore, standard curves showed good agreement between predicted and observed OS or PFS between 1, 3, and 5-year survival (Fig. 6f,g).

Validation of the 7 genes risk score model based on GEO dataset. We select GSE39582 ($1 \leq OS$, time ≤ 120 , months) as the validation data set. We plotted a heatmap of the gene expression distribution of the 532 COAD samples (Fig. 7a). KM analyses showed that the validation set has a high risk score corresponded with poorer overall survival (Fig. 7b). The predictive accuracies of AUC results were 0.718, 0.742, and 0.5695 for the 1-year, 3-year, and 5-year ROC curves, respectively (Fig. 7c). We plot scatterplots and histograms to describe the survival status and risk score distributions of these COAD samples (Fig. 7d).

Immune analysis. We analyzed genes related to the immune system, looked for genes that were differentially expressed and drew boxplots. IRF3 and VEGFB expression increased in the high-risk group and vice versa (Fig. 8a). The distribution of these genes in the two group of COAD samples is shown in Fig. 8b. The occurrence and development of COAD is affected by the tumor microenvironment, and T cells, memory CD4 cells, M0 macrophages, and plasma cells are the most common tumor-infiltrating immune cells (Fig. 8c,d). As a result of the immunological analysis, statistically significant differences in the frequencies of six types of immune cells were observed between the favorable and poor prognosis groups (Fig. 8e–g). It has been suggested that the two groups differed in the ratio of mesenchymal cells to immune cells, possibly resulting in different tumor purities. These results suggest that immune infiltration and immune microenvironment are important for OS in COAD patients.

Correlation of risk score with drug sensitivity. We evaluated the value of risk scores in predicting drug sensitivity in different cancer types. We tested 30 drugs in the GDSC2 database whose risk scores were significantly associated with drug sensitivity according to Spearman correlation analysis. The risk score is negatively sensitive to five drugs including Uprosertib, Venetoclax, Fludarabine, Buparlisib, and Osimertinib, and positively correlated with sensitivity to 25 drugs, including Oxaliplatin, 5-Fluorouracil, Cisplatin, Gemcitabine, Camptothecin, Irinotecan and others (Fig. 9a,b,f–h). In conclusion, the establishment of the risk score helps us

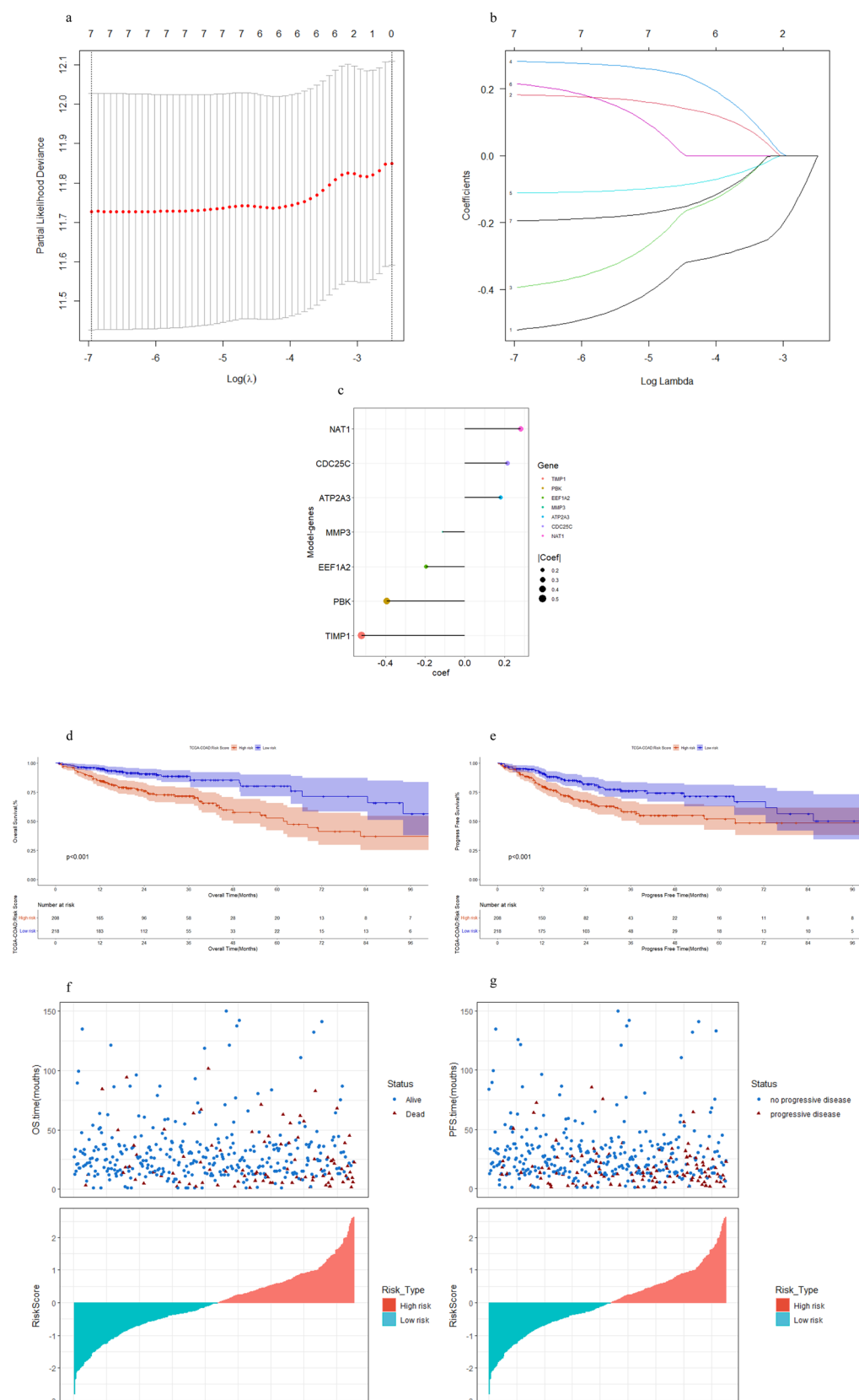


Figure 5. Construction of the risk score. (a,b) The least absolute shrinkage and selection operator (LASSO) method of anoikis-related genes associated with prognosis, (c) The risk score for predicting the survival and prognosis of patients with COAD, (d) Kaplan Meier plot of the anoikis signature and overall survival, (e) Kaplan Meier plot of the anoikis signature and progress-free survival, (f) distribution of risk score, overall survival status, (g) distribution of risk score, progress-free survival status.

