Construction of a novel prognostic model for gastric cancer based on pharmacokinetics-related genes and comprehensive prognostic analysis

Yu Zhang [ID] *, Kai Jia, Yuntong Guo, Xiaole Ma, Tian Yao, Feng Wu* and He Huang*

Department of Gastrointestinal Surgery, First Hospital of Shanxi Medical University, Taiyuan, Shanxi, China

Background: Absorption, distribution, metabolism, and excretion of drugs-related genes (ADMERGs), pivotal in cancer occurrence, development, and chemotherapy resistance, lack investigation in gastric cancer (GC). Thus, this study aims to build a prognostic model for gastric cancer utilizing ADMERGs.

Methods: The GC-related datasets, including TCGA-GC, GSE62254, GSE163558 and GSE13911, as well as 298 ADMERGs, were retrieved in this study. Prognostic risk models associated with ADME were developed utilizing univariate Cox analysis, followed by additional refinement using the least absolute shrinkage and selection operator (LASSO). The entire pool of gastric cancer (GC) patient samples was partitioned into high and low-risk categories, delineated by the median value of their respective risk scores. Within these two distinct groups, we conducted enrichment analysis, immune infiltration, and prognostic evaluation of ADME-related prognostic genes to uncover their molecular mechanisms in GC. The construction of ceRNA regulatory networks was undertaken to analyse the prognostic gene regulatory mechanisms. We analyzed single-cell data in GC to investigate the mechanisms driving its onset and progression at the cellular level. Additionally, we validated the expression trends of prognostic genes in clinical samples using RT-qPCR.

Results: A prognostic model for GC was established and validated, comprising five genes (*UGT1A1, ADH4, ADH1B, CYP19A1,* and *GPX3*). The levels of infiltration of 21 immune cells exhibited significant disparities between the two risk groups, such as central memory CD4 T cells, activated B cells, and mast cells. There was a notable positive correlation between the risk scores and mast cells and plasmacytoid dendritic cells. In the high-risk group, the TIDE scores were heightened. The single-cell dataset showed significant under-expression of *ADH1B, ADH4, CYP19A1*, and *GPX3* in tumor samples. Finally, RT-qPCR showed that all the prognostic genes except for ADH4 were under-expressed in tumor tissues.

Conclusion: We have developed and validated an innovative prognostic risk model for GC, revealing that elevated ADMERGs risk scores are indicative of unfavorable prognosis and diminished immunotherapy response. These findings

furnish molecular evidence regarding the participation of ADMERGs in modulating the immune microenvironment and therapeutic responsiveness in GC.

1 Introduction

Gastric cancer (GC) is a significant worldwide health challenge, being the fifth most prevalent form of cancer and the third leading cause of cancer-related fatalities. The disease presents an urgent threat to both public health and global economies due to its high occurrence and death rates (De Martel et al., 2020; Rawla and Barsouk, 2019). The GC is commonly correlated with environmental variables such as infection by *Helicobacter pylori*, dietary choices, and lifestyle, but the diffuse type tends to have a more aggressive clinical progression and is typically associated with genetic factors (Smyth et al., 2020). The clinical manifestation of GC exhibits considerable heterogeneity, with symptoms that span from vague dyspepsia to more severe indications, hence adding complexity to the management and prognosis (Yang et al., 2021). Advancements in the understanding of the development of GC have resulted in better methods for diagnosis and treatment. However, there are still considerable obstacles to overcome.

ADME, which stands for absorption, distribution, metabolism, and excretion, encompasses the fundamental processes in pharmacokinetics (Furuta et al., 2005). It outlines the path a medication takes from its introduction into the body to its removal. The latest progress in the research of ADME related genes (ADMERGs) has emphasized their vital functions in the regulation of drug-metabolizing enzymes, transport proteins, and nuclear receptors, hence exerting a major impact on these pharmacokinetic processes (Aitken et al., 2008; Fletcher et al., 2010). ADMERGs have been found to have important roles in the genesis and progression of GC. Genetic variations in specific metabolic enzyme genes and transport protein genes can have an influence on the way chemotherapeutic medications are metabolized and distributed in the body. This, in turn, can have an impact on the effectiveness of treatment and the development of drug resistance. Furthermore, the presence of ADMERGs can impact the availability and toxicity of drugs, ultimately affecting how patients respond to treatment and their overall prognosis (Rodrigues, 2022). While research has revealed certain functions of ADMERGs in GC, the specific methods by which they operate are still not well understood. Hence, it is imperative to do additional study on the correlation between ADMERGs and the prognosis of GC to develop more effective tailored treatments.

In this paper, differentially expressed genes (DEGs) associated with ADME were found using the TCGA-GC dataset from the TCGA database, and candidate genes were obtained by taking intersections with ADEMRGs. Prognostic genes closely associated with GC outcome were identified by univariate Cox regression analysis and Least Absolute Shrinkage and Selection Operator (LASSO) regression. Subsequently, a predictive risk model was created using prognostic genes to classify patients based on risk scores. In addition, a nomogram was created to accurately predict the prognosis of GC patients. To investigate the biological mechanisms of these prognostic genes, enrichment analysis,

immune infiltration, drug sensitivity analysis, ceRNA network construction, and analysis of prognostic gene expression patterns at the single-cell level were also performed. The findings will provide novel insights into the molecular mechanisms behind GC and propose prospective targets for therapeutic intervention, as well as serve as a helpful reference for future research on the etiology and treatment strategies of GC.

2 Materials and methods

2.1 Data sources

For the development of a predictive model for GC, we accessed transcriptome data and corresponding clinical pathology data from The Cancer Genome Atlas (TCGA, http://portal.gdc.cancer.gov/). This dataset consists of 375 tumor samples and 32 normal samples and serves as the foundational training set, of these, 350 tumor samples had survival information. The datasets GSE62254 (GPL570) including 300 tumor tissue samples and GSE13911 including 69 samples were used as the validation set (Cristescu et al., 2015; Oh et al., 2018) and were obtained from Gene Expression Omnibus database (GEO, https://www.ncbi.nlm.nih.gov/geo/). A collection of 298 ADMERGs was sourced from the PharmaADME Consortium (http://www.pharmaadme.org) (Tang et al., 2022).

2.2 Identification and enrichment analysis of candidate genes

We employed the R package DESeq2 (version 1.34.0) to carry out differential expression analysis between tumor and normal samples within our training set. Genes that exhibited a | $\log_2$FoldChange| > 2 and an adj.P-value <0.05 were identified as significant and were selected based on these criteria (Love et al., 2014). To obtain candidate genes, the intersecting set of DEGs and ADMERGs was determined using the R package ggvenn (version 1.2.2). Enrichment analysis, which encompasses Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG), was conducted using the R package clusterProfiler (version 3.18.1), with an adjusted p-value threshold of less than 0.05 (Yu et al., 2012). The Oncobox database was utilized to assess the KEGG pathway activation level of candidate genes, adj.P < 0.05 was used as the screening threshold.

2.3 A prognostic risk model linked to ADME was constructed and validated

The candidate genes associated with prognosis were identified using univariate Cox regression analyses, employing the glmnet package (version 4.1-2) (HR ≠ 1 & P < 0.05), using 10-fold cross-

validation, GC samples were randomly and equally divided into 10 subsets. In each iteration, one subset was selected as the validation set, while the remaining nine subsets served as the training set. The LASSO-Cox model was fitted using the training set, and the Cox deviation of the model was evaluated on the validation set. After 10 iterations, the average deviations of all λ values on the validation sets were compared, and λ.min, which minimized the cross-validation error, was selected as the final regularization parameter. Finally, the optimal λ value was determined, and the final LASSO-Cox model was constructed using all training data to screen out key prognostic gene features. The proportional hazards (PH) assumption was then employed to verify these genes, using a p-value threshold of greater than 0.05. Subsequently, we utilized LASSO regression analysis from the R package glmnet to determine the final prognostic genes (Friedman et al., 2010). Subsequently, a risk score was calculated for each GC patient, which was based on the expression (expr) and risk coefficients (coef) of these prognostic genes, as follows:

$$Riskscore = \sum_{i=1}^{n} coef\left(gene_i\right) * expr\left(gene_i\right)$$

Patients with GC were categorized into either a high- or low-risk groups based on the median risk score. Additionally, KM curves (P < 0.05) were plotted according to the high and low risk groups using survminer (version 0.4.9) (Liu et al., 2021) to determine the difference in survival between the high and low risk groups. The receiver operating characteristic (ROC) curves (2, 3, 4, 5, 6 years) were generated utilizing the survival and survival ROC packages (version 1.0.3.1) (Maeser et al., 2021) to assess the model's accuracy (AUC >0.6). In the validation set, the same method was employed. Subsequently, the cor function from the R package "stats" was employed to assess correlations between prognostic genes and established GC (gastric cancer) prognostic markers, including MSI (microsatellite instability), PD-L1 expression, and HER2 status, using both Spearman and Pearson correlation analyses.

2.4 Correlation between clinical features and risk score

The utilization of the Wilcoxon rank-sum test (P < 0.05) allowed for the evaluation of variations in risk scores among distinct clinical subgroups and the investigation of patient distribution into high and low-risk groups across these subgroups, taking into account clinical factors including Age, Gender, T/N/M stage, and Stage.

2.5 Establishment and validation of a nomogram

Via univariate and multivariate Cox regression analyses utilizing the R package survival (P > 0.05; HR≠1), identified independent prognostic factors influencing GC, with the results confirmed by the PH hypothesis test (P > 0.05). Subsequently, a prognostic nomogram was developed employing the rms package (version 6.2-0) (Luo et al., 2024). This was followed by the generation of decision curve analysis (DCA), calibration curves and ROC curves to gauge the accuracy and reliability of the nomogram.

2.6 Gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA)

Across risk groups, GSEA was executed via the R package clusterProfiler, employing c2.cp.kegg.V7.0.symbols.gmt as the reference gene set. Meanwhile, the analysis applied a stringent significance threshold of P < 0.05 and a False Discovery Rate (FDR) of <0.25 for the screening process. GSVA was performed utilizing the R package "GSVA" (version 1.42.0) (Hänzelmann et al., 2013). To solidify the findings, a Spearman correlation analysis was meticulously conducted to elucidate the intricate relationship between clinical manifestations, risk scores, and biological pathways.

2.7 Immune cell infiltration analysis

In the training set, employing single-sample GSEA (ssGSEA) to assess the abundance of 28 immune infiltrating cells and 2 stromal cells. Following this, a stringent Wilcoxon test (with a significance threshold of P < 0.05) was applied with a significance threshold of P < 0.05 to identify significant differences in the abundance of immune infiltrating cells between two groups. Furthermore, to delve deeper into the association between risk score and differentially infiltrating cells, a Spearman correlation analysis was conducted (with |R| > 0.4 and P < 0.05). Next, the cor function from the R package stats was used to perform Pearson correlation analysis between prognostic genes and differentially infiltrated immune cell subsets.

2.8 Immunotherapy and prediction of chemotherapy response

We assessed the varying expression of 15 m6A RNA methylation genes and 24 immune checkpoint genes between two risk groups, aiming to forecast the potential effectiveness of immune checkpoint blockade (ICB) in patients with different risk levels (P < 0.05). Additionally, we calculated the Tumor Immune Dysfunction and Exclusion (TIDE) score for each GC patient using the TIDE database (http://tide.dfci.harvard.edu/). To further predict the clinical response to chemotherapy, we evaluated the sensitivity of drugs using the R package pRRophetic (version 0.5) (Geeleher et al., 2014) (P < 0.05). Next, the cor function and cor.test significance test built into R language were used, and the Spearman correlation analysis method was adopted to calculate the correlation between prognostic genes and immune checkpoint genes (p < 0.05).

