

Construction and Validation of a Prognostic Prediction Model for Gastric Cancer Using a Series of Genes Related to Lactate Metabolism

Siyu Wang

Sun Yat-sen University Cancer Center

Yuxin Wang

The First Clinical College of Jilin University

Xin Wang

The First Affiliated Hospital of Sun Yat-sen University

Shuqiang Yuan (✉ yuanshq@sysucc.org.cn)

Sun Yat-Sen University Cancer Center <https://orcid.org/0000-0003-3661-5069>

Research Article

Keywords: Gastric cancer, Lactate metabolism, Prognosis, Prediction model

Posted Date: April 28th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-449860/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Gastric cancer is one of the most common clinical malignant tumors worldwide, with high morbidity and mortality. The commonly used TNM staging and some common biomarkers have a certain value in predicting the prognosis of GC patients, but they gradually failed to meet the clinical demands. Therefore, we aim to construct a prognostic prediction model for GC patients.

Methods: A total of 350 cases were included as TCGA-STAD entire cohort, including TCGA-STAD training cohort (n=176) and TCGA-STAD testing cohort (n=174). GSE15459 (n=191), GSE62254 (n=300) and some cases in our center (n=12) were for external validation.

Results: Through differential expression analysis and univariate Cox regression analysis in TCGA-STAD training cohort, we screened out 5 genes among 600 genes related to lactate metabolism for the construction of our prognostic prediction model. The internal and external validations showed the same result, that is, patients with higher risk score were associated with poor prognosis (all $P < 0.05$), and our model works well without regard of patients' age, gender, tumor grade, clinical stage or TNM stage, which supports the availability, validity and stability of our model. Gene function analysis, tumor-infiltrating immune cells analysis, tumor microenvironment analysis and clinical treatment exploration were performed to improve the practicability of the model, and hope to provide a new basis for more in-depth study of the molecular mechanism for GC and for clinicians to formulate more reasonable and individualized treatment plans.

Conclusions: We screened out and used 5 genes related to lactate metabolism to develop a prognostic prediction model for GC patients based on them. The prediction performance of the model is confirmed by a series of bioinformatics and statistical analysis.

1 Introduction

As a serious global health problem, the global morbidity of gastric cancer (GC) ranks fifth and it's the fourth leading cause of cancer-related death^[1], with more than one million people newly diagnosed with it worldwide^[2]. The majority of cases of gastric cancer are usually not diagnosed until an advanced stage, usually leading to a poor prognosis^[3, 4]. Although overall survival (OS) for GC patients have improved greatly due to combined treatment^[4, 5], it is still difficult to evaluate the prognosis of GC patients due to the great differences in individual prognosis^[6].

At present, tumor-node-metastasis (TNM) staging system has been generally applied to predicting the prognosis of GC and more and more prognostic prediction models have also been constructed based on the general data and clinicopathological data of GC patients^[7]. However, there is a certain lag in time of these data, resulting in its less accuracy.

In addition, a series of traditional biomarkers (such as CEA, CA 12 – 5, CA 19 – 9 and CA 72 – 4) and novel biomarkers (such as microRNAs, long non-coding RNAs and circular RNAs) have been exploited to assist

in the diagnosis and prognostic prediction of GC and achieve good clinical effects^[8]. Unfortunately, most of novel biomarkers show shortcomings due to limitation by high-throughput technology and some of them are still at the organ level^[9]. Some researchers have found several sorts of genes related to GC and explored their prognostic value, but the prognostic value of an individual gene is still not ideal and perfect^[10, 11]. With the rapid development of precision medicine and gene sequencing technology, multiple genes signature will be more and more applied to the construction of prognostic prediction models to achieve more prospective, accurate and individualized prediction performance.

Reprogramming of energy metabolism and evading immune are two emerging hallmarks of cancers^[12], and the former is a common feature of most tumor cells^[13]. Tumor cells were found to upmake a large amount of glucose and metabolize into lactate under oxygen-rich conditions, known as "Warburg effect"^[14]. The same phenomenon was also found in GC cells^[15]. Lactate, as a signaling molecule, is discovered to play important roles in the regulation of the metabolic pathways and the immune response within the tumor microenvironment^[16], which is closely related to the occurrence, invasion, metastasis, drug resistance and poor prognosis of GC^[15]. Sonveaux found that lactate promotes endothelial cell division, thus promoting angiogenesis^[17]. Some researchers have successfully constructed corresponding metabolism-related gene signatures for prognostic prediction of gliomas^[18, 19], GC^[20] and lung adenocarcinomas (LUAD)^[21]. All the above indicates the availability and feasibility of applying lactate metabolic genes to predicting the prognosis of GC, which is the theoretical basis for us to construct a prognostic prediction model for GC patients.

As lactate metabolism is a complex biological process involving multiple genes, the predictive value of those models based on multiple metabolism related genes will be better than those based on a single gene. Therefore, we aimed and managed to construct a prognostic prediction model by screening a series of genes related to lactate metabolism to better predict the prognosis of GC patients.

2 Methods

2.1 Acquisition of genes related to lactate metabolism

Genes related to lactate metabolism, known as LM genes, were collected in GeneCards database (<https://www.genecards.org/>), which provides a wide range of information about human genes. The term "lactate metabolism" was used as a keyword for the search, and genes with relevance scores > 12 were considered to be LM genes.

2.2 Collection of datasets

The RNA-seq data and clinical characteristics of TCGA cohorts (ACC, BLCA, BRCA, CESC, CHOL, COAD, DLBC, ESCA, GBM, HNSC, KICH, KIRC, KIRP, LAML, LGG, LIHC, LUAD, LUSC, MESO, OV, PAAD, PCPG, PRAD, READ, SARC, SKCM, STAD, TGCT, THCA, THYM, UCEC, UCS and UVM) were downloaded from Xena website (<https://xenabrowser.net/datapages/>). At the same time, we also downloaded the TCGA Pan-

Cancer Datasets of Immune Subtype, Immune Signature Scores (Denise Wolf et al), Stemness Score (DNA methylation based) and Stemness Score (RNA based) from Xena website. The gene-expression profile matrixes of the STAD cohort GEO: GSE15459 and GSE62254 were collected from GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) for external validation. The "limma" package was used to do the normalization for the expression profiles. The average expression level was retained for repeated genes. Moreover, Sequencing data was obtained from surgical tissue specimens of patients in our center. The specimens included gastric cancer tissue and gastric cancer peritoneal metastasis tissue, and the tissue specimens of these patients were diagnosed as GC and GC peritoneal metastasis by pathologist in our center.