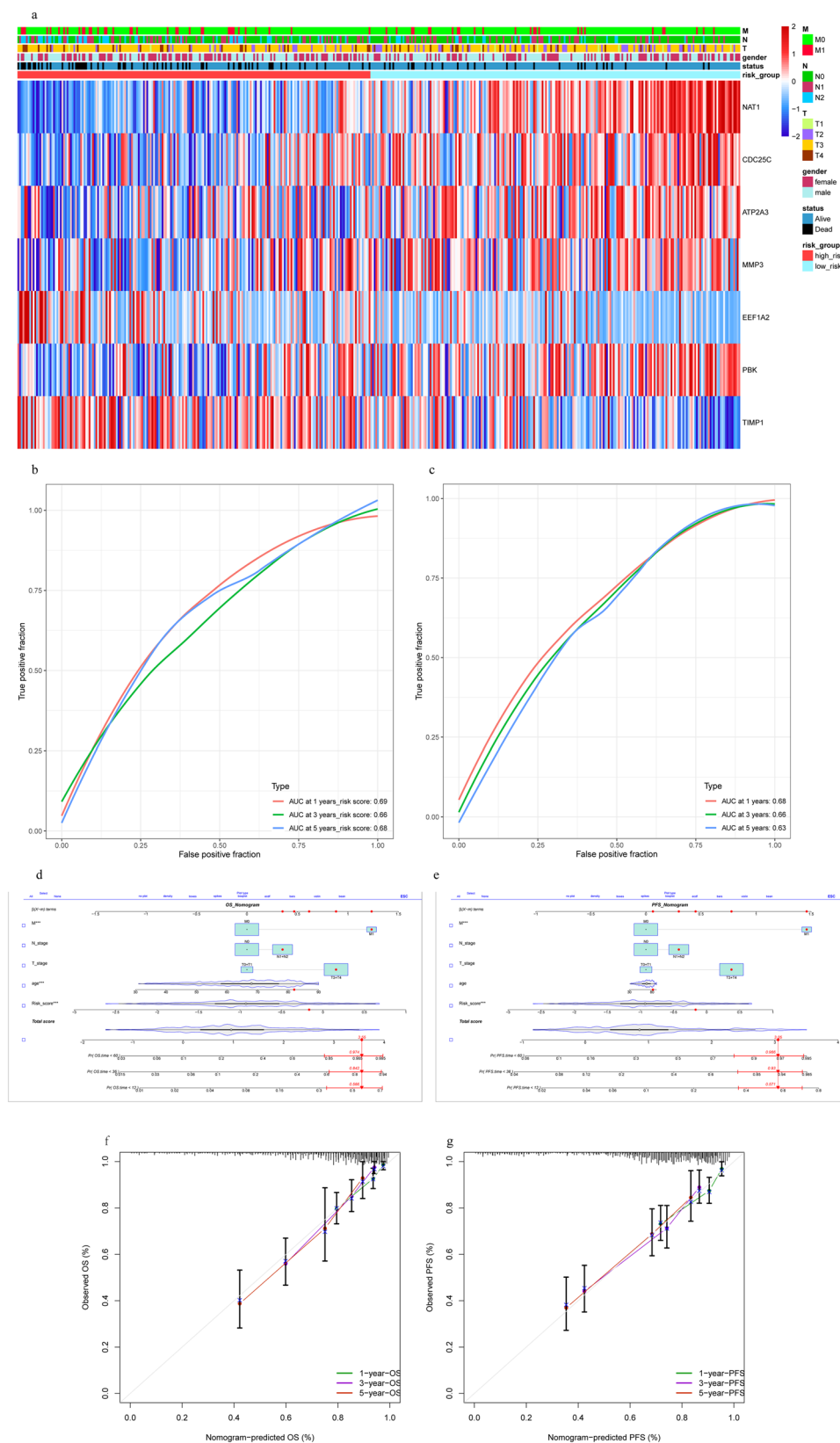


Figure 6. Nomogram development and validation. **(a)** Heatmap of model-genes identified, Legend on the top right indicates the characteristic of the samples, **(b)** ROCs for 1-year, 3-year and 5-year survival prediction, **(c)** ROCs for 1-year, 3-year and 5-year progression-free survival prediction. Nomogram to predict the 1-year, 3-year and 5-year **(d)** OS and **(e)** PFS rate of COAD patients. Calibration curve for the **(f)** OS and **(g)** PFS nomogram model in COAD. A dashed diagonal line represents the ideal nomogram.