2.9 The ceRNA regulatory network analysis

To investigate the molecular regulatory mechanisms between high-risk and low-risk groups, in the training set, utilizing DESeq2 (v.1.34.0) (Love et al., 2014) to identify differentially

expressed miRNAs (miRNA1), and lncRNAs (lncRNA1) between GC high-risk group and low-risk group ($|\log_2 FoldChange| > 1$ and adj.P.value<0.05). Additionally, miRDB (http://www.mirdb.org/miRDB/policy.html) databases were utilized to predict prognostic genes miRNAs (miRNA2). Key miRNAs were obtained by intersecting miRNA1 and miRNA2 using the package ggvenn. Furthermore, upstream lncRNAs (lncRNA2) of key miRNAs were predicted using the miRNet(https://www.mirnet.ca) databases. Key lncRNAs were obtained by intersecting lncRNA1 and lncRNA2 with the ggvenn package. Finally, we constructed a ceRNA regulatory network using these key molecules.

2.10 The scRNA-seq data analysis

The scRNA-seq data underwent quality control measures implemented through the R package Seurat (version 4.1.0) (Hao et al., 2021). Exclusion criteria as follows: cells with less than 200 genes expressed, genes detected in fewer than 3 cells, and cells with a mitochondrial gene proportion above 20%. Additionally, cells expressing less than 101 or more than 6000 genes were filtered out. Subsequently, the Seurat package's integrated was then performed to correct for batch effects. To reduce dimensionality and cluster cells, we used the RunPCA function with the t-SNE algorithm. Following clustering, we used the FindAllMarkers function and the singleR package (Aran et al., 2019) to annotate each cell. After annotation to obtain the cell types, a bar stacking plot was drawn to show the distribution of different cell types in tumor and normal samples, as well as UMAP plots and prognostic genes in each cell type distribution and expression in each cell type. Finally, the Wilcoxon test was used to compare the differences in the expression of prognostic genes between gastric cancer tumor tissue samples and normal samples, as well as the expression of biomarkers in annotated cell types.

2.11 Reverse transcription quantitative PCR (RT-qPCR)

From the Department of gastrointestinal surgery in First Hospital of Shanxi Medical University, we collected the tumor (n = 5) and para-carcinoma tissues (n = 5) of patients with GC. All donor patients provided and signed off on the informed consent, which was approved by the Ethics Committee of our hospital (Ethics Review No:KYLL-2024-075). The total RNA of all samples was isolated in terms of TRIzol Reagent (Ambion, Shanghai, China). Subsequently, total RNA was used to reverse transcription via the SweScript First Strand cDNA synthesis kit (Servicebio, Wuhan, China). Then, the primers of prognostic genes were shown in Supplementary Table S1. The qPCR was proceeded using the Universal Blue SYBR Green qPCR Master Mix (Servicebio, Wuhan, China) on the CFX96TM PCR System (BIO-RAD, U.S.A.). The relative expression of these prognostic genes was calculated based on the $2^{-\Delta\Delta CT}$ method (Rao et al., 2013), with an endogenous control GAPDH.

2.12 Statistical analysis

All bioinformatics analyses were performed using the R language (version 4.2.2). P < 0.05 was considered meaningful and significant.
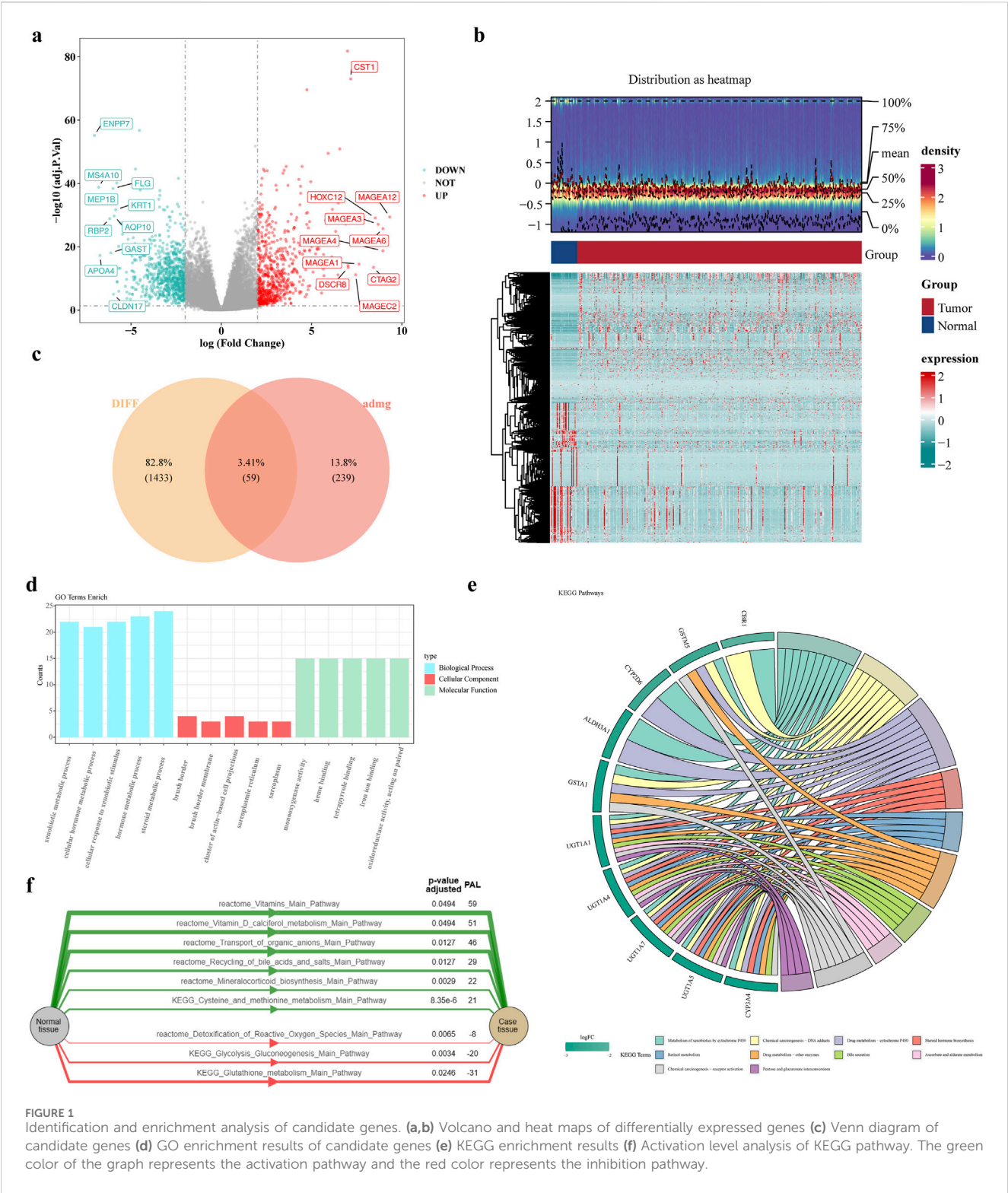
3 Results

3.1 Identification and enrichment analysis of candidate genes

A total of 1,492 DEGs were identified, with 683 upregulated and 809 downregulated in the GC patients (Figures 1a,b). Moreover, 59 candidate gene were obtained by intersecting ADMERGs and DEGs using a Venn diagram (Figure 1c). Following this, candidate genes GO enrichment analysis results showed, the enriched biological process categories included cellular hormone metabolic process, cellular response to xenobiotic stimulus, hormone metabolic process, xenobiotic metabolic process, and steroid metabolic process. The primary enriched cellular component categories included brush border membrane, sarcoplasmic reticulum, brush border, sarcoplasm, cluster of actin-based cell projections. The primary enriched molecular function categories included monooxygenase activity, tetrapyrrole binding, and so on (Figure 1d). The results of the KEGG enrichment analysis indicated that the candidate genes were predominantly enriched in metabolic pathways, metabolism of xenobiotics by chemical carcinogenesis-DNA adducts, drug metabolism-cytochrome P450, retinol metabolism, and steroid hormone biosynthesis (Figure 1e). Pathway activation level analysis showed that six pathways exhibited significant activation in GC samples, include vitamin, transport of organic anions, recycling of bile acids and salts, mineralocorticoid biosynthesis, cysteine and methionine metabolism main pathway. And three pathways were significantly inhibited, including detoxification of reactive oxygen species, glycolysis gluconeogenesis, glutathione metabolism main pathway (Figure 1f).

3.2 Construction of a prognostic models related to ADME

To obtain genes associated with GC prognosis in the training set, using the univariate Cox regression analysis and verified by the PH hypothesis test, resulting in the screening out of 5 genes associated with survival (Figure 2a). Subsequently, we utilized the LASSO for further refinement, resulting in the identification of 5 prognostic genes: UGT1A1, ADH4, ADH1B, CYP19A1, GPX3 (Figures 2b,c). Analyze the correlation between these prognostic genes and known GC prognostic markers such as MSI, PD-L1, HER2. The results showed that UGT1A1, ADH4 were significantly negatively correlated with PD-L1, and UGT1A1, GPX3, ADH1B were significantly correlated with HER2 (Supplementary Figure S1). Among them, UGT1A1 was significantly positively correlated with HER2, GPX3, ADH1B were significantly negatively correlated with HER2, and GPX3, ADH1B, ADH4 were significantly negatively correlated with MSI. The ADMERGs

FIGURE 1
Identification and enrichment analysis of candidate genes. **(a,b)** Volcano and heat maps of differentially expressed genes **(c)** Venn diagram of candidate genes **(d)** GO enrichment results of candidate genes **(e)** KEGG enrichment results **(f)** Activation level analysis of KEGG pathway. The green color of the graph represents the activation pathway and the red color represents the inhibition pathway.

plays a dual role in metabolic reprogramming and immune microenvironment regulation, significantly affecting key prognostic markers of gastric cancer (PD-L1, HER2, MSI). UGT1A1, GPX3, ADH1B, and ADH4 can serve as cross pathway regulatory nodes and can be integrated into a "metabolic immune prognostic model" in the future to optimize precision treatment strategies for gastric cancer.