2.3 DEG analysis and gene function analysis

By using the "limma" package, those LM genes mentioned above were investigated with differential expression analysis in TCGA-STAD cohort and visualized as heatmap plots. $|\text{Fold change (FC)}| > 0$ and p value (adj. p) < 0.05 were considered to be statistically apparent for distinguishing differential expression genes (DEGs). We selected the intersection of DEGs and LM genes, known as LM-DEGs, for further analysis.

The analysis of gene function is an important procedure in translating molecular findings from high-throughput data into biological significance. By using the "clusterProfiler" package in R, we perform statistical analysis and visualize functional profiles of LM genes, including Genetic Ontology (GO) annotation analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment pathway analysis. Adj. P value < 0.05 was considered to be statistically significant.

2.4 The construction of prognostic gene signature

By performing the univariate Cox proportional hazards regression analysis on each LM-DEG, we selected genes evidently associated with OS in TCGA-STAD training cohort. Then we constructed a multivariable model with LM genes. We calculate the risk score by using those genes with nonzero coefficients. A prognostic risk score was calculated for each patient by using the following formula:

Risk score = expression level of gene 1 * j_1 + expression level of gene 2 * j_2 + ... + expression level of gene x * j_x , where j represents the coefficient.

We used the "survminer" package to select the optimal cut-off value in risk score to divide TCGA-STAD patients into high and low risk groups. The identical formula and identical cut-off value were selected to two GEO datasets and our own sequencing data for validation.

Performing univariate and multivariate Cox proportional hazards regression analyses to investigate whether the LM-DEGs-based prognostic model was an independent prognostic factor. To assess the survival differences between two groups, we constructed the KM survival curve and used the log-rank test. At the same time, to examine the specificity and sensitivity of the prognostic performance, we explored the ROC curve analysis. The AUC values indicated discrimination.

2.5 GSEA

To find the possible pathways and molecular mechanisms in the high-risk group and low-risk group in TCGA-STAD cohorts, we use the analysis of GSEA (<https://www.gsea-msigdb.org/gsea/index.jsp>). Gene sets were considered significantly enriched with a p value of < 0.05 after 1000 permutation.

2.6 Investigation of tumor-infiltrating immune cells

To explore the relationship between the immune-cell characteristics and the model and genes, we considered the CIBERSORT method to calculate the immune infiltration statuses in the samples from the TCGA-STAD dataset. By using the Wilcoxon signed-rank test, we explored the differences in immune infiltrating cell content between high and low risk groups of the constructed model.

2.7 Tumor microenvironment analysis

To explore the immune microenvironment of GC patients, we applied the ESTIMATE algorithm in “estimate” and “limma” package of R Software to calculate the Immune Score and Stromal Score. The Estimate Score was used to describe tumor purity. This analysis was based on the gene expression data retrieved from TCGA-STAD datasets.

2.8 Exploration of the significance of the model in clinical treatment

To explore the model for gastric cancer treatment in the clinic, we calculated the IC_{50} of common approved chemotherapeutic drugs in the TCGA-STAD datasets by using “pRRophetic” package of R Software. The difference between the high and low risk groups in the IC_{50} was compared by using Wilcoxon signed-rank test.

2.9 NCI-60 analysis

We downloaded the drug-related data from the NCI-60 database, which contains data on 60 different cancer cell lines in nine different types of cancer. F5, MTTP, SERPINE1, CYP19A1 and SLC52A3 mRNA expression levels and z scores for cell sensitivity data (GI_{50}) were collected from 59 cell lines and were used to explore the relationship between gene expression and drug sensitivity by using Pearson correlation. In the correlation analysis, we analyzed the drug response of 262 drugs approved by FDA or on clinical trials.

3 Result

3.1 Calculation of the risk score and construction of the prognostic prediction model

A total of 350 cases of TCGA-STAD were downloaded from Xena as the entire cohort, which were further randomly divided into TCGA-STAD training cohort ($n = 176$) and TCGA-STAD testing cohort ($n = 174$). The

former was for the construction of our prognostic prediction model and the latter was for internal validation (Fig. 1).

After search and screening on GeneCards database, we collected 600 LM genes. Among them, 376 differential expression genes (known as LM-DEGs) were screened out after differential expression analysis between cancer and normal tissues in TCGA-STAD training cohort. The gene-expression profiles of the top 60 of them are displayed in Fig. 2A. After univariate Cox regression analysis, finally, we obtained 5 genes (F5, MTTP, SERPINE1, CYP19A1 and SLC52A3, known as model genes) with nonzero coefficients. The expressions of the 5 genes were all significantly different between GC and normal tissues in TCGA-STAD training cohort (all $P < 0.01$, Fig. 2B-F), verifying the rationality and feasibility of using these 5 genes to develop our prediction model:

Risk score = $F5 * 0.1496 + MTTP * 0.1838 + SERPINE1 * 0.1581 + CYP19A1 * 0.4849 + SLC52A3 * (-0.3032)$.

The risk scores of each patient were calculated separately using the above formula, and all these patients were divided into high and low risk group according to the optimal cut-off value we calculated.

3.2 Evaluation and internal validation of the prognostic prediction model

In training cohort, Kaplan-Meier (KM) survival analysis revealed that patients of high risk group obviously had worse prognosis compared to those of low risk group, with significant difference ($P < 0.001$, Fig. 3A). And the corresponding receiver operating characteristic (ROC) curve of 5-year survival was shown in Fig. 3D, with the area under curve (AUC) of 0.763. Univariate and multivariate Cox regression analyses in TCGA-STAD training cohort showed that the risk score was confirmed to be an independent prognostic factor for patients' survival (both $P < 0.05$, Fig. 4A, 4D).

For internal validation, the patients in TCGA-STAD testing cohort were also divided into high and low risk group according to the same cut-off value, showing significant difference in KM survival analysis ($P < 0.001$, Fig. 3B) and with 5-year AUC of 0.781 (Fig. 3E). In entire cohort, our prognostic prediction model also had a good performance (Fig. 3C, 3F).