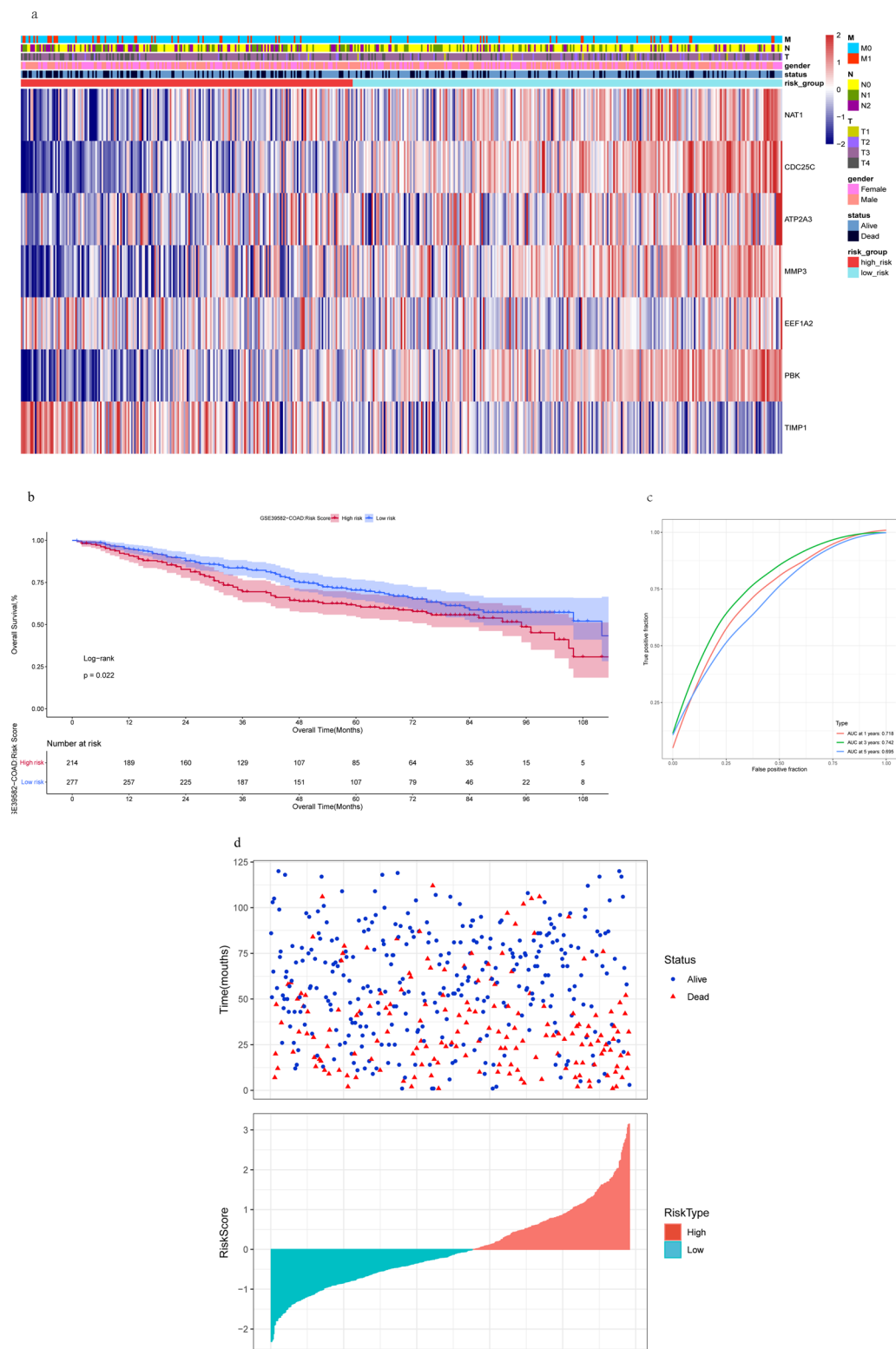


Figure 7. Validation of the seven-gene Signature. **(a)** Heatmap of model-genes identified in validation dataset, Legend on the top right indicates the characteristic of the samples, **(b)** Kaplan Meier plot of the anoikis signature and overall survival in validation dataset. **(c)** ROCs for 1-year, 3-year and 5-year survival prediction in validation dataset. **(d)** Distribution of risk score, overall survival status.