3.3 Assessment of prognostic characteristics

Risk scores were computed for all patient samples in the training set, leading to their classification into either a high-risk or low-risk group, rested on the median value. The high-risk group exhibited decreased survival rates and less survival times in contrast to the low-risk group. Additionally, a notable rise in the number of deaths
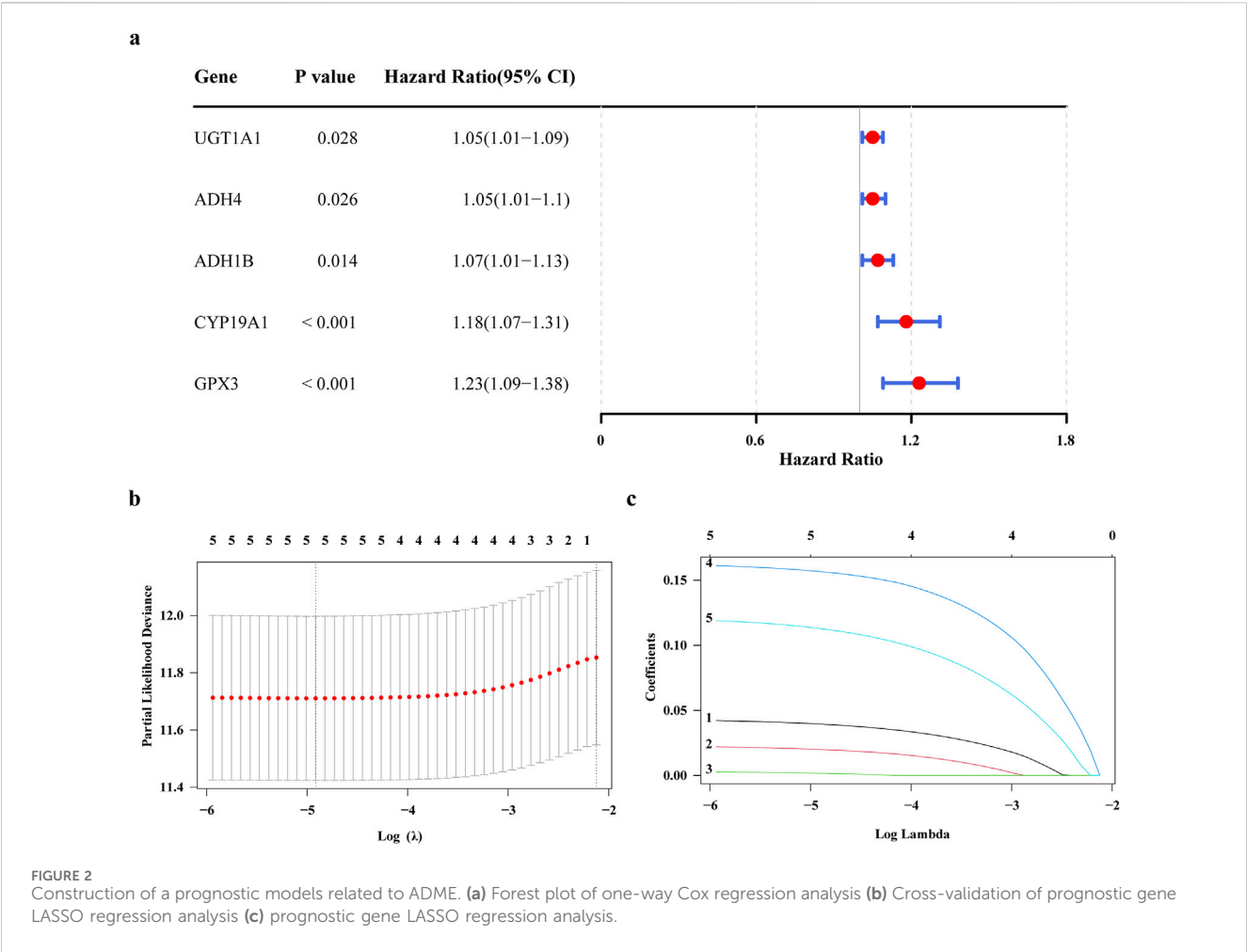
FIGURE 2
Construction of a prognostic models related to ADME. **(a)** Forest plot of one-way Cox regression analysis **(b)** Cross-validation of prognostic gene LASSO regression analysis **(c)** prognostic gene LASSO regression analysis.

was observed with increasing risk scores among the samples (Figures 3a,b). Afterward, KM curve showed significantly lower survival in the high-risk group than in the low-risk group, as illustrated in Figure 3c (P < 0.005). In the training set, the area under curve (AUC) values for the risk score were 0.64, 0.63, 0.69, 0.69, and 0.77 at 2, 3, 4, 5, and 6 years, respectively, as presented in Figure 3d. The same method was employed to verify the model's accuracy and applicability in the test set, with the risk survival status of GC patients was also displayed (Figures 3e–g). The AUC values were 0.59, 0.62, 0.62, 0.63, and 0.64 at 2, 3, 4, 5 and 6 years, in the test set (Figure 3h), suggesting that the risk model has good predictive power.

3.4 Correlation of risk score and other clinicopathological features

To investigate the involvement of risk score in GC, we assessed their association with clinical characteristics. Utilizing the clinical data obtained from TCGA, GC patients were categorized into distinct subgroups. Our analysis unveiled noteworthy disparities in risk scores between the M0 and M1 groups (Figures 4a–c).

3.5 The nomogram was constructed and validated

Risk score, age, and stage were significant independent predictors of patient outcome (P < 0.05) (Figures 5a–c). We created a nomogram based on these independent prognostic factors (Figure 5d). Figure 5e exhibits the calibration curves for the nomogram, illustrating the probability of Overall Survival (OS) at 2, 3, 4, 5, and 6 years. In addition, the ROC and DCA curve was generated to evaluate the nomogram's accuracy (Figures 5f,g). The AUC values for the nomogram surpassed 0.6 across 2, 3, 4, 5, and 6 years, suggesting a positive predictive performance. All AUC values were greater than 0.6 and the decision curve of the nomogram was higher than any of the independent prognostic factors, suggesting that a nomogram combining risk scores with clinical aspects is valuable for the diagnosis of GC.

3.6 GSEA and GSVA analysis

In this study, we identified 48 pathways that were noticeably abundant between the two groups. The high-risk group exhibited
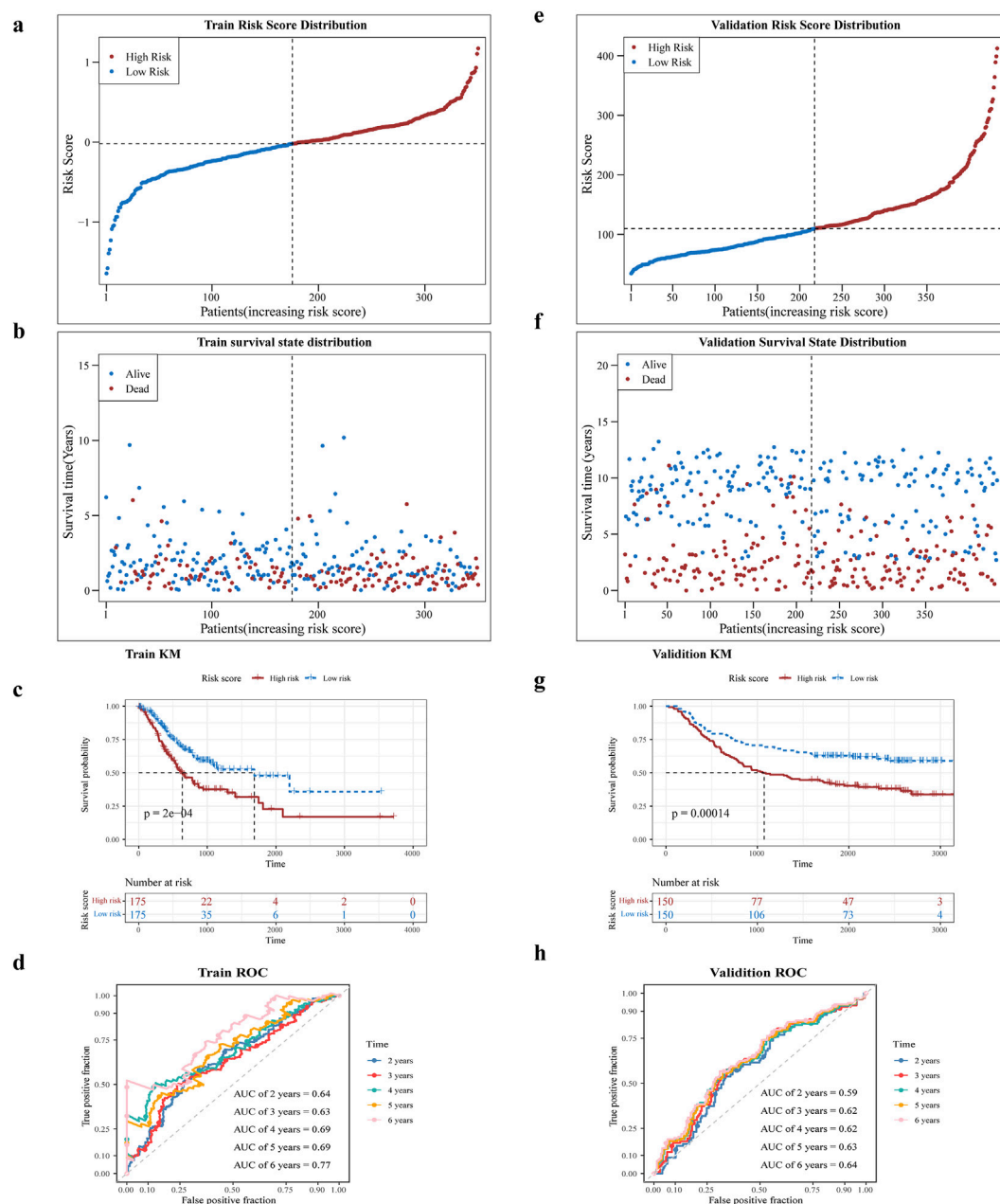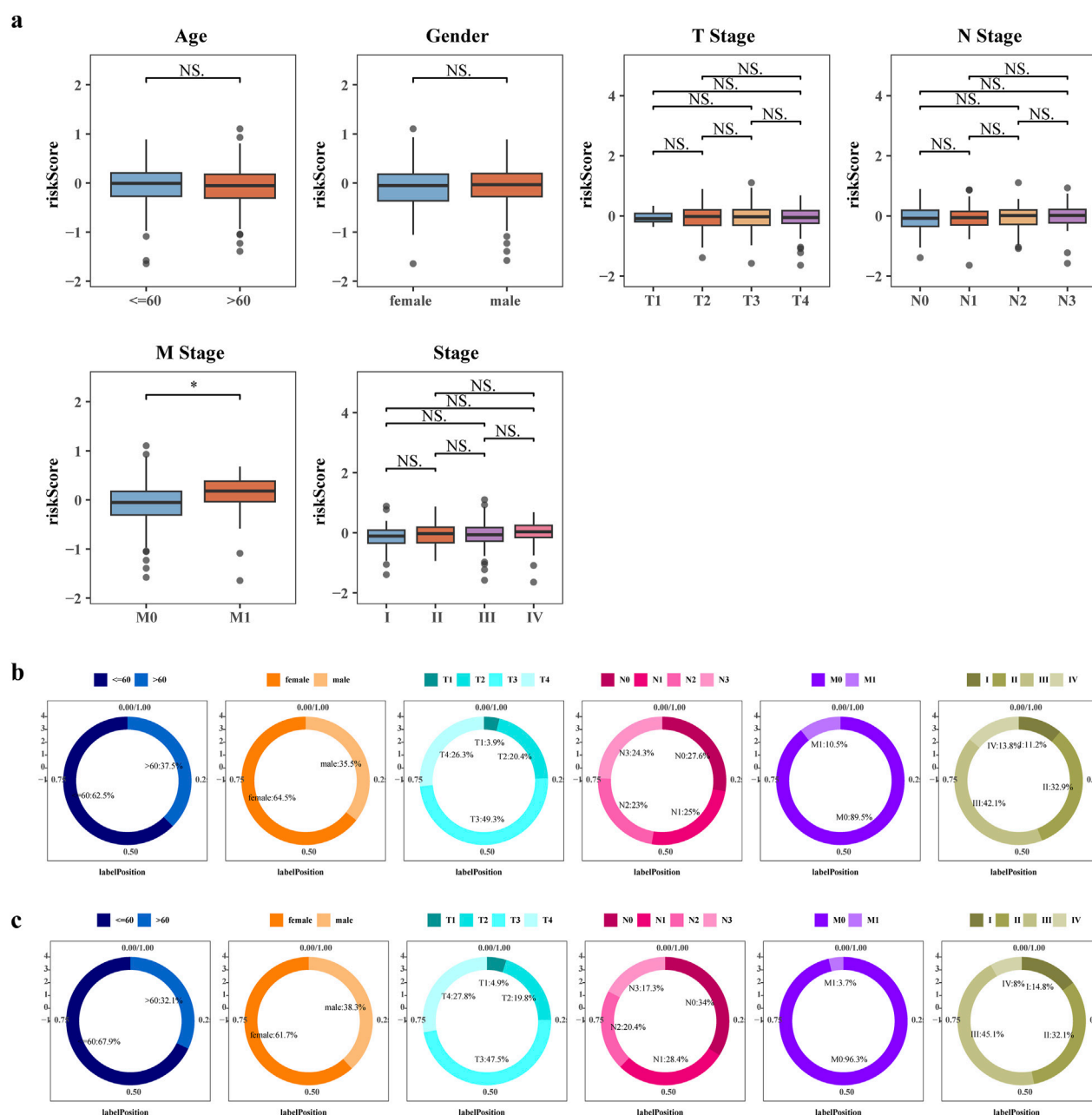
FIGURE 3
Assessment of prognostic characteristics. **(a,b)** Risk profiles of the training set GC patients in high and low risk subgroups will and scatter plots. **(c)** Prognostic gene K-M survival analysis. The horizontal axis is the total survival time (days) and the vertical axis is the survival probability; red color represents the high-risk group and blue color represents the low-risk group. **(d)** ROC curves of training set GC patients at 2, 3, 4, 5, and 6 years **(e,f)** Risk curves and scatter plots of high- and low-risk subgroups of validation set GSE62254 samples. **(g)** K-M survival analysis of high- and low-risk subgroups of validation set GSE62254 samples **(h)** ROC curves of validation set GSE62254 patients at 2, 3, 4, 5, and 6 years.

significant enrichment in pathways such as the calcium signaling pathway and complement and coagulation cascade. Conversely, the low-risk group showed significant enrichment in pathways including DNA replication, spliceosome, ribosome, nucleotide excision repair, and the cell cycle, (Figure 6a). Figure 6b illustrates the correlation among risk scores, clinical features, and typical biological pathways. This finding suggested a substantial positive correlation between the risk score and the type II interferon response, as shown in Figure 6c.