3.3 Performance of the prognostic prediction model on different subgroups

All the patients in TCGA-STAD entire cohort were further divided into two subgroups according to their age (> 65 or ≤ 65 years old), gender (male or female), tumor grade (G1-2 or G3), clinical stage (stage I-II or stage III-IV), T stage (T1-2 or T3-4), N stage (N0 or N1-3) and M stage (M0 or M1). All the corresponding KM curves were shown in Fig. 5. It should be specially explained that between high and low risk patients with M1, the result of log-rank test did not reach the level of statistical significance ($P = 0.056$), but with

obviously separated survival curve. We think it might be caused by lack of numbers of cases. While in other subgroups, there were all significant differences between high and low risk groups (all $P < 0.05$), revealing that our prognostic prediction model works well without regard of patients' age, gender, tumor grade, clinical stage or TNM stage, that is, the patients in high risk group were usually associated with a poor prognosis.

3.4 External validation of the prognostic prediction model

According to the same cut-off value in TCGA-STAD training cohort, we did KM survival analysis on three independent datasets: GSE15459 ($n = 191$), GSE62254 ($n = 300$) and some cases in our center ($n = 12$). The results showed that the patients in high risk group were significantly associated with lower OS (all $P < 0.05$, Fig. 6), which was consistent with findings in TCGA-STAD training cohort. As for the 12 cases in our center, similar results were obtained regardless of whether the tissues were taken from GC or peritoneal metastasis.

To evaluate the sensitivity and specificity of the prognostic prediction model, external validation in the independent dataset of 32 types of TCGA tumors were performed. According to the results of KM survival analysis, it was found that our signature could significantly distinguish high and low risk patients with one of the following 20 types of TCGA tumors respectively: adrenocortical carcinoma (ACC), bladder urothelial carcinoma (BLCA), colon adenocarcinoma (COAD), glioblastoma multiforme (GBM), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), acute myeloid leukemia (LAML), lower grade glioma (LGG), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), mesothelioma (MESO), pancreatic adenocarcinoma (PAAD), pheochromocytoma and paraganglioma (PCPG), sarcoma (SARC), skin cutaneous melanoma (SKCM), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), thymoma (THYM), uterine corpus endometrial carcinoma (UCEC) and uveal melanoma (UVM) (all $P < 0.05$; Fig. 7A-T).

3.5 Gene function analysis and gene set enrichment analysis (GSEA)

To further explore the potential biological functions of LM-DEGs, we performed GO annotation analysis and KEGG enrichment pathway analysis to visualize their functional profiles. The GO analysis showed that the top 5 enriched terms in BP group were "energy derivation by oxidation of organic compounds", "small molecule catabolic process", "ATP metabolic process", "cellular respiration" and "electron transport chain"; the top 3 enriched terms in CC group were "mitochondrial matrix", "mitochondrial inner membrane" and "mitochondrial protein complex"; the top 2 enriched terms in MF group were "electron transfer activity" and "tetrapyrrole binding" (Fig. 8A, 8B). Meanwhile, KEGG analysis showed that LM-DEGs were mainly enriched in "thermogenesis", "carbon metabolism" and "oxidative phosphorylation" (Fig. 8D, 8E). The above 13 enriched terms are all associated with cellular energy metabolism.

Furthermore, the transcript message of patients stratified by risk score into high and low risk groups were analyzed by GSEA (Fig. 8C, 8F). "Collagen containing extracellular matrix", "extracellular matrix structural constituent", "calcium signaling pathway", "focal adhesion" and "neuroactive ligand receptor interaction" were more enriched in high risk group, while "mitochondrial membrane organization", "mitochondrial protein complex", "mitochondrial respiratory chain complex assembly" and "mitochondrial transport" were more enriched in low risk group, suggesting that mitochondrial function of GC cells in high risk group was inhibited to a certain extent.

3.6 Model genes analysis

To evaluate the impact of the 5 model genes on patients' OS individually, the optimal cut-off values of the expression levels of the 5 model genes were calculated according to their expression levels separately, and patients in TCGA-STAD entire cohort were divided into high and low expression groups to explore the impact of the 5 model genes on patients' survival. The results of KM survival analysis showed that all these 5 model genes exerted significant effects on patients' survival individually (all $P < 0.05$, Fig. 9). High expressions of F5, MTTP, SERPINE1 and CYP19A1 may be risk factors for GC, while high expression of SLC52A3 may be a protective factor, which were consistent with our risk score formula.

3.7 Investigation of tumor-infiltrating immune cells

We also analyzed immune infiltration levels of several specific types of immune cells. Macrophages M1, NK cells activated, T cells CD8 and T cells follicular helper were more enriched in low risk group, while Mast cells activated and NK cells resting were more enriched in high risk group (all $P < 0.05$, Fig. 10A-F). The KM survival analysis showed that high level of Mast cells activated was associated with a poor prognosis, and the high levels of NK cells activated, T cells CD8 and T cells follicular helper were associated with a better prognosis (both $P < 0.05$, Fig. 10G-J).

3.8 Tumor microenvironment analysis

To further explore the association of the 5 model genes' expressions with immune microenvironment, we further compared their expressions among 5 immune subtypes in TCGA-STAD entire cohort. As shown in Fig. 11A, high expressions of F5, SERPINE1 and CYP19A1 were associated with Immune C6 (TGF- β dominant), and high expression of MTTP was associated with Immune C3 (Inflammatory). However, high expression of SLC52A3 was usually associated with Immune C2 (IFN- γ dominant).

The Estimate Score and Stromal Score calculated by "estimate" package showed the patients in high risk group were associated with significantly higher Estimate Score and Stromal Score, indicating their poor prognosis (both $P < 0.001$, Fig. 11B). The results of the correlation between the expressions of the 5 model genes and RNAss, DNAss, Stromal Score, Immune Score or Estimate Score in TCGA-STAD cohort showed that higher expression of SLC52A3 was associated with lower Estimate Score ($P < 0.001$), indicating SLC52A3 correlated with reduced risk of GC (Fig. 11C).

We also analyzed the correlation between the expressions of the 5 model genes and immune checkpoint molecules among TCGA pan-cancer, as shown in the scatter plots of Fig. 11D. The results showed that the higher expression of SERPINE1 was associated with higher levels of immune checkpoint molecules (PD-1, PD-L1 and CTLA4), indicating that SERPINE1 might promote tumor development by down-regulating immune response ($P < 0.01$).