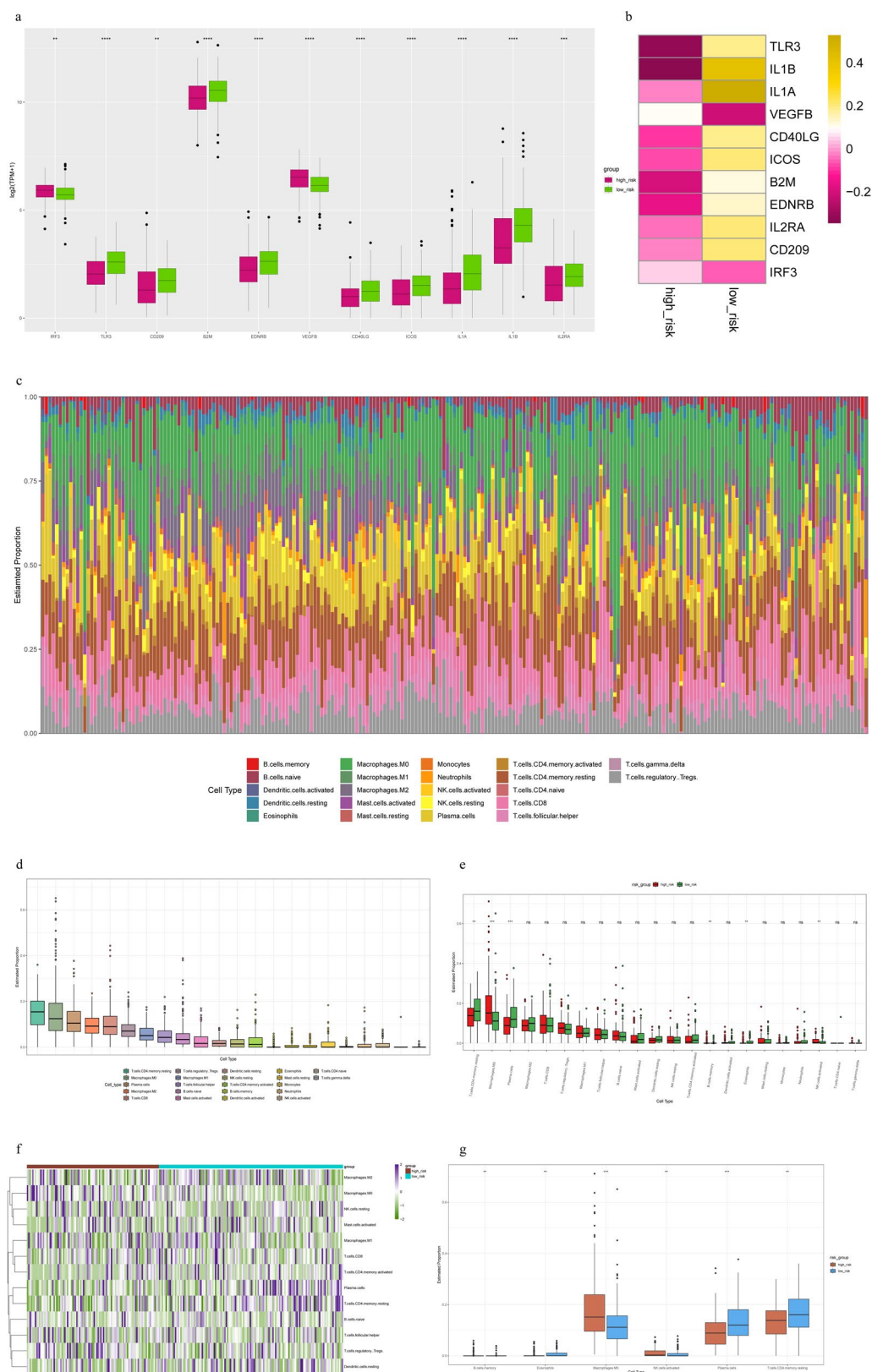


Figure 8. Immune analysis. **(a,b)** Comparisons of immune-related genes between the two risk groups in COAD patients. **(c,d)** Immune infiltration of 22 different types of immune cells. **(e)** Immune cell component between high-risk group and low-risk group. **(f)** The landscape of immune infiltration in two risk groups for COAD patients. **(g)** Boxplot visualizing differentially immune infiltration. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

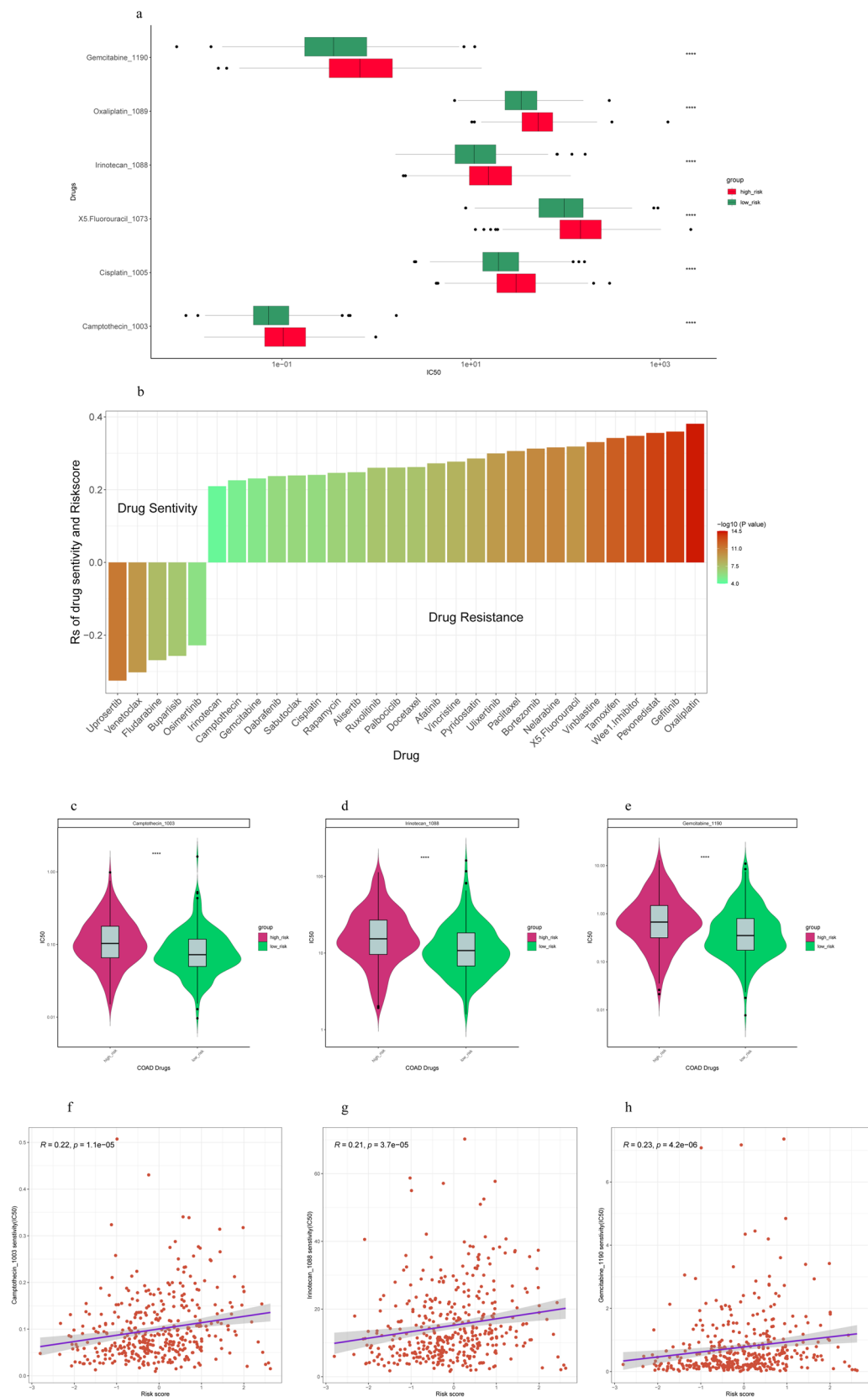


Figure 9. Drug sensitivity. **(a)** Comparisons of 6 COAD chemotherapy drugs on GDSC2 database. **(b)** Assessing drug sensitivity of COAD tumor based on the risk score. **(c–e)** Boxplot visualizing differentially drugs. **(f–h)** The correlation analysis between risk score and the IC50 of 3 chemotherapy drugs.