3.7 Immune microenvironment analysis

The findings revealed significant differences in abundance for 21 immune cells and 2 stromal cells (fibroblasts and endothelial cells) between the two risk groups, for instance effector memory CD8 T cell, activated B cell, central memory CD4 T cell, central memory CD8 T cell, etc., (Figure 7a). In addition, a positive relationship was found between the risk score and the abundance of cell types including mast cells, plasmacytoid dendritic cells, and effector

FIGURE 4
Correlation of risk score and other clinicopathological features. **(a)** Boxplot of differences in risk scores between clinical subgroups. Differences in the distribution of patients in the high risk groups **(b)** and low risk groups **(c)** between clinical subgroups.

memory CD4 T cells (Figure 7b). Subsequently, correlation analysis was conducted between prognostic genes and differential immune cells. The results showed that ADH1B, GPX3, and most immune cells had strong positive correlations, while UGT1A1, ADH4, CYP19A1 had weak correlations with some immune cells (Supplementary Figure S2).

Figure 7c demonstrates substantial variances in the expression levels of 24 immune checkpoints among the two risk groups, and overexpressed in the high-risk group. Comparing m6A-related genes (m6A-RGs), 15 differentially expressed genes were identified. Except for FTO and IGF2BP1, the rest of the differential m6A genes had increased expression in the low-risk group (Figure 7d). Subsequently, we analyzed correlations between prognostic and immune checkpoint genes. Results

showed: UGT1A1 was significantly negatively correlated with PDCD1, CD274, PDCD1LG2, CTLA4, and LAG3; ADH4 was significantly negatively correlated with CD274; ADH1B was significantly positively correlated with TIGIT and PDCD1LG2, negatively with CD274; CYP19A1 was significantly positively correlated with HAVCR2, negatively with LAG3; GPX3 was significantly positively correlated with HAVCR2, LAG3, PDCD1LG2, and PDCD1 (Supplementary Figure S3). This indicates that these prognostic genes are deeply involved in the balance between the activity and inhibition of the immune microenvironment, providing new strategies for precise stratification and targeted metabolic sensitization in tumor immunotherapy.
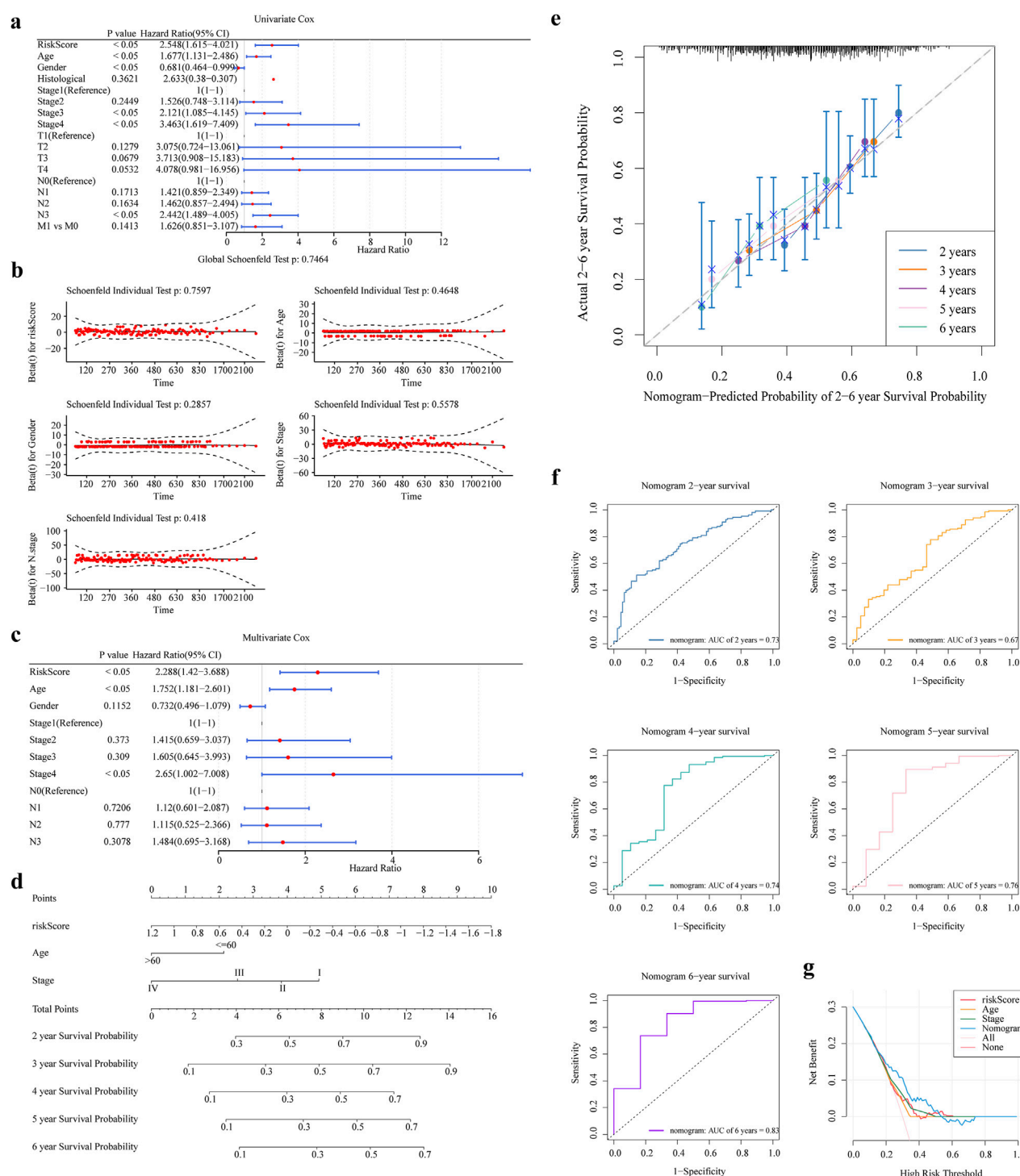
FIGURE 5
The nomogram was constructed and validated. **(a,b)** One-way Cox Analysis and PH Hypothesis Test **(c)** Results of Multifactor Cox Analysis **(d)** Nomogram Columns **(e)** Columns 2, 3, 4, 5, and 6 Years Calibration Curves. **(f,g)** Columns 2, 3, 4, 5, and 6 Years ROC and DCA Curves.

3.8 Immunotherapy and prediction of chemotherapy response

This research aimed to explore the utility of the risk score in informing treatment strategy decisions for patients. We found that

the low-risk group demonstrated enhanced responsiveness to the drugs QS11 and ABT.888, whereas the high-risk category showed increased sensitivity for DMOG and VX.702 ($P < 0.05$) (Figures 8a–c). The Spearman association analysis demonstrated a positive correlation between BIBW2992 and the risk score, as well as a

FIGURE 6
GSEA and GSVA analysis. **(a)** GSEA enrichment analysis of differential genes in high and low risk groups **(b)** ssGSEA biological pathway analysis **(c)** Correlation analysis of risk scores and biological pathways.

FIGURE 7
Immune microenvironment analysis. **(a)** Differential violin plots of immune infiltrating cells in samples from high and low risk groups **(b)** Correlation analysis between risk scores and differential immune cells **(c)** Differential violin plots of immune checkpoints in samples from high and low risk groups **(d)** Differential violin plots of m6A-related genes in samples from high and low risk groups.
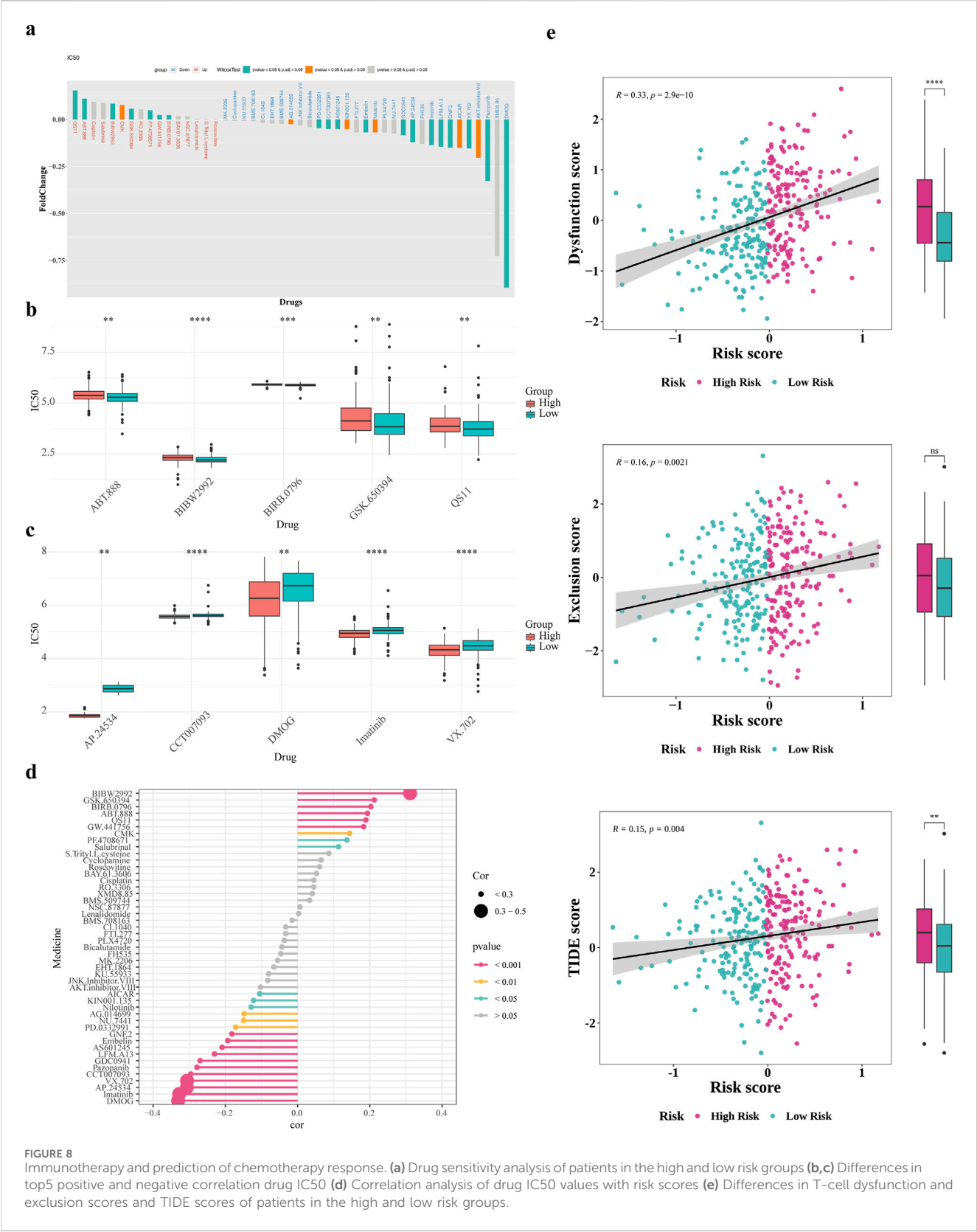
negative correlation with DMOG, Imatinib, AP.24534, and VX.702. (Figure 8d). Besides, patients in the high-risk group revealed significantly higher Dysfunction and TIDE scores compared to those in the low-risk group (Figure 8e).