3.9 Exploration of the significance of the model in clinical treatment

We analyzed the estimated IC_{50} of the 9 chemotherapeutic drugs (Camptothecin, Cisplatin, Mitomycin C, Vinblastine, Docetaxel, Doxorubicin, Erlotinib, Paclitaxel and Sorafenib), the results showed that the patients in high risk group had significantly higher estimated IC_{50} levels of the 4 chemotherapeutic drugs (Camptothecin, Cisplatin, Mitomycin C and Vinblastine, all $P < 0.05$, Fig. 12A-D), revealing their lower drug sensitivity for high risk patients. While there were no significant declines of drug sensitivity among Docetaxel, Doxorubicin, Erlotinib, Paclitaxel and Sorafenib (Fig. S12).

Scatter plots were also drawn to study the correlation between the expressions of the 5 model genes and drug sensitivity. We found that the expression of the 5 model genes were related to increased or decreased drug sensitivity of a number of chemotherapeutic drugs, see Fig. 12E for more details.

4 Discussion

Gastric cancer (GC) is one of the most common clinical malignant tumors worldwide, with high morbidity and mortality. The commonly used TNM staging and some common biomarkers, such as CEA, CA 12 - 5, CA 19 - 9 and CA 72 - 4, have a certain value in predicting the prognosis of GC patients, but they gradually failed to meet the clinical demands at present because of their lag in time and there're great differences in individual prognosis. Therefore, the construction of a prospective, accurate and individualized prognostic prediction model for GC is one of the most urgent clinical tasks at the present stage.

After differential expression analysis and univariate/multivariate Cox regression analysis, we screen out 5 genes as the model genes (F5, MTPP, SERPINE1, CYP19A1 and SLC52A3). In previous studies, Liu, Y have found that F5 gene is significantly upregulated in GC tissues, and may be a potential prognostic biomarker for GC^[22]. SERPINE1, as a tumor-promoting gene of gastric cancer^[23], its high expression is significantly related to the poor prognosis of GC patients^[24], which can be used as a biomarker for the diagnosis and prognosis of GC alone^[25]. CYP19A1 was found to be correlated with the poor prognosis of lung cancer^[26] and colon/rectal cancer^[27], unfortunately there is no research on GC.

After calculating the risk score and successfully constructing the prognostic prediction model, to ensure its effectiveness and stability, the study first used TCGA-STAD testing cohort for internal verification, and used three independent datasets for external verification. These results all suggest that patients in high

risk group were negatively correlated with their OS. We also evaluated the impact of the 5 model genes on patients' OS individually, all the results were consistent with our risk score formula.

To study whether the differences in clinical information of patients have an impact on the predictive performance of the model, we further divided TCGA-STAD entire cohort into several subgroups according to their age, gender, tumor grade, clinical stage and TNM stage. The KM analysis results were all in line with the expected results, showing that our model is not easily disturbed by the above clinical information of patients, suggesting its wide range of clinical predictive value.

For further analysis of the specificity and sensitivity of the model, 32 types of TCGA tumors were selected for external verification of the model. The results of the KM analysis showed that a higher risk score was associated with a lower OS among 22 TCGA tumors. Our explanation for this result is that these tumors may have similar metabolic characteristics and share some same nodes in their metabolic pathways. The identification of these nodes may contribute to a more in-depth study of the mechanism of these tumors and update the existing prognostic prediction models.

In addition, GO analysis and KEGG analysis were performed to further explore the potential biological functions of LM-DEGs, and it was found that all these enriched terms are associated with cellular energy metabolism. For deeper study of the biological functions, we performed GSEA to explore the different enrichment pathways between high and low risk groups in TCGA-STAD entire cohort. Interestingly, it was found that all the "mitochondrial-related" pathways were more enriched in low risk group, suggesting that mitochondrial function of GC cells in high risk group was inhibited to a certain extent, which is consistent with the fact that even under the condition of sufficient oxygen, GC cells still absorb a large amount of glucose and metabolize it into lactate^[28].

Importantly, tumor-infiltrating immune cells seems to have more and more significant prognostic value of tumors^[29, 30], including GC^[31]. In our study, Mast cells activated were more enriched in high risk group and led to poor prognosis. It is found in previous studies that Mast cells enriched in tumor tissue can produce a large amount of pro-angiogenic factors, which promote angiogenesis and lead to poor prognosis^[32, 33]. Sammarco found the same phenomenon in GC^[34]. Macrophages M1, NK cells activated, T cells CD8 and T cells follicular helper were found to be more abundant in patients of low risk group, and the latter three significantly affected patients' OS. Husain found tumor-derived lactate inhibits NK cell function^[35]. Fischer found tumor cells can increase surrounding lactate concentration and further inhibit the function of cytotoxic T cells^[36]. Interestingly, M1 and M2 macrophages were found to have almost opposite effects on tumor patients' prognosis, and the former has been proved to improve patients' OS^[37].

As for tumor microenvironment analysis, Baumann found lactate induces TGF- β expression^[38], and some other researchers' results showed TGF- β promotes GC malignancy^[39] and metastasis^[40], which can be suppressed by attenuating TGF- β ^[41]. It is worth noting in our study that SERPINE1 and CYP19A1, may act as oncogenic genes, were found to associated with Immune C6 (TGF- β dominant), which is consistent with the above researches. SLC52A3 was found to be associated with Immune C2 (IFN- γ dominant) in our

study. Intriguingly, IFN- γ was once recognized as a tumor-killing cytokine, but there's controversial in recent years^[42] and its mechanism is not clear. Based on the ESTIMATE algorithm, patients in high risk group were associated with significantly higher Estimate Score and Stromal Score, indicating a correlation with poor prognosis. And higher expression of SLC52A3 was associated with lower Estimate Score, indicating better correlation. Part of the above results, as well as the results of immune checkpoint molecules, are not in line with expectations, which we think may be related to the complexity and polyvalence of metabolic pathways. On the whole, however, these results accord with our expectations.

According to our drug sensitivity analysis, Camptothecin, Cisplatin, Mitomycin C and Vinblastine showed lower drug sensitivity among GC patients in high risk group compared with those of low risk, while there're no significant reduced drug sensitivity among Docetaxel, Doxorubicin, Erlotinib, Paclitaxel and Sorafenib. Meanwhile, the expression of the 5 model genes were found to be related to increased or decreased drug sensitivity of a number of chemotherapeutic drugs. These results can provide a new and fresh basis for clinicians to formulate individualized chemotherapy regimens for GC patients in high risk group based on our prognostic prediction model.