consider appropriate and effective treatment strategies. The IC50 of three drugs in the high-risk group was lower than low-risk group (Fig. 9c–e). These three drugs are Camptothecin, Irinotecan, Gemcitabine, which indicates that they are expected to become potential drugs for the treatment of COAD.

Expression of model genes in clinical samples. Compared with tumor adjacent tissues (Fig. 10), EEF1A2, PBK and TIMP1 showed significant upregulation in colon adenocarcinoma, while ATP2A3, CDC25C,

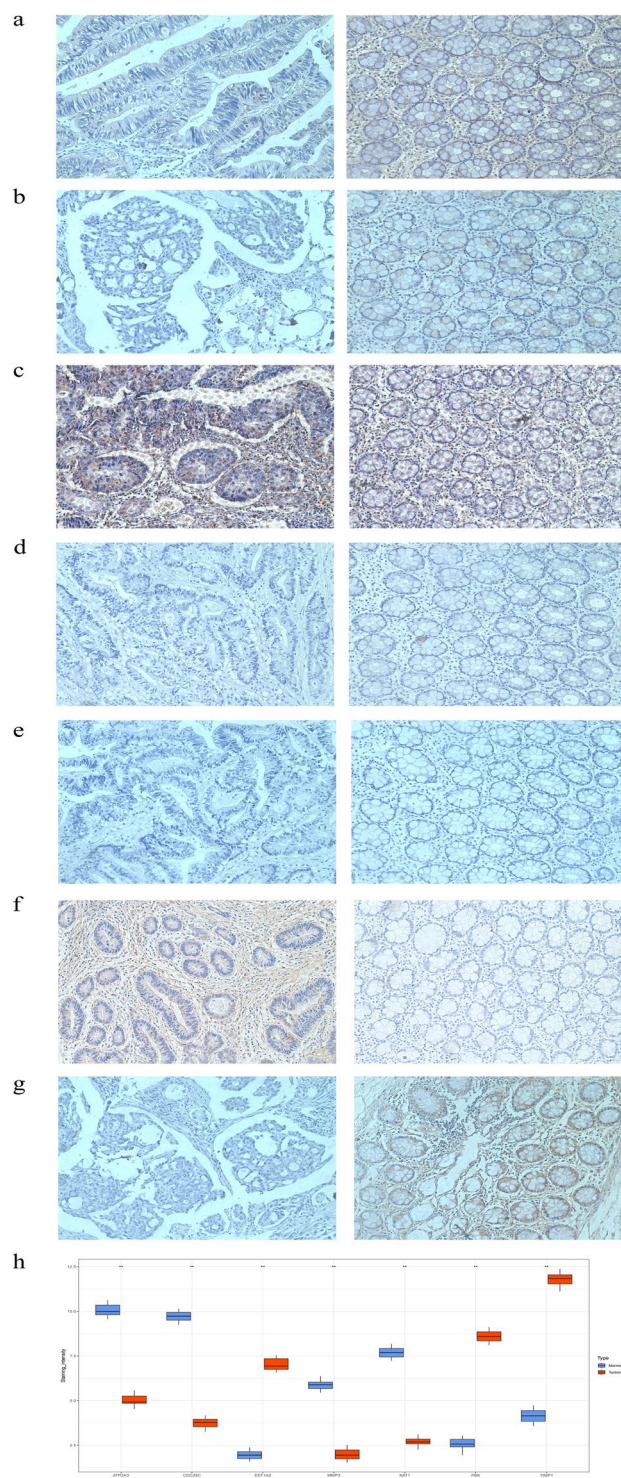


Figure 10. Model genes expression in COAD validated by clinical samples. Representative images and quantification of IHC staining with (a) ATP2A3, (b) CDC25C, (c) EEF1A2, (d) TIMP1, (e) NAT1, (f) PBK, and (g) MMP3. Magnification: 20x. p-values were obtained by unpaired *t* test. All data are represented by mean $\pm$ SD.

MMP3 and NAT1 showed significant downregulation. This outcome, is consistent with the results of the risk score model established by us.