3.9 The ceRNA regulatory network analysis

In the training set, DESeq2 identified differentially expressed miRNA1 and lncRNA1 between GC high-risk

FIGURE 8
Immunotherapy and prediction of chemotherapy response. **(a)** Drug sensitivity analysis of patients in the high and low risk groups **(b,c)** Differences in top5 positive and negative correlation drug IC50 **(d)** Correlation analysis of drug IC50 values with risk scores **(e)** Differences in T-cell dysfunction and exclusion scores and TIDE scores of patients in the high and low risk groups.

group and low-risk group (Figures 9a–d). Using miRDB databases were predicted miRNA2, the Venn diagram illustrated the identification of four key miRNAs, mir590,

mir7152, mir4420 and mir5000 were obtained (Figure 9e). The lncRNA2 of the key miRNAs was predicted using the miRNet databases, lncRNA1 and lncRNA2 were intersected
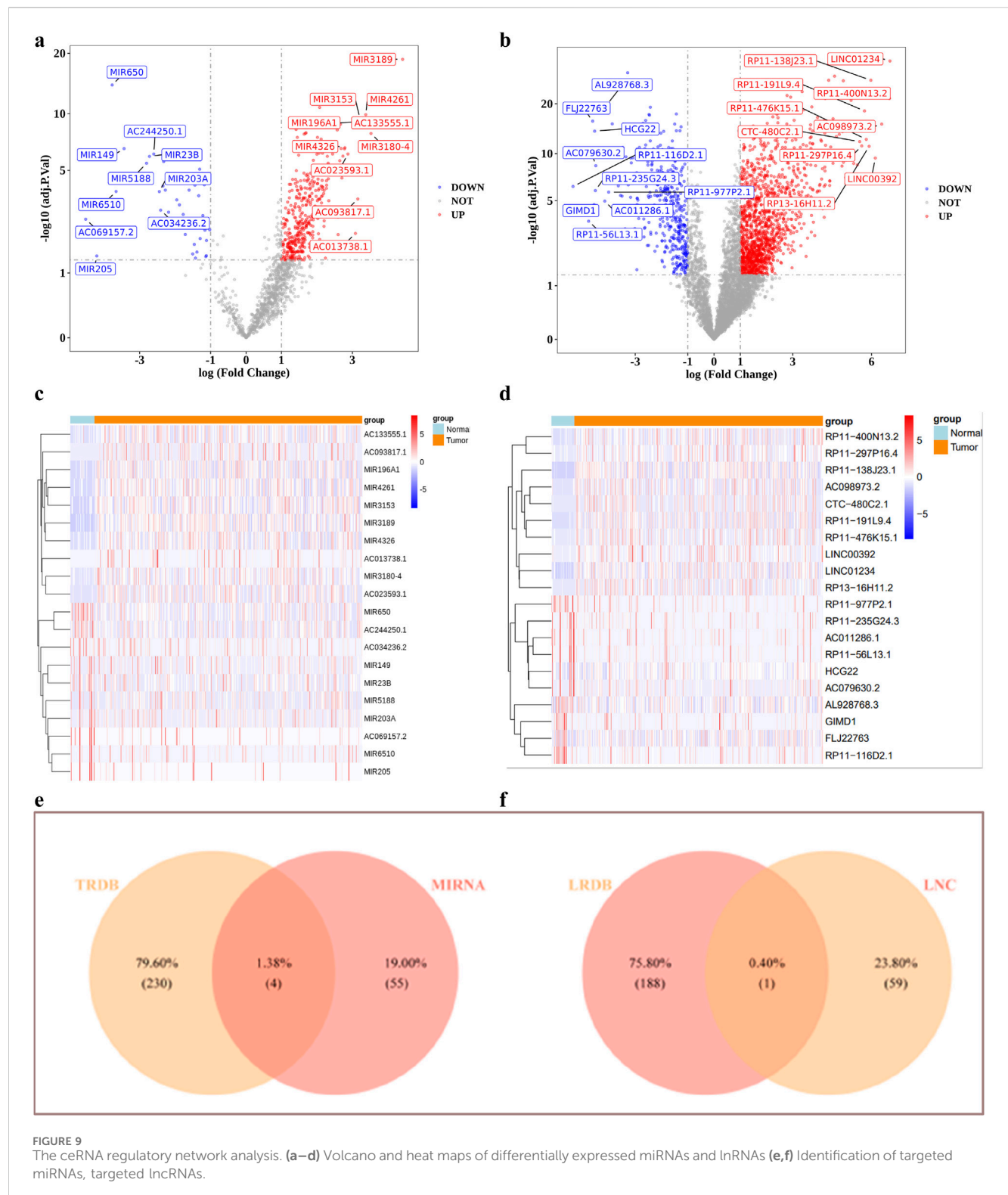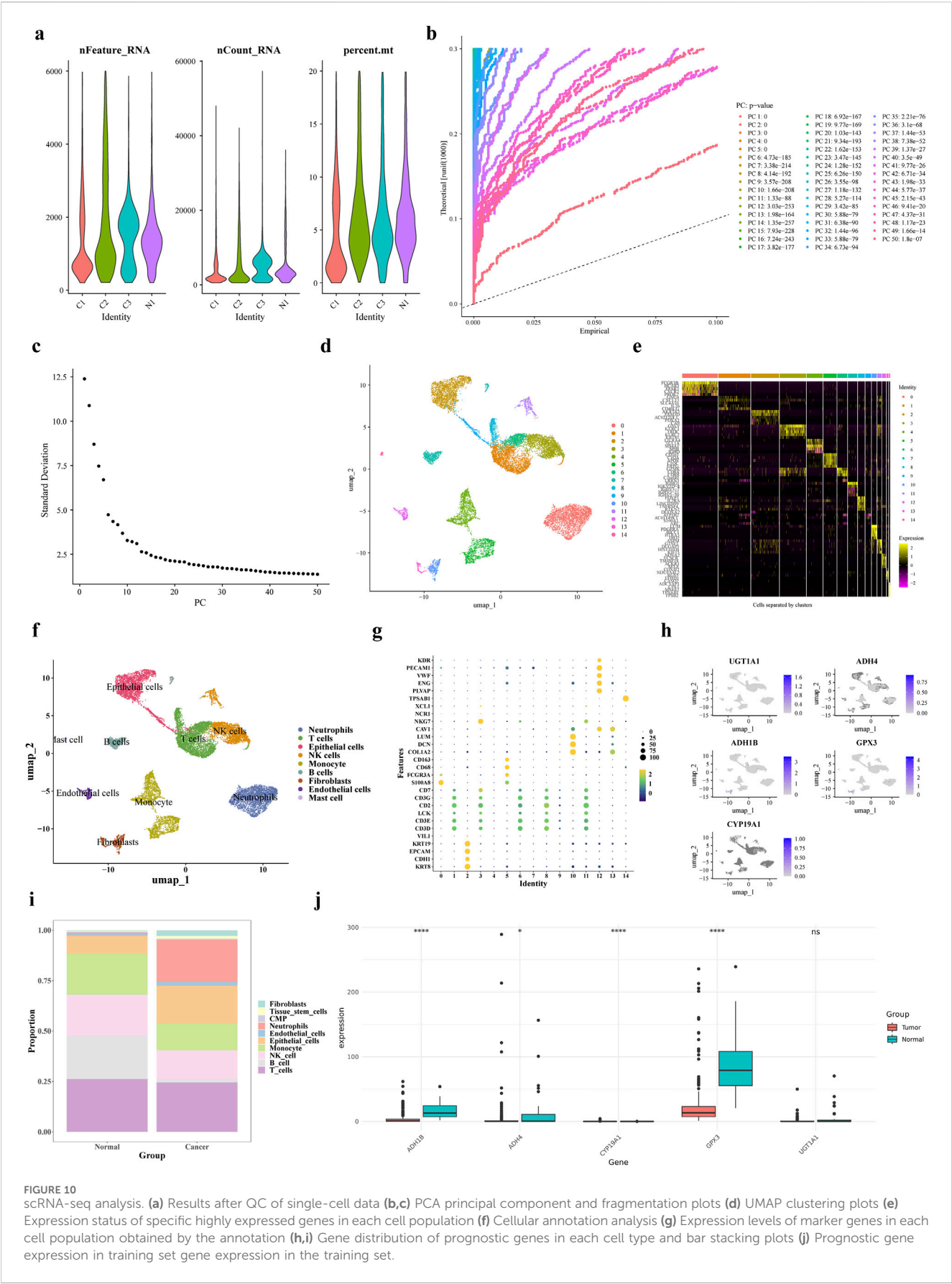
FIGURE 9
The ceRNA regulatory network analysis. (a–d) Volcano and heat maps of differentially expressed miRNAs and lnRNAs (e,f) Identification of targeted miRNAs, targeted lncRNAs.

to obtain a key lncRNA (MIR100HG) (Figure 9f). Among them, mir590, mir7152, mir4420 and mir5000 were significantly upregulated in the GC high-risk group (P < 0.05). Additionally, MIR100HG lncRNA was also significantly upregulated in the GC high-risk group (P < 0.05) (Supplementary Tables S2, S3).

3.10 scRNA-seq analysis

The scRNA-seq data were processed using the R package "Seurat" for filtering. Then calculated nFeature-RNA, nCount-RNA and percent.mt (Figure 10a). The top 50 principal components with statistical significance from the PCA analysis

FIGURE 10
scRNA-seq analysis. **(a)** Results after QC of single-cell data **(b,c)** PCA principal component and fragmentation plots **(d)** UMAP clustering plots **(e)** Expression status of specific highly expressed genes in each cell population **(f)** Cellular annotation analysis **(g)** Expression levels of marker genes in each cell population obtained by the annotation **(h,i)** Gene distribution of prognostic genes in each cell type and bar stacking plots **(j)** Prognostic gene expression in training set gene expression in the training set.

were chosen for further analysis using UMAP the clustering analysis identified 15 distinct cell clusters (Figures 10b–d). We annotated 9 cell clusters based on marker genes from the CellMarker database and the R package "singleR": T cells (CD3D, CD3E, LCK, CD2, CD3G, CD7), B cells (identified by SingleR auxiliary comments), Monocytes, NK cells, intermediate granule cells, endothelial cells, fibroblastic vascular cells, epithelial cells and mast cells (TPSAB1) (Figures 10e–g).