However, some limitations exist in our study. Most of our data was downloaded from the public database, with the clinical information of some patients incomplete, and there's only a small number of cases from our center. At the same time, as a retrospective study, information bias is inevitable, so it is necessary to prove the clinical predictive value of the model through more convincing prospective studies in future clinical practice.

5 Conclusion

In this study, we excavated and screened a series of LM-DEGs and developed a prognostic prediction model for GC patients based on these genes. Through a series of bioinformatics and statistical analysis, our model has been validated to have a good performance to predict the prognosis of GC patients, and can guide clinicians to provide more effective and individualized therapeutic schemes for GC patients.

Declarations

Ethics approval and consent to participate: Not applicable

Consent for publication: Yes

Availability of data and materials: Yes

Competing interests: The authors declare no competing interests.

Acknowledgements: Not applicable

Funding: This research was not solicited and did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' contributions

Siyu Wang: Investigation, Methodology, Visualization, Software, Roles/Writing - original draft. **Yuxin Wang:** Data curation, Roles/Writing - original draft, Writing - review & editing. **Xin Wang:** Data curation. **Shuqiang Yuan:** Conceptualization, Resources, Supervision.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. *CA Cancer J Clin*, 2021.
2. Thrift AP, El-Serag HB. Burden of Gastric Cancer[J]. *Clin Gastroenterol Hepatol*. 2020;18(3):534–42.
3. Katai H, Ishikawa T, Akazawa K, et al. Five-year survival analysis of surgically resected gastric cancer cases in Japan: a retrospective analysis of more than 100,000 patients from the nationwide registry of the Japanese Gastric Cancer Association (2001–2007)[J]. *Gastric Cancer*. 2018;21(1):144–54.
4. Tan Z. Recent Advances in the Surgical Treatment of Advanced Gastric Cancer: A Review[J]. *Med Sci Monit*. 2019;25:3537–41.
5. Kiyokawa T, Fukagawa T. Recent trends from the results of clinical trials on gastric cancer surgery[J]. *Cancer Commun (Lond)*. 2019;39(1):11.
6. Costa NR, Gil Da Costa RM, Medeiros R. A viral map of gastrointestinal cancers[J]. *Life Sci*. 2018;199:188–200.
7. Muneoka Y, Akazawa K, Ishikawa T, et al. Nomogram for 5-year relapse-free survival of a patient with advanced gastric cancer after surgery[J]. *Int J Surg*. 2016;35:153–9.
8. Necula L, Matei L, Dragu D, et al. Recent advances in gastric cancer early diagnosis[J]. *World J Gastroenterol*. 2019;25(17):2029–44.
9. Ye DM, Xu G, Ma W, et al. Significant function and research progress of biomarkers in gastric cancer[J]. *Oncol Lett*. 2020;19(1):17–29.
10. Guan K, Liu X, Li J, et al. Expression Status And Prognostic Value Of M6A-associated Genes in Gastric Cancer[J]. *J Cancer*. 2020;11(10):3027–40.
11. Pan JH, Zhou H, Cooper L, et al. LAYN Is a Prognostic Biomarker and Correlated With Immune Infiltrates in Gastric and Colon Cancers[J]. *Front Immunol*. 2019;10:6.
12. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation[J]. *Cell*. 2011;144(5):646–74.
13. Martinez-Outschoorn UE, Peiris-Pages M, Pestell RG, et al. Cancer metabolism: a therapeutic perspective[J]. *Nat Rev Clin Oncol*. 2017;14(2):113.
14. Warburg O, Wind F, Negelein E. The Metabolism of Tumors in the Body[J]. *J Gen Physiol*. 1927;8(6):519–30.
15. Liu Y, Zhang Z, Wang J, et al. Metabolic reprogramming results in abnormal glycolysis in gastric cancer: a review[J]. *Onco Targets Ther*. 2019;12:1195–204.

16. Ippolito L, Morandi A, Giannoni E, et al. Lactate: A Metabolic Driver in the Tumour Landscape[J]. *Trends Biochem Sci.* 2019;44(2):153–66.
17. Sonveaux P, Copetti T, De Saedeleer CJ, et al. Targeting the lactate transporter MCT1 in endothelial cells inhibits lactate-induced HIF-1 activation and tumor angiogenesis[J]. *PLoS One.* 2012;7(3):e33418.
18. Liu YQ, Chai RC, Wang YZ, et al. Amino acid metabolism-related gene expression-based risk signature can better predict overall survival for glioma[J]. *Cancer Sci.* 2019;110(1):321–33.
19. Wu F, Zhao Z, Chai RC, et al. Prognostic power of a lipid metabolism gene panel for diffuse gliomas[J]. *J Cell Mol Med.* 2019;23(11):7741–8.
20. Luo T, Li Y, Nie R, et al. Development and validation of metabolism-related gene signature in prognostic prediction of gastric cancer[J]. *Comput Struct Biotechnol J.* 2020;18:3217–29.
21. He L, Chen J, Xu F, et al. Prognostic Implication of a Metabolism-Associated Gene Signature in Lung Adenocarcinoma[J]. *Mol Ther Oncolytics.* 2020;19:265–77.
22. Liu Y, Liao XW, Qin YZ, et al. Identification of F5 as a Prognostic Biomarker in Patients with Gastric Cancer[J]. *Biomed Res Int.* 2020;2020: 9280841.
23. Yang JD, Ma L, Zhu Z. SERPINE1 as a cancer-promoting gene in gastric adenocarcinoma: facilitates tumour cell proliferation, migration, and invasion by regulating EMT[J]. *J Chemother.* 2019;31(7–8):408–18.
24. Li L, Zhu Z, Zhao Y, et al. FN1, SPARC, and SERPINE1 are highly expressed and significantly related to a poor prognosis of gastric adenocarcinoma revealed by microarray and bioinformatics[J]. *Sci Rep.* 2019;9(1):7827.
25. Liao P, Li W, Liu R, et al. Genome-scale analysis identifies SERPINE1 and SPARC as diagnostic and prognostic biomarkers in gastric cancer[J]. *Onco Targets Ther.* 2018;11:6969–80.
26. Zhang J, Yin Y, Niu XM, et al. CYP19A1 gene polymorphisms and risk of lung cancer[J]. *J Int Med Res.* 2013;41(3):735–42.
27. Slattery ML, Lundgreen A, Herrick JS, et al. Variation in the CYP19A1 gene and risk of colon and rectal cancer[J]. *Cancer Causes Control.* 2011;22(7):955–63.
28. Warburg O. The Chemical Constitution of Respiration Ferment[J]. *Science.* 1928;68(1767):437–43.
29. Angell H, Galon J. From the immune contexture to the Immunoscore: The role of prognostic and predictive immune markers in cancer[J]. *Curr Opin Immunol.* 2013;25(2):261–7.
30. Fridman WH, Pagès F, Sautès-Fridman C, et al. The immune contexture in human tumours: Impact on clinical outcome[J]. *Nat Rev Cancer.* 2012;12(4):298–306.
31. Zeng D, Zhou R, Yu Y, et al. Gene expression profiles for a prognostic immunoscore in gastric cancer[J]. *Br J Surg.* 2018;105(10):1338–48.
32. Mukai K, Tsai M, Saito H, et al. Mast cells as sources of cytokines, chemokines, and growth factors[J]. *Immunol Rev.* 2018;282(1):121–50.