Discussion

Anoikis is a type of programmed cell death that occurs when cells lose their proper extracellular matrix, thereby disrupting integrin attachment²⁴. This is an important mechanism to prevent dysplastic cells from proliferating and attaching to inappropriate substrates²⁵. Anoikis is also an important body defense that prevents detached cells from attaching to new substrates in the wrong place, thereby preventing their growth and shrinkage²⁶. Anoikis is critical for tissue homeostasis and development as it prevents colonization elsewhere by detached epithelial cells²⁷. Typically, detachment of normal epithelial cells results in loss of key survival factors and programmed cell death, termed anoikis²⁸. Resistance to anoikis, a specific mode of apoptosis activated during cell division, is a critical step in cancer metastasis²⁹. Several studies have shown that the initiation of apoptosis depends on both intrinsic and extrinsic signaling pathways³⁰. Anoikis can be activated by various intracellular signals, such as DNA damage and endoplasmic reticulum stress, and mitochondria play an important role in the regulation of apoptosis³¹. This impaired function of hypoxia is a hallmark of tumor cells and can lead to tumor invasion and migration, distant metastasis, and drug resistance^{32–34}. Anoikis is now of particular interest to the scientific community because anchorage-dependent growth and epithelial-mesenchymal transition, two properties associated with Anoikis resistance, are key steps in tumor progression and dissemination of metastatic cancer cells^{35–39}. However, the effects of anoikis-related genes on COAD invasion, migration and drug resistance and their roles in predicting COAD prognosis are rarely studied. Malagobadan S found that the overexpression of this novel miRNA, miR-6744-5p promotes insomnia in Luminal A and triple-negative breast cancer cell lines and directly targets the NAT1 enzyme⁴⁰. Cai Jiaqin & research found that the expression of NAT1 was significantly reduced in colorectal cancer, which was independently related to the poor prognosis of colorectal cancer patients. NAT1 can exert anti-tumor activity by inhibiting the phosphorylation of pi3k/Akt/mTOR signaling pathway⁴¹. Dixie E found that expression of EEF1A2 did not alter sensitivity to anoikis in SK-OV-3 cells⁴². As an epithelial-mesenchymal transition gene, EEF1A2 was found to be strongly associated with the clinical outcome of COAD and has been used as a biomarker and therapeutic drug target in the literature^{43,44}. Cui Tang' research results showed that HOXB13 plays a tumor-promoting role in HCC cells, promotes HCC drug resistance by up-regulating CDC25C, and improves the ability of cells to resist anoikis⁴⁵. These studies found that CDC25C may be an important target gene in patients with COAD because they are related to metabolism, cell cycle and tumor progression^{46,47}. Mariana Toricelli found that Timp1 confers cell survival through activation of the PDK1 pathway and that Timp1 and AKT cooperate to confer resistance to anoikis in metastatic melanoma cells⁴⁸. TIMP1 stimulates cell proliferation and has anti-apoptotic functions⁴⁹. In colorectal cancer, upregulation of TIMP1 in tumor tissue compared with normal tissue⁵⁰ is considered to be an independent prognostic factor for disease free survival⁵¹. MMP3 (Matrix Metalloproteinase 3) is a protein-coding gene, and MMP3-related diseases include Coronary Heart Disease 6 and conjunctivochalasis. Its associated pathways include downstream GPCR signaling and trimerization of collagen chains. Akira Koshino demonstrated that PBK is an immunohistochemical marker of good clinical outcome in CRC patients. PBK-mediated upregulation of cell proliferation and inhibition of CRC cell migration and invasion have also been demonstrated⁵². Overexpression of the PBK gene is associated with tumorigenesis⁵³. Immunohistochemical analysis of PBK/TOPK expression can be used as an independent marker for the prognosis of CRC patients^{54,55}. ATP2A3 encodes a SERCA pump that pumps Ca²⁺ into the lumen of the ER. One study showed that Serca3 expression is regulated by the proximal ATP2A3 promoter during the induction of epithelial cell differentiation⁵⁶. Previous studies have shown low or no expression of SERCA3 in colon cancer cell lines⁵⁷. Therefore, increased expression of SERCA3 in gastrointestinal cancer may be a prognostic biomarker. Brouland et al. We found that SERCA3 levels were inversely correlated with poorly differentiated epithelial cells. Furthermore, SERCA3 expression is lower in adenocarcinomas⁵⁸. These results suggest a strong correlation between transition from adenomatous polyposis to adenocarcinoma, COAD development, and aberrant SERCA3 expression⁵⁹. Previous studies have shown that these model genes are related to the occurrence and progression of most tumors, which is consistent with our research results.

Here, we demonstrate global genetic alterations associated with anoikis at the gene and transcriptional levels and demonstrate the cross-correlation of COAD. We developed a risk assessment model to predict prognosis and patient response to immunotherapy and targeted therapies. Examining the differential expression of anoikis-related genes in COAD will not only improve our understanding of the aggressiveness of COAD, but will also help in the development of more personalized and accurate immunotherapy regimens. In this experiment, we constructed a prognostic COAD risk assessment model based on the differential expression of anoikis-related genes. We investigated the value of risk scores in predicting the response of COAD to immunotherapy and analyzed the differential expression of immune cells in high-risk and low-risk cancers. The results demonstrate the usefulness of using this risk assessment model to predict patient response to immunotherapy. Examining risk score-based treatment outcomes in COAD patients, we found an interaction between risk score and drug sensitivity. Regulation of apoptosis, signal transduction and metabolism play an active role in the treatment of COAD.

Conclusion

In this study, we systematically generated and evaluated a COAD risk score based on seven anoikis-related genes. Systematic assessment of risk scores contributes to a better understanding of invasiveness and leads to more individualized and precise treatment strategies. Our model showed good performance in both the training and validation datasets and was further supported by immunohistochemistry.

Our study has profound implications for future research and clinical practice. Our model can also enable the identification of potential therapeutic targets and biomarkers for colon adenocarcinoma. In addition, our model

can enable the classification of patients into different risk levels and the customization of treatment options based on their individual prognosis. Our study fosters the progress of precision oncology and personalized medicine for colon adenocarcinoma patients.

Data availability

TCGADData Portal: <https://portal.gdc.cancer.gov/> (accessed on 15 March 2022); GEO Datasets: <https://www.ncbi.nlm.nih.gov/gds/> (accessed on 18 December 2022); UCSC Xena: <https://xenabrowser.net/> (accessed on 15 December 2022).