Subsequently, we conducted an analysis of each cell type proportion in the GC samples, as illustrated in Figures 10h,i. The T cells proportion was the largest in both tumor and control groups, while fibroblasts had the lowest proportion. In tumor tissues, neutrophil proportion was higher and B cells and T cells were lower compared to normal cells. The levels of expression for ADH1B, ADH4, CYP19A1, and GPX3 were notably decreased in tumor samples from the single-cell dataset, as depicted in Figure 10j. Subsequently, the expression levels of prognostic genes were compared and validated using the GSE13911 dataset, and the results showed that the expression trends of four prognostic genes, ADH1B, ADH4, CYP19A1, and GPX3, were consistent with them (Supplementary Figure S4). The Wilcoxon test was used to compare the expression of prognostic genes between GC tumor tissues and control samples in the TCGA training set and GSE13911 dataset. Results showed that four prognostic genes (ADH1B, ADH4, UGT1A1, and GPX3) exhibited significantly lower expression in tumor samples compared to normal samples (Supplementary Figure S5a). However, CYP19A1 showed upregulated expression in the TCGA training set but downregulated expression in the GSE13911 validation set, which might be attributed to differences in dataset characteristics or sample heterogeneity (Supplementary Figure S5b). Additionally, we analyzed the differential expression of prognostic genes across distinct cell types. As shown in Supplementary Figures S5a,b, GPX3 and ADH1B were more highly expressed in Stromal cells compared to other cell types. The Wilcoxon test was used to compare the differential expression of prognostic genes in annotated cell types between disease and normal physiological conditions. UGT1A1 showed significant expression differences between disease and normal conditions in T cells and Monocytes; ADH4 exhibited significant differences in Monocytes, B cells, Epithelial_cells, and Stromal cells; GPX3 showed significant variations in T_cells, Plasma, Epithelial_cells, and Stromal cells; ADH1B had significant differences in Plasma, Stromal cells, and Epithelial_cells; and CYP19A1 displayed significant expression differences in Stromal cells between disease and normal conditions. (Supplementary Figures S5c–g).
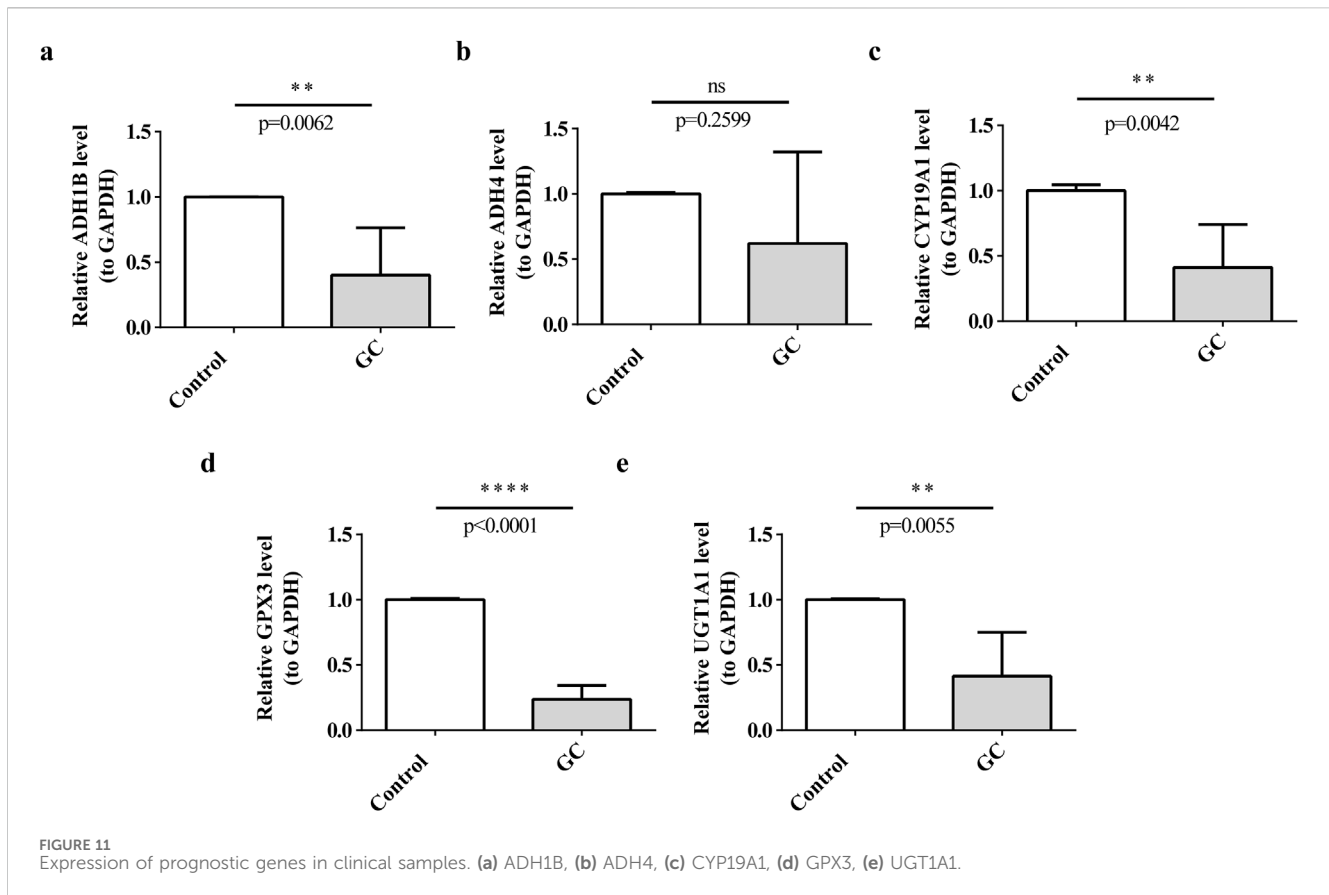
3.11 Expression level verification of prognostic genes

To explore the expression differences of prognostic genes in the clinic samples, the tumors and para-carcinoma tissues were collected for RT-qPCR. Interestingly, all the prognostic genes except for ADH4 were under-expressed in tumors (P < 0.05), and there was no significant expression difference of ADH4 between the tumors and para-carcinoma tissues (Figures 11a–e).

4 Discussion

GC is a leading cause of cancer deaths worldwide, and ADMERGs are increasingly implicated in its progression and therapy. Studying ADMERGs effect on GC can help predict patient prognosis and therapy options. This study found differentially expressed ADMERGs in combination with the TCGA-GC dataset. Five prognostic genes (UGT1A1, ADH4, ADH1B, CYP19A1, and GPX3) were then identified using univariate Cox regression and LASSO regression. Subsequently, Risk scores for GC patients were then obtained based on the prognostic genes and combined with a clinically created nomogram. In recent years, numerous studies have attempted to construct prognostic models for gastric cancer (GC) based on multi-omics features, covering multiple research dimensions such as immune-related genes, m6A modifications, metabolic regulation, and ferroptosis mechanisms. For example, Ma et al. developed a model based on differentially expressed immune-related genes (DEIRGs) to predict overall survival in GC patients and assess immune infiltration levels, though it did not account for the influence of metabolic or drug response pathways (Ma et al., 2024). Similarly, Peng et al. established a prognostic model for the GC tumor microenvironment based on m6A regulator-related genes (Peng et al., 2024). Feng et al. analyzed six tumor-associated metabolic pathways. Wen et al. developed a gastric cancer (GC) model based on ferroptosis-related genes (comprising six genes), which demonstrated prognostic value for patient survival but lacked in-depth investigation into immune phenotypes. Additionally, some studies have constructed models based on different cancer stages. For instance, Liu et al. developed prognostic models for early- and late-stage cancers. Compared to these models, the prognostic model constructed in this study based on five ADME-related genes not only more accurately predicts patient survival by incorporating covariates such as age and clinical stage (T, N, M), thereby providing a basis for developing personalized treatment strategies for gastric cancer (GC) at different stages, but also explores differences in genomic variation, immune microenvironment, and drug sensitivity, investigates the molecular regulatory mechanisms of prognostic genes, and observes the distribution and expression patterns of these genes in single-cell datasets, offering new reference evidence and theoretical support for GC treatment.

UDP glucuronosyltransferase family 1 member A1(UGT1A1) helps detoxify and eliminate toxins from the body and outside sources by glucuronidation. Chemical degradation associated to UGT1A1 expression alterations causes digestive tract cancer. Several studies suggested that UGT1A1 glucuronidases cancer-causing compounds to make them water-soluble and excretable. Chemicals that cause stomach lining cancer are broken down by UGT1A1. And UGT1A1 expression or activity may alter stomach lining cancer-causing material elimination and GC risk. Chemical accumulation from UGT1A1 dysfunction promotes DNA damage and tumor risk, while genetic variations in UGT1A1 can affect GC risk via modifying the enzyme's chemical elimination (Pereira et al., 2022). Additionally, the detoxification function of UGT1A1 has been confirmed to be closely associated with the risk of digestive system tumors. In this study, it showed significant correlations with PD-L1 and HER2, suggesting its potential important role in immune regulation (Tomono et al., 2023; Sathe et al., 2024). Alcohol

FIGURE 11
Expression of prognostic genes in clinical samples. (a) ADH1B, (b) ADH4, (c) CYP19A1, (d) GPX3, (e) UGT1A1.

dehydrogenase 4 (Class II, ADH4) and alcohol dehydrogenase 1B (Class I, ADH1B), as key enzymes in ethanol metabolism, play a central role in liver ethanol detoxification by catalyzing the oxidation of alcohol to acetaldehyde. Notably, the ADH4-mediated conversion of ethanol to acetaldehyde is an essential step in liver detoxification (PMID: 16801720). However, excessive alcohol consumption significantly increases the risk of gastric cancer (GC), which is closely related to ADH4 (Freedman et al., 2007). The acetaldehyde produced by ADH4 catalysis exhibits high reactivity. On one hand, it can bind to DNA to form adducts, directly inducing gene mutations and compromising the stability of genetic material (Wang et al., 2023). On the other hand, acetaldehyde generated by ADH4 metabolism in gastric tissue can also disrupt the gastric mucosal barrier, creating pathological conditions for carcinogenesis (Nieminen and Salaspuro, 2018). Additionally, polymorphisms in the ADH1B gene significantly influence alcohol metabolism efficiency. Individuals carrying specific variants exhibit reduced acetaldehyde clearance capacity, leading to increased acetaldehyde accumulation during heavy drinking and thereby further elevating the risk of gastric cancer (GC) (Aida et al., 2013; Cui et al., 2009). However, in this study, ADH4 did not show significant downregulation in tumor tissues, suggesting that its expression may be subject to more complex regulatory mechanisms - a phenomenon that has not been adequately discussed in existing research. The aromatase gene cytochrome P450 family 19 subfamily A member 1(CYP19A1) can convert androgens into estrogens, and reproductive tissue growth and function depend on estrogens (Mo et al., 2013). The hormone-sensitive pathways of estrogen signaling

are connected to stomach cancer. The estrogen produced by CYP19A1 activates estrogen receptors in gastric tissue, altering cell growth and death. Our study further reveals its low expression in GC tissues, complementing previous literature reporting that its variants may disrupt hormone signaling pathways and thereby promote tumor growth (Frycz et al., 2017). At the level of genetic variation, genetic differences in the CYP19A1 gene can affect aromatase activity, which can affect estrogen levels and gastric epithelial cell hormonal control (Miyoshi et al., 2003). Glutathione peroxidase 3(GPX3), an antioxidant enzyme, reduces hydrogen peroxide and organic hydroperoxides to protect cells. GC tissues show GPX3 downregulation, which increases oxidative stress and tumor development (Chen et al., 2011). Hypermethylation of the GPX3 promoter in GC decreases the output of this protective enzyme, leading to loss of GPX3 activity and causing oxidative damage and malignancy in gastric epithelial cells. Also, an increase in reactive oxygen species (ROS) can damage DNA and lead to cancer, so GPX3 is needed to counteract ROS (Lan et al., 2017). Due to their roles in carcinogen detoxification (UGT1A1), alcohol metabolism and acetaldehyde production (ADH4, ADH1B), hormonal regulation (CYP19A1), and oxidative stress protection (GPX3), these five genes affect GC development. These gene abnormalities can increase GC risk by causing DNA damage, inflammation, metabolic abnormalities and aberrant cell proliferation. These key genes (UGT1A1, ADH4, ADH1B, CYP19A1, and GPX3) are crucial for the construction of our gastric cancer (GC) prognostic model. However, based on the

risk scores derived from the prognostic genes and combined with clinical data, the nomogram analysis revealed that the current model's AUC values ranged between 0.59 and 0.77, failing to exceed 0.8. Nevertheless, existing literature reports that the AUC values of GC-related models vary from 0.560 to 0.989 (Xu et al., 2024), indicating that our newly developed model possesses a certain ability to distinguish between favorable and poor prognoses in patients and holds some clinical guidance significance. However, its accuracy may still be relatively low. Further optimization by incorporating additional clinical information is necessary to enhance its predictive accuracy and generalizability, thereby facilitating more personalized treatment strategies.