33. Paolino G, Corsetti P, Moliterni E, et al. Mast cells and cancer[J]. *G Ital Dermatol Venereol*. 2019;154(6):650–68.
34. Sammarco G, Varricchi G, Ferraro V, et al. Mast Cells, Angiogenesis and Lymphangiogenesis in Human Gastric Cancer[J]. *Int J Mol Sci*, 2019, 20(9).
35. Husain Z, Huang Y, Seth P, et al. Tumor-derived lactate modifies antitumor immune response: effect on myeloid-derived suppressor cells and NK cells[J]. *J Immunol*. 2013;191(3):1486–95.
36. Fischer K, Hoffmann P, Voelkl S, et al. Inhibitory effect of tumor cell-derived lactic acid on human T cells[J]. *Blood*. 2007;109(9):3812–9.
37. Jackute J, Zemaitis M, Pranys D, et al. Distribution of M1 and M2 macrophages in tumor islets and stroma in relation to prognosis of non-small cell lung cancer[J]. *BMC Immunol*. 2018;19(1):3.
38. Baumann F, Leukel P, Doerfelt A, et al. Lactate promotes glioma migration by TGF-beta2-dependent regulation of matrix metalloproteinase-2[J]. *Neuro Oncol*. 2009;11(4):368–80.
39. Im JY, Kim DM, Park H, et al. VGLL1 phosphorylation and activation promotes gastric cancer malignancy via TGF-beta/ERK/RSK2 signaling[J]. *Biochim Biophys Acta Mol Cell Res*. 2021;1868(1):118892.
40. Zhang X, Zhang P, Shao M, et al. SALL4 activates TGF-beta/SMAD signaling pathway to induce EMT and promote gastric cancer metastasis[J]. *Cancer Manag Res*. 2018;10:4459–70.
41. Liu J, Dai X, Guo X, et al. Circ-OXCT1 Suppresses Gastric Cancer EMT and Metastasis by Attenuating TGF-beta Pathway Through the Circ-OXCT1/miR-136/SMAD4 Axis[J]. *Onco Targets Ther*. 2020;13:3987–98.
42. Song M, Ping Y, Zhang K, et al. Low-Dose IFNgamma Induces Tumor Cell Stemness in Tumor Microenvironment of Non-Small Cell Lung Cancer[J]. *Cancer Res*. 2019;79(14):3737–48.

Figures

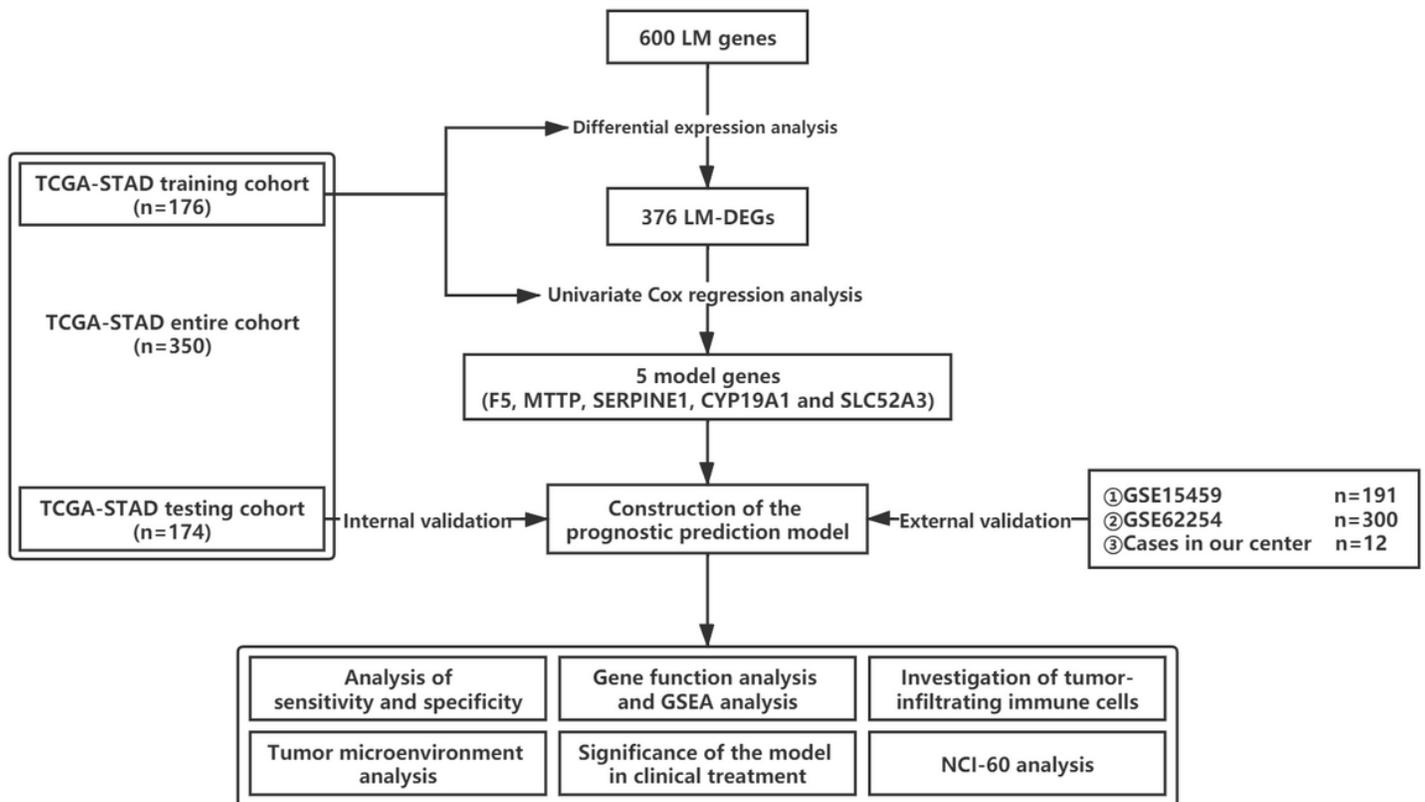


Figure 1

Flow diagram of the whole study.