Received: 28 January 2023; Accepted: 18 August 2023

Published online: 25 August 2023

References

- Feng, R. M., Zong, Y. N., Cao, S. M. & Xu, R. H. Current cancer situation in China: Good or bad news from the 2018 Global Cancer Statistics?. *Cancer Commun. (Lond)*. **39**(1), 22 (2019).
- Hossain, M. S. *et al.* Colorectal cancer: A review of carcinogenesis, global epidemiology, current challenges, risk factors, preventive and treatment strategies. *Cancers (Basel)* **14**(7), 1732 (2022).
- Chen, W. *et al.* Cancer statistics in China, 2015. *CA Cancer J. Clin.* **66**(2), 115–132 (2016).
- Van Cutsem, E. *et al.* ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Ann. Oncol.* **27**(8), 1386–1422 (2016).
- Han, H. J. *et al.* Fibronectin regulates anoikis resistance via cell aggregate formation. *Cancer Lett.* **508**, 59–72 (2021).
- Fidler, I. J. The pathogenesis of cancer metastasis: The ‘seed and soil’ hypothesis revisited. *Nat. Rev. Cancer* **3**(6), 453–458 (2003).
- Paoli, P., Giannoni, E. & Chiarugi, P. Anoikis molecular pathways and its role in cancer progression. *Biochim. Biophys. Acta* **1833**(12), 3481–3498 (2013).
- Kim, Y. N., Koo, K. H., Sung, J. Y., Yun, U. J. & Kim, H. Anoikis resistance: An essential prerequisite for tumor metastasis. *Int. J. Cell Biol.* **2012**, 306879 (2012).
- Simpson, C. D., Anyiwe, K. & Schimmer, A. D. Anoikis resistance and tumor metastasis. *Cancer Lett.* **272**(2), 177–185 (2008).
- Jin, L. *et al.* The PLAG1-GDH1 axis promotes anoikis resistance and tumor metastasis through CamKK2-AMPK signaling in LKB1-deficient lung cancer. *Mol. Cell* **69**(1), 87–997 (2018).
- Wheeler, L. J. *et al.* CBX2 identified as driver of anoikis escape and dissemination in high grade serous ovarian cancer. *Oncogenesis* **7**(11), 92 (2018).
- Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**(7), e47 (2015).
- Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**(18), 2847–2849 (2016).
- Yu, G., Wang, L. G., Yan, G. R. & He, Q. Y. DOSE: An R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* **31**(4), 608–609 (2015).
- Yu, G. Gene ontology semantic similarity analysis using GOSemSim. *Methods Mol. Biol.* **2117**, 207–215 (2020).
- Yu, G. *et al.* GOSemSim: An R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* **26**(7), 976–978 (2010).
- Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2**(3), 100141 (2021).
- Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**(8), 1184–1191 (2009).
- Durinck, S. *et al.* BioMart and bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**(16), 3439–3440 (2005).
- Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **28**(11), 1747–1756 (2018).
- Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**(1), 1–22 (2010).
- Robin, X. *et al.* pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **12**, 77 (2011).
- Maeser, D., Gruener, R. F. & Huang, R. S. oncoPredict: An R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Brief Bioinform.* <https://doi.org/10.1093/bib/bbab260> (2021).
- Taddei, M. L., Giannoni, E., Fiaschi, T. & Chiarugi, P. Anoikis: An emerging hallmark in health and diseases. *J. Pathol.* **226**(2), 380–393 (2012).
- Li, S. *et al.* S100A8 promotes epithelial-mesenchymal transition and metastasis under TGF-beta/USF2 axis in colorectal cancer. *Cancer Commun.* **41**(2), 154–170 (2021).
- Klymchenko, A. S. & Kreder, R. Fluorescent probes for lipid rafts: From model membranes to living cells. *Chem. Biol.* **21**(1), 97–113 (2014).
- Bakir, B., Chiarella, A. M., Pitarresi, J. R. & Rustgi, A. K. EMT, MET, plasticity, and tumor metastasis. *Trends Cell Biol.* **30**(10), 764–776 (2020).
- Nirmala, J. G. & Lopus, M. Cell death mechanisms in eukaryotes. *Cell Biol. Toxicol.* **36**(2), 145–164 (2020).
- Gilmore, A. P. Anoikis. *Cell Death Differ.* **12**(Suppl 2), 1473–1477 (2005).
- Li, S. *et al.* Shear stress promotes anoikis resistance of cancer cells via caveolin-1-dependent extrinsic and intrinsic apoptotic pathways. *J. Cell. Physiol.* **234**(4), 3730–3743 (2019).
- Zhi, Z. *et al.* Non-canonical phosphorylation of Bmf by p38 MAPK promotes its apoptotic activity in anoikis. *Cell Death Differ.* **29**(2), 323–336 (2022).
- Jin, L. *et al.* Phosphorylation-mediated activation of LDHA promotes cancer cell invasion and tumour metastasis. *Oncogene* **36**(27), 3797–3806 (2017).
- Jiang, K. *et al.* MOB2 suppresses GBM cell migration and invasion via regulation of FAK/Akt and cAMP/PKA signaling. *Cell Death Dis.* **11**(4), 230 (2020).
- Kim, H. *et al.* Ginsenosides Rk1 and Rg5 inhibit transforming growth factor-beta1-induced epithelial-mesenchymal transition and suppress migration, invasion, anoikis resistance, and development of stem-like features in lung cancer. *J. Ginseng Res.* **45**(1), 134–148 (2021).
- Talukdar, S. *et al.* MDA-9/Syntenin regulates protective autophagy in anoikis-resistant glioma stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **115**(22), 5768–5773 (2018).
- Corbet, C. *et al.* TGFbeta2-induced formation of lipid droplets supports acidosis-driven EMT and the metastatic spreading of cancer cells. *Nat. Commun.* **11**(1), 454 (2020).