Similar to the findings of previous studies, the findings of this research demonstrate the complex correlation between the prognostic gene profile linked to ADME activities and the immune microenvironment in GC (Avolio et al., 2025; Li et al., 2024). The results emphasize notable disparities in the prevalence of immunological and stromal cells among the two indicated risk groups, with specific focus on immune cells such as Effector memory CD8 T cells, Activated B cells, and Central memory T cells. The presence of these differences highlights the potential impact of the immune landscape on the prognosis of patients with GC and the processes that drive tumor growth. The presence of a positive association between the risk score and the quantity of cell types such mast cells, plasmacytoid dendritic cells, and effector memory CD4 T cells indicates that the high-risk group may have a milieu that suppresses the immune system to a greater extent (Lazăr et al., 2018). This could exacerbate the unfavorable outcome reported in these individuals, as such an environment may enhance tumor evasion from immune monitoring and promote cancer development and spread (Yang et al., 2021). The significant variations in the levels of expression of 24 immunological checkpoints, such as the increased expression of BTLA, CD200, and CD40LG in the high-risk group, provide additional evidence that immune evasion mechanisms are more active in patients with higher risk scores. Therefore, utilizing personalized patient risk scores to tailor more effective immunotherapies may represent a promising direction, and future studies could incorporate clinical data from immunotherapy treatments to validate the predictive capability of TIDE scores in real-world therapeutic applications.

Furthermore, the contrasting expression of m6A-RGs in the two risk groups indicates that RNA methylation may have a crucial impact on regulating the immune response and affecting the clinical outcomes of patients with GC. The excessive production of m6A-RGs in the high-risk group may likely cause the disruption of immune-related pathways, leading to the observed changes in immune cell infiltration and checkpoint expression (Zhang et al., 2016). The medication sensitivity assessments offer useful insights into prospective therapy methods for various risk groups. Our research suggested that patients classified as low-risk are more prone to positive responses to medications such as QS11 and ABT.888. These treatments specifically target hypoxia pathways and inflammatory signaling, respectively. On the other hand, the high-risk group was more sensitive to drugs such as DMOG and VX.702. The Spearman association study provides more evidence, indicating a positive correlation between the risk score and sensitivity to BIBW2992, which is an EGFR inhibitor. Conversely, there is a negative correlation between the risk score and the drugs

DMOG, Imatinib, AP.24534, and VX.702. The findings indicate that the prognostic gene signature not only has the ability to predict patient outcomes, but also has the potential to inform treatment decisions in clinical practice. Patients categorized as high-risk may get advantages from treatments that specifically target immunological checkpoints or EGFR signaling, and individuals in the high-risk group may also respond better to medicines that target hypoxia and inflammatory pathways. ESCC is characterized by mutations in the p53 gene, which enhance the production of the AGAP1 protein. This protein, in turn, increases the synthesis of exosomes, leading to the accelerated growth and spread of cancer cells (Feng et al., 2023). According to a study, QS11 has the ability to hinder the activity of AGAP1, which in turn reduces the growth and spread of ESCC cells. Furthermore, research has identified somatic (P53) mutations in GC as well (Han et al., 2024). The study found that low-risk patients showed considerable enrichment in pathways such as DNA replication and nucleotide excision repair. It is hypothesized that patients with a low risk profile may have a higher likelihood of seeing positive outcomes from gene- and protein-level therapy.

The integration of differential expression analysis, mRNA-miRNA-lncRNA interaction prediction, and scRNA-seq provides a comprehensive view of the molecular and cellular heterogeneity in GC (Li et al., 2025; Wan et al., 2015). Significant differences in miRNAs (e.g., mir590, mir7152, mir4420, mir5000) were identified between the high-risk and low-risk gastric cancer (GC) groups. These miRNAs exhibited significantly upregulated expression in the high-risk group, suggesting their potential role in suppressing oncogenic pathways to inhibit GC tumor growth, which aligns with existing literature. Similarly, the lncRNA (MIR100HG) was also markedly upregulated in the high-risk group. Previous studies have demonstrated that MIR100HG serves as a reliable prognostic biomarker associated with GC cell proliferation, migration, and invasion. The differentially expressed miRNAs and lncRNAs uncovered in this study may play pivotal roles in GC progression. They hold promise as prognostic markers and could provide novel therapeutic avenues for GC treatment.

ScRNA-seq data from GC tissues showed significant cell diversity using Seurat R package. ScRNA-seq data also showed significant differences in the expression levels of key prognostic genes such ADH1B, ADH4, CYP19A1, and GPX3 across cell types. The lower expression of these genes in tumor tissues suggests a role in GC onset and progression. For instance, decreased ADH1B and ADH4 expression in tumor cells can lead to acetaldehyde accumulation, which can promote cancer, DNA damage, and tumor growth. Decreased GPX3 may increase oxidative stress, which promotes gastric epithelial cell malignancy (Ren et al., 2021).

GC development's molecular mechanisms can be understood by studying miRNA-lncRNA interactions and single-cell data. The identified miRNAs, lncRNAs, and target genes may be GC biomarkers. Additionally, they could be targeted for new therapeutic methods. Customized treatment techniques that account for tumors' diverse cell and molecular profiles are crucial because GC tissues contain distinct types of cells. ADME-related prognostic genes vary in expression in different cell types, suggesting that they can affect the metabolic milieu of the tumor microenvironment (TME) and targeted therapeutic efficacy (Gu et al., 2015).

Regarding the limitations of current GC diagnostic methods (such as gastroscopy and tumor marker detection) in terms of sensitivity and specificity, this study identified five ADME-related prognostic genes (UGT1A1, ADH4, ADH1B, CYP19A1, and GPX3) that exhibit unique expression patterns in tumor tissues (UGT1A1, ADH1B, CYP19A1, and GPX3 show significant downregulation). These findings hold potential for developing new tools and strategies to improve GC diagnosis and treatment. First, a multiplex gene quantification assay based on minimally invasive samples (peripheral blood, gastric juice, exfoliated cells) could be developed for early screening of high-risk populations (e.g., individuals with family history or chronic *H. pylori* infection). This tool could compensate for the limitations of gastroscopy, reduce missed diagnoses, and improve early detection rates. Second, detecting ADME gene expression (e.g., downregulation of UGT1A1, ADH1B, GPX3) in gastroscopically suspicious lesions could enhance diagnostic accuracy through combined assessment, guide further examinations, and reduce false-negative results. Additionally, an ADME risk score could effectively stratify GC patients into high-risk and low-risk groups. Significant differences were observed between the two groups regarding immune microenvironment, drug sensitivity, and prognosis. The high-risk group (with poorer prognosis) requires aggressive treatment strategies, including neoadjuvant chemotherapy combined with targeted therapy (e.g., EGFR inhibitors), radical surgery, and adjuvant therapy, to reduce recurrence risk. In contrast, the low-risk group (with better prognosis) should avoid overtreatment, with surgical intervention or low-toxicity adjuvant chemotherapy being preferred options, along with intensified postoperative surveillance. These findings provide valuable guidance for developing precision medicine strategies.

Due to the relatively limited sample size in the training, validation and scRNA-seq datasets, particularly the scarcity of clinically normal tissue samples—our study may be susceptible to batch effects or population bias. This constraint likely hindered the model's ability to fully capture the comprehensive features and distribution patterns of the data, partially compromising the stability and reliability of the results (AUC >0.6 but <0.8) and limiting the generalizability of findings to broader populations. Similarly, the inconsistent results between RT-qPCR and single-cell data are likely attributable to insufficient sample size and sample heterogeneity. Furthermore, insufficient sample resources led to an extremely small clinical validation cohort (n = 5 for tumor/adjacent tissues), and the regulatory mechanisms of the ceRNA network relied solely on bioinformatic predictions without experimental validation, weakening the robustness of conclusions. Moving forward, we will further investigate the differential activity of these genes in specific cell populations. This will involve experimental validation of the ceRNA network's regulatory mechanisms through approaches such as RNA interference (RNAi), overexpression assays, constructing animal models, and gene knockout experiments. These studies aim to elucidate the precise role of the ceRNA network in gastric cancer (GC) initiation and progression, as well as its potential crosstalk with other regulatory pathways. Finally, we will attempt to obtain additional single-cell sample data for analysis and collect extra clinical samples to validate our research findings via qPCR. Meanwhile, we will specifically measure the relationship between

TIDE scores in the high-risk group and actual immune therapy responses to further verify the predictive value of TIDE scores. These efforts are expected to provide more reliable and actionable insights for improving GC treatment strategies.

This work concludes by offering a thorough investigation of the correlation between ADME-related prognostic genes and the immunological microenvironment in GC. The recognition of unique immunological profiles and medication sensitivities linked to various risk groups emphasizes the possibility of tailored treatment approaches in GC. Additional investigation is necessary to examine the fundamental mechanisms that cause these connections and to confirm these discoveries in medical environments, with the ultimate objective of enhancing patient outcomes through customized therapy methods. Our research helps us grasp GC's complex molecular and cellular networks. The found miRNAs, lncRNAs, and prognostic genes offer promising potential for further research and therapeutic applications, such as targeted medicines and personalized GC treatment.

5 Conclusion

This study effectively discovered predictive genes associated to ADME in GC by thorough bioinformatics analysis of data from the TCGA and GEO databases. The results provide useful knowledge on the genetic basis of GC and emphasize the potential of these genes as biomarkers for the detection, prediction of outcome, and targets for treatment. Through the integration of differential expression analysis, miRNA-lncRNA interaction prediction, and single-cell RNA sequencing data, we have discovered noteworthy connections between these prognostic genes, the tumor microenvironment, and cellular heterogeneity in GC.

The identified genes and their regulatory networks serve as a new reference framework for future study in GC, namely in comprehending the disease's course and in formulating tailored treatment methods. In order to further understand how these prognostic genes, contribute to the development of GC, it is necessary to conduct additional experimental research to uncover the exact mechanisms involved. Moreover, the practical implementation of these discoveries, which involves creating specific treatments, shows potential for enhancing the identification and management of patients with GC.

Data availability statement

The datasets analyzed for this study can be found in the TCGA database and GEO database [http://portal.gdc.cancer.gov/ and https://www.ncbi.nlm.nih.gov/geo/], accession number TCGA-GA, GSE62254 and GSE13911.