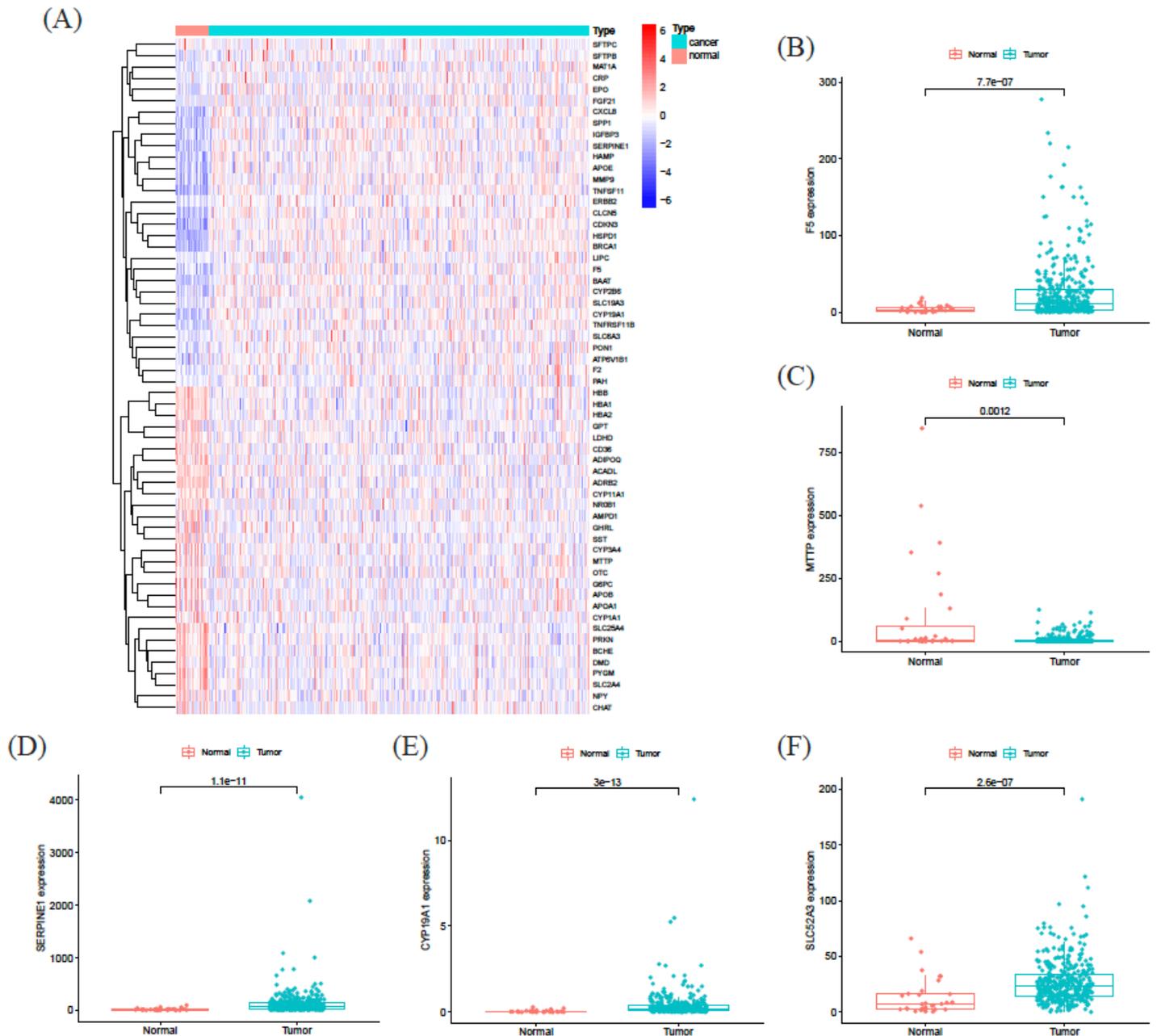


Figure 2

Screening of the model genes. (A) The gene-expression profiles of the top 60 DEGs. Red indicates genes with higher expression and blue indicates genes with lower expression. DEGs: differentially expressed genes. (B-F) Gene expressions of F5, MTTT, SERPINE1, CYP19A1 and SLC52A3 between GC and normal tissues, respectively.