37. Ye, G. *et al.* Nuclear MYH9-induced CTNNB1 transcription, targeted by staurosporin, promotes gastric cancer cell anoikis resistance and metastasis. *Theranostics* **10**(17), 7545–7560 (2020).
38. Yoon, S. J. & DeNicola, G. M. IL1RAP pulls a double shift in the cysteine factory. *Cancer Discov.* **11**(11), 2679–2681 (2021).
39. Yu, Y. *et al.* CircCEMIP promotes anoikis-resistance by enhancing protective autophagy in prostate cancer cells. *J. Exp. Clin. Cancer Res.* **41**(1), 188 (2022).
40. Malagobadan, S., Ho, C. S. & Nagoor, N. H. MicroRNA-6744-5p promotes anoikis in breast cancer and directly targets NAT1 enzyme. *Cancer Biol. Med.* **17**(1), 101–111 (2020).
41. Cai, J. *et al.* NAT1 is a critical prognostic biomarker and inhibits proliferation of colorectal cancer through modulation of PI3K/Akt/mTOR. *Fut. Oncol.* **17**(19), 2489–2498 (2021).
42. Joung, E. K. *et al.* Expression of EEF1A1 is associated with prognosis of patients with colon adenocarcinoma. *J. Clin. Med.* **8**(11), 1903 (2019).
43. Yuan, Q., Zhang, W. & Shang, W. Identification and validation of a prognostic risk-scoring model based on sphingolipid metabolism-associated cluster in colon adenocarcinoma. *Front. Endocrinol. (Lausanne)* **13**, 1045167 (2022).
44. Tang, C., Qiu, S., Mou, W., Xu, J. & Wang, P. Excessive activation of HOXB13/PIMREG axis promotes hepatocellular carcinoma progression and drug resistance. *Biochem. Biophys. Res. Commun.* **623**, 81–88 (2022).
45. Wang, W., Liu, H., Wang, S., Hao, X. & Li, L. A diterpenoid derivative 15-oxospiramylactone inhibits Wnt/beta-catenin signaling and colon cancer cell tumorigenesis. *Cell Res.* **21**(5), 730–740 (2011).
46. Wu, C. *et al.* Targeting AURKA-CDC25C axis to induce synthetic lethality in ARID1A-deficient colorectal cancer cells. *Nat. Commun.* **9**(1), 3212 (2018).
47. Gao, H. & Xing, F. A novel signature model based on mitochondrial-related genes for predicting survival of colon adenocarcinoma. *BMC Med. Inform. Decis. Mak.* **22**(1), 277 (2022).
48. Toricelli, M. *et al.* Timp1 promotes cell survival by activating the PDK1 signaling pathway in melanoma. *Cancers* **9**(4), 37 (2017).
49. Macedo, F. C. *et al.* A prospective cohort study of TIMP1 as prognostic biomarker in gastric and colon cancer. *Chin. Clin. Oncol.* **11**(6), 43 (2022).
50. Niewiarowska, K. *et al.* Diagnostic significance of TIMP-1 level in serum and its immunohistochemical expression in colorectal cancer patients. *Pol. J. Pathol.* **65**(4), 296–304 (2014).
51. Song, G. *et al.* TIMP1 is a prognostic marker for the progression and metastasis of colon cancer through FAK-PI3K/AKT and MAPK pathway. *J. Exp. Clin. Cancer Res.* **35**(1), 148 (2016).
52. Koshino, A. *et al.* PBK enhances cellular proliferation with histone H3 phosphorylation and suppresses migration and invasion with CDH1 stabilization in colorectal cancer. *Front. Pharmacol.* **12**, 772926 (2021).
53. Gao, T. *et al.* Novel selective TOPK inhibitor SKLB-C05 inhibits colorectal carcinoma growth and metastasis. *Cancer Lett.* **445**, 11–23 (2019).
54. Su, T. C. *et al.* Cytoplasmic, nuclear, and total PBK/TOPK expression is associated with prognosis in colorectal cancer patients: A retrospective analysis based on immunohistochemistry stain of tissue microarrays. *PLoS One* **13**(10), e0204866 (2018).
55. Meng, Y. *et al.* Identification and validation of a novel prognostic gene model for colorectal cancer. *Comput. Math. Methods Med.* **2022**, 9774219 (2022).
56. Meneses-Morales, I. *et al.* Epigenetic regulation of the human ATP2A3 gene promoter in gastric and colon cancer cell lines. *Mol. Carcinog.* **58**(6), 887–897 (2019).
57. Flores-Peredo, L., Rodriguez, G. & Zarain-Herzberg, A. Induction of cell differentiation activates transcription of the Sarco/Endoplasmic Reticulum calcium-ATPase 3 gene (ATP2A3) in gastric and colon cancer cells. *Mol. Carcinog.* **56**(2), 735–750 (2017).
58. Brouland, J. P. *et al.* The loss of sarco/endoplasmic reticulum calcium transport ATPase 3 expression is an early event during the multistep process of colon carcinogenesis. *Am. J. Pathol.* **167**(1), 233–242 (2005).
59. Cheng, X., Wei, Y., Fu, Y., Li, J. & Han, L. A novel enterocyte-related 4-gene signature for predicting prognosis in colon adenocarcinoma. *Front. Immunol.* **13**, 1052182 (2022).

Acknowledgements

We thank the study participants and research staff for their contributions and commitment to this study.

Author contributions

G.H. takes responsibility for the content of manuscript, including the data and analysis. Conception and design: G.H. and J.L.; Development of methodology: G.H. and Y.Z.; Acquisition of data: G.H. and L.L.; Analysis and interpretation of data: G.H.; writing—review and/or revision of the manuscript: G.H., Z.Y. and X.Q.; Administration: Z.Y. and X.Q.; study supervision: G.H. and J.L. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-40907-x>.

Correspondence and requests for materials should be addressed to K.L. or H.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023