Ethics statement

The studies involving humans were approved by The First Hospital of Shanxi Medical University. The studies were conducted in accordance with the local legislation and

institutional requirements. The human samples used in this study were acquired from primarily isolated as part of your previous study for which ethical approval was obtained. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

Funding

Acknowledgments

Conflict of interest

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

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2025.1541401/full#supplementary-material

References

Aida, J., Yokoyama, A., Shimomura, N., Nakamura, K. I., Ishikawa, N., Terai, M., et al. (2013). Telomere shortening in the esophagus of japanese alcoholics: relationships with chromoendoscopic findings, aldh2 and adh1b genotypes and smoking history. *Plos One* 8 (5), e63860. doi:10.1371/journal.pone.0063860

Aitken, A. E, Richardson, T. A., and Morgan, E. T. (2008). Regulation of drug-metabolizing enzymes and transporters in infection, inflammation, and cancer. *Annu. Rev. Pharmacol. Toxicol.* 46 (1), 123–149. doi:10.1146/annurev.pharmtox.46.120604.141059

Aran, D., Looney, A. P., Liu, L., Wu, E., Fong, V., Hsu, A., et al. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* 20 (2), 163–172. doi:10.1038/s41590-018-0276-y

Avolio, P. P., Garbin, A., Pandini, M., Carriero, R., Kunderfranco, P., Elefante, M. G., et al. (2025). Immune profiling at the single-cell level to evaluate the immune microenvironment in bladder cancer: insights from non-muscle invasive and muscle-invasive disease. *J. Clin. Oncol.* 43 (5suppl), 842. doi:10.1200/jco.2025.43.5_suppl.842

Chen, B., Rao, X., House, M. G., Nephew, K. P., Cullen, K. J., and Guo, Z. (2011). GPx3 promoter hypermethylation is a frequent event in human cancer and is associated with tumorigenesis and chemotherapy response[J]. *Cancer Lett.* 5309 (1), 37–45. doi:10.1016/j.canlet.2011.05.013

Cristescu, R., Lee, J., Nebozhyn, M., Kim, K. M., Ting, J. C., Wong, S. S., et al. (2015). Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat. Med.* 21 (5), 449–456. doi:10.1038/nm.3850

Cui, R., Kamatani, Y., Takahashi, A., Usami, M., Hosono, N., Kawaguchi, T., et al. (2009). Functional variants in adh1b and aldh2 coupled with alcohol and smoking synergistically enhance esophageal cancer risk. *Gastroenterology* 137 (5), 1768–1775. doi:10.1053/j.gastro.2009.07.070

De Martel, C., Georges, D., Bray, F., Ferlay, J., and Clifford, G. M. (2020). Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *Lancet Glob. health* 8 (2), e180–e190. doi:10.1016/S2214-109X(19)30488-7

Feng, R., Yin, Y., Wei, Y., Li, Y., Li, L., Zhu, R., et al. (2023). Mutant p53 activates hnRNPA2B1-AGAP1-mediated exosome formation to promote esophageal squamous cell carcinoma progression. *Cancer Lett.* 562, 216154. doi:10.1016/j.canlet.2023.216154

Fletcher, J. I., Haber, M., Henderson, M. J., and Norris, M. D. (2010). Abc transporters in cancer: more than just drug efflux pumps. *Nat. Rev. Cancer* 10 (2), 147–156. doi:10.1038/nrc2789

Freedman, N. D., Abnet, C. C., Leitzmann, M. F., Mouw, T., Subar, A. F., Hollenbeck, A. R., et al. (2007). A prospective study of tobacco, alcohol, and the risk of esophageal and gastric cancer subtypes. *Am. J. Epidemiol.* 165 (12), 1424–33. doi:10.1093/aje/kwm051

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models *via* coordinate descent. *J. Stat. Softw.* 33 (1), 1–22. doi:10.18637/jss.v033.i01

Frycz, B. A., Murawa, D., Borejsza-Wysocki, M., Wichtowski, M., Spychała, A., Marciniak, R., et al. (2017). mRNA expression of steroidogenic enzymes, steroid hormone receptors and their coregulators in gastric cancer. *Oncol. Lett.* 13 (5), 3369–3378. doi:10.3892/ol.2017.5881

Furuta, T., Shirai, N., Sugimoto, M., Nakamura, A., Hishida, A., and Ishizaki, T. (2005). Influence of cyp2c19 pharmacogenetic polymorphism on proton pump inhibitor-based therapies. *Drug Metabolism & Pharmacokinetics* 20 (3), 153–167. doi:10.2133/dmpk.20.153

Geeleher, P., Cox, N., and Huang, R. S. (2014). PRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS one* 9 (9), e107468. doi:10.1371/journal.pone.0107468

Gu, W., Gao, T., Sun, Y., Zheng, X., Wang, J., Ma, J., et al. (2015). LncRNA expression profile reveals the potential role of lncRNAs in gastric carcinogenesis[J]. *Cancer Biomark.* 15 (3), 249–58. doi:10.3233/cbm-150460

Han, X., Jia, X., Sheng, C., Li, M., Han, J., Duan, F., et al. (2024). A comparison analysis of the somatic mutations in early-onset gastric cancer and traditional gastric cancer. *Clin. Res. hepatology gastroenterology* 48 (2), 102287. doi:10.1016/j.clinre.2024.102287

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinforma.* 14, 7. doi:10.1186/1471-2105-14-7

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., 3rd, Zheng, S., Butler, A., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* 184 (13), 3573–3587.e29. doi:10.1016/j.cell.2021.04.048

Lan, X., Xing, J., Gao, H., Li, S., Quan, L., Jiang, Y., et al. (2017). Decreased expression of selenoproteins as a poor prognosticator of gastric cancer in Humans[J]. *Biol. Trace Elem. Res.* 178 (1), 22–28. doi:10.1007/s12011-016-0908-8

Lazăr, D. C., Avram, M. F., Romoşan, I., Cornianu, M., Tăban, S., and Goldiş, A. (2018). Prognostic significance of tumor immune microenvironment and immunotherapy: novel insights and future perspectives in gastric cancer. *World J. Gastroenterol.* 24 (32), 3583–616. doi:10.3748/wjg.v24.i32.3583

Li, X., Tang, B., Yujie, O., Xu, C., and Yuan, S. (2025). Single-cell rna sequencing analysis reveals cancer-associated fibroblast signature for prediction of clinical outcomes and immunotherapy in gastric cancer. *J. Immunothe.* 48 (2), 15. doi:10.1097/CJI.0000000000000539

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8

Li, Q., Zhang, C., Ren, Y., Qiao, L., Xu, S., Li, K., et al. (2024). A novel platelets-related gene signature for predicting prognosis, immune features and drug sensitivity in gastric cancer[J]. *Front. Immunol.* doi:10.3389/fimmu.2024.1477427

Liu, T. T., Li, R., Huo, C., Li, J. P., Yao, J., Ji, X. L., et al. (2021). Identification of CDK2-related immune forecast model and cerna in lung adenocarcinoma, a pan-cancer analysis. *Front. cell Dev. Biol.* 9, 682002. doi:10.3389/fcell.2021.682002

Luo, Q., Lai, R., Su, M., Wu, Z., Feng, H., and Zhou, H. (2024). Risk factors and a predictive model for the occurrence of adverse outcomes in patients with new-onset refractory status epilepsy. *Front. Mol. Neurosci.* 17, 1360949. doi:10.3389/fnmol.2024.1360949

Ma, X. T., Liu, X., Ou, K., and Yang, L. (2024). Construction of an immune-related gene signature for overall survival prediction and immune infiltration in gastric cancer. *World J. Gastrointest. Oncol.* 16 (3), 919–932. doi:10.4251/wjgo.v16.i3.919

Maeser, D., Gruener, R. F., and Huang, R. S. (2021). OncoPredict: an R package for predicting *in vivo* or cancer patient drug response and biomarkers from cell line screening data. *Brief. bioinform* 22 (6), bbab260. doi:10.1093/bib/bbab260

Miyoshi, Y., Ando, A., Hasegawa, S., Ishitobi, M., Yamamura, J., Irahara, N., et al. (2003). Association of genetic polymorphisms in cyp19 and cyp1a1 with the oestrogen receptor-positive breast cancer risk. *Eur. J. Cancer* 39 (17), 2531–2537. doi:10.1016/j.ejca.2003.08.017

Mo, S. J., Ke, S. R., Zhang, J. T., and Yang, S. L. (2013). Progress in research of polymorphisms of genes associated with susceptibility to esophageal cancer in chinese han population. *World Chinese J. Digestol.* 21 (21), 2045–2050. doi:10.11569/wcjd.v21.i21.2045

Nieminen, M. T., and Salaspuro, M. (2018). Local Acetaldehyde-An essential role in alcohol-related upper gastrointestinal tract carcinogenesis. *Cancers (Basel)* 10 (1), 11. doi:10.3390/cancers10010011

Oh, S. C., Sohn, B. H., Cheong, J. H., Kim, S. B., Lee, J. E., Park, K. C., et al. (2018). Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype. *Nat. Commun.* 9 (1), 1777. doi:10.1038/s41467-018-04179-8

Peng, B., Lin, Y., Yi, G., Lin, M., Xiao, Y., Qiu, Y., et al. (2024). Comprehensive landscape of m6A regulator-related gene patterns and tumor microenvironment infiltration characterization in gastric cancer. *Sci. Rep.* 14 (1), 16404. doi:10.1038/s41598-024-66744-0

Pereira, E. E. B., Leitão, L. P. C., Andrade, R. B., Modesto, A. A. C., Fernandes, B. M., Burbano, R. M. R., et al. (2022). UGT1A1 gene polymorphism contributes as a risk factor for lung cancer: a pilot study with patients from the amazon. *Genes (Basel)* 13 (3), 493. doi:10.3390/genes13030493

Rao, X., Huang, X., Zhou, Z., and Lin, X. (2013). An improvement of the 2(-delta delta CT) method for quantitative real-time polymerase chain reaction data analysis. *Biostat. Bioinforma. Biomath.* 3 (3), 71–85.

Rawla, P., and Barsouk, A. (2019). Epidemiology of gastric cancer: global trends, risk factors, and prevention. *Przegl Gastroenterol.* 14 (1), 26–38. doi:10.5114/pg.2018.80001

Ren, Y., Zhang, B., Xu, X., and Zhang, L. (2021). Dissecting the single-cell transcriptome network in superficial/deep tumor tissues of diffuse gastric cancer. *Research Square [Preprint].* doi:10.21203/rs.3.rs-1130074/v1

Rodrigues, A. C. (2022). "ADME genes: definition, function, and clinical significance," in *Pharmacogenomics in drug discovery and development: from bench to bedside.* Springer, 35–60.

Sathe, A. G., Singh, I., Singh, P., Diderichsen, P. M., Wang, X., Chang, P., et al. (2024). Population pharmacokinetics of Sacituzumab Govitecan in patients with metastatic triple-negative breast cancer and other solid tumors. *Clin. Pharmacokinet.* 63 (5), 669–681. doi:10.1007/s40262-024-01366-3

Smyth, E. C., Nilsson, M., Grabsch, H. I., van Grieken, N. C., and Lordick, F. (2020). Gastric cancer. *Lancet* 396 (10251), 635–648. doi:10.1016/S0140-6736(20)31288-5

Tang, X., Li, R., Wu, D., Wang, Y., Zhao, F., Lv, R., et al. (2022). Development and validation of an adme-related gene signature for survival, treatment outcome and immune cell infiltration in head and neck squamous cell carcinoma. *Front. Immunol.* 13, 905635. doi:10.3389/fimmu.2022.905635

Tomono, H., Taniguchi, H., Fukuda, M., Ikeda, T., Nagashima, S., Akagi, K., et al. (2023). Phase II study of IRInotecan treatment after COmbined chemo-immunotherapy for extensive-stage small cell lung cancer: protocol of IRICO study. *Thorac. Cancer* 14 (28), 2890–2894. doi:10.1111/1759-7714.15097

Wan, X., Ding, X., Chen, S., Song, H., Jiang, H., and Fang, Y. (2015). The functional sites of mirnas and lncrnas in gastric carcinogenesis. *Tumor Biol.* 36 (2), 521–532. doi:10.1007/s13277-015-3136-5

Wang, X., Zhang, Y., Wang, Y., Liu, J., Xu, X., Liu, J., et al. (2023). Associations between alcohol metabolism-related gene polymorphisms and the risk of digestive system cancers: a meta-analysis. *BMC cancer* 23, 568. doi:10.1186/s12872-023-03472-9

Xu, L., Lyu, J., Zheng, X., and Wang, A. (2024). Risk prediction models for gastric cancer: a scoping review. *J. Multidiscip. Healthc.* 17, 4337–4352. doi:10.2147/JMDH.S479699

Yang, Y., He, W., Wang, Z. R., Wang, Y. J., Li, L. L., Lu, J. Z., et al. (2021). Immune cell landscape in gastric cancer. *Biomed Res. Int.* 2021, 1930706. doi:10.1155/2021/1930706

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics-a J. Integr. Biol.* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Zhang, C., Samanta, D., Lu, H., Bullen, J. W., and Semenza, G. L. (2016). Hypoxia induces the breast cancer stem cell phenotype by hif-dependent and alkbh5-mediated m6a-demethylation of nanog mrna. *Proc. Natl. Acad. Sci. USA* 113 (14), E2047. doi:10.1073/pnas.1602883113