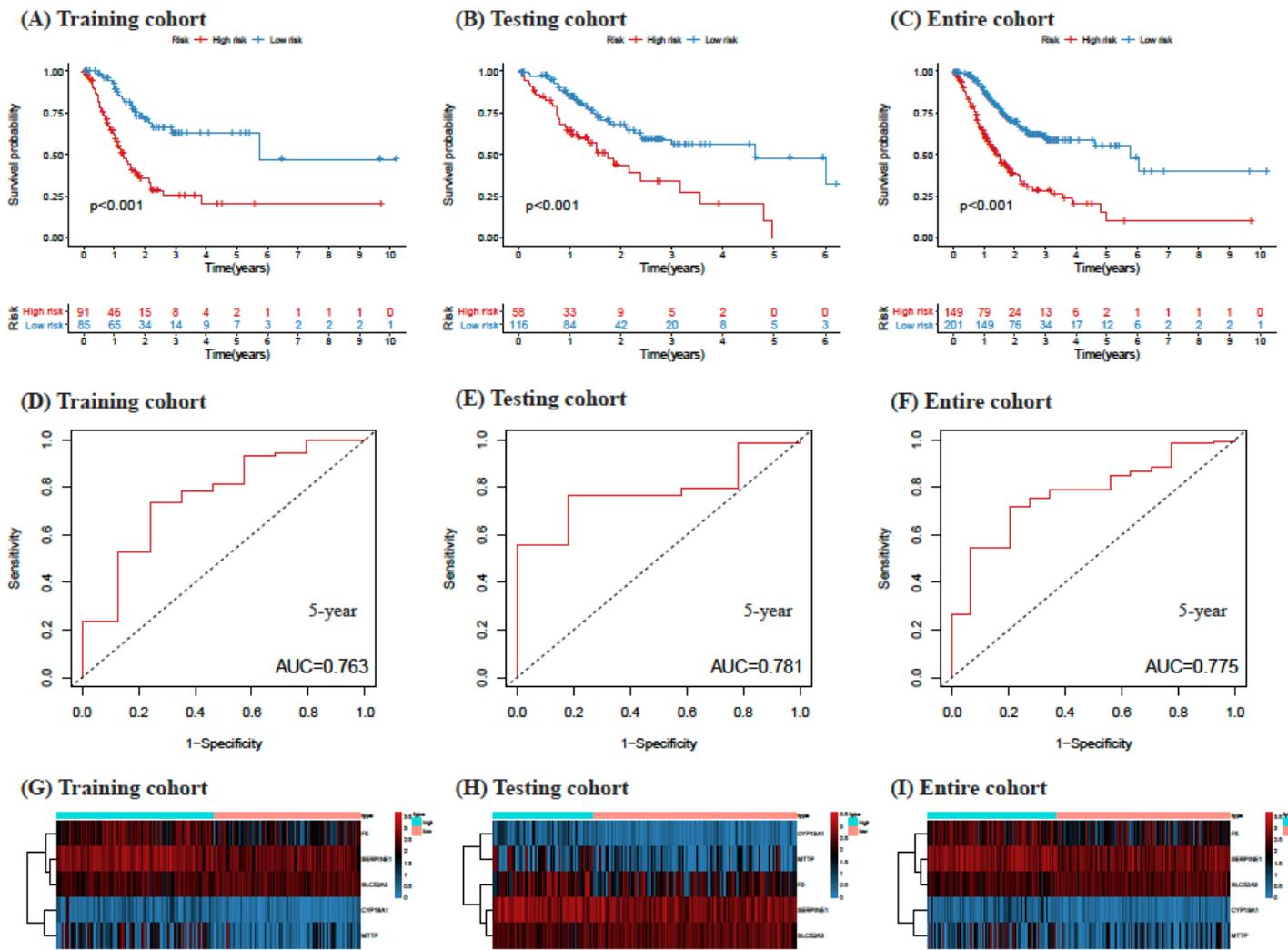


Figure 3

Kaplan-Meier (KM) survival analysis, receiver operating characteristic (ROC) analysis and heatmaps of prognostic prediction model in (A, D) training cohort, (B, E) testing cohort and (C, F) entire cohort, respectively. Distribution of model gene expressions (F5, MTTP, SERPINE1, CYP19A1 and SLC52A3) in (G) training cohort, (H) testing cohort and (I) entire cohort, respectively.

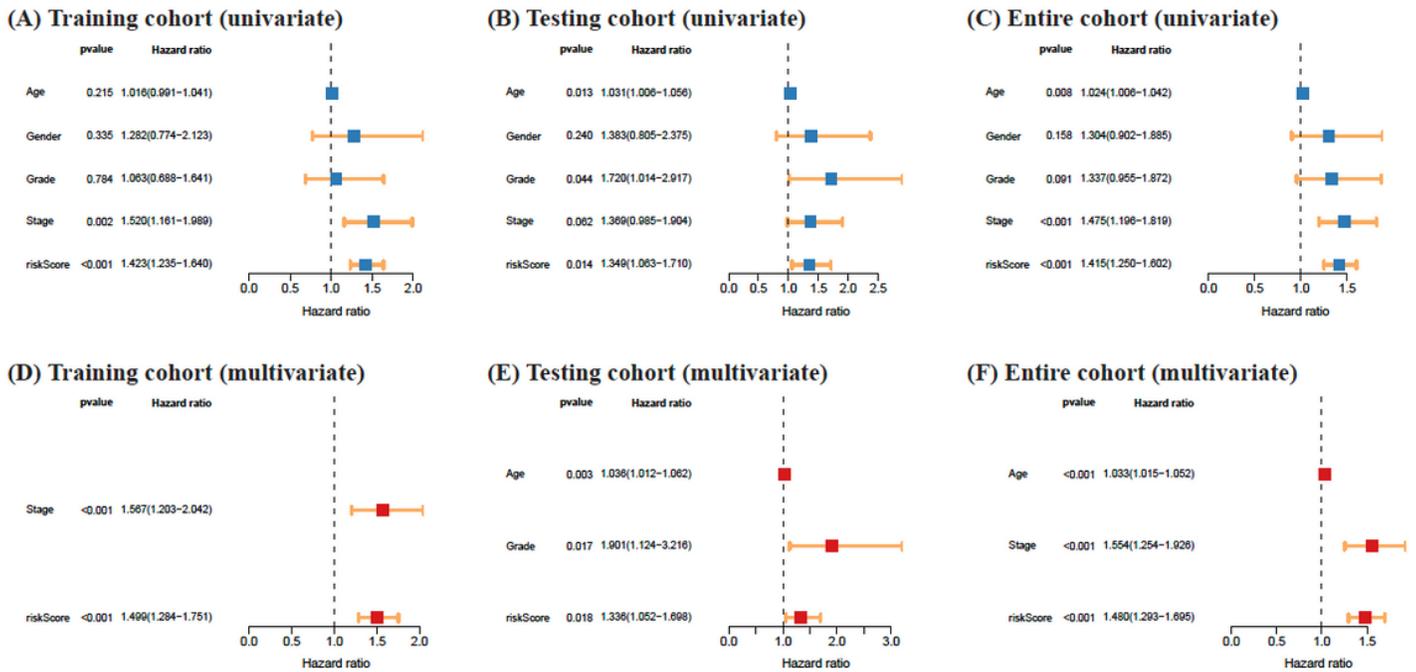


Figure 4

Univariate and multivariate Cox regression analyses in training, testing and entire cohorts. (A) Training cohort (univariate); (B) Testing cohort (univariate); (C) Entire cohort (univariate); (D) Training cohort (multivariate); (E) Testing cohort (multivariate); (F) Entire cohort (multivariate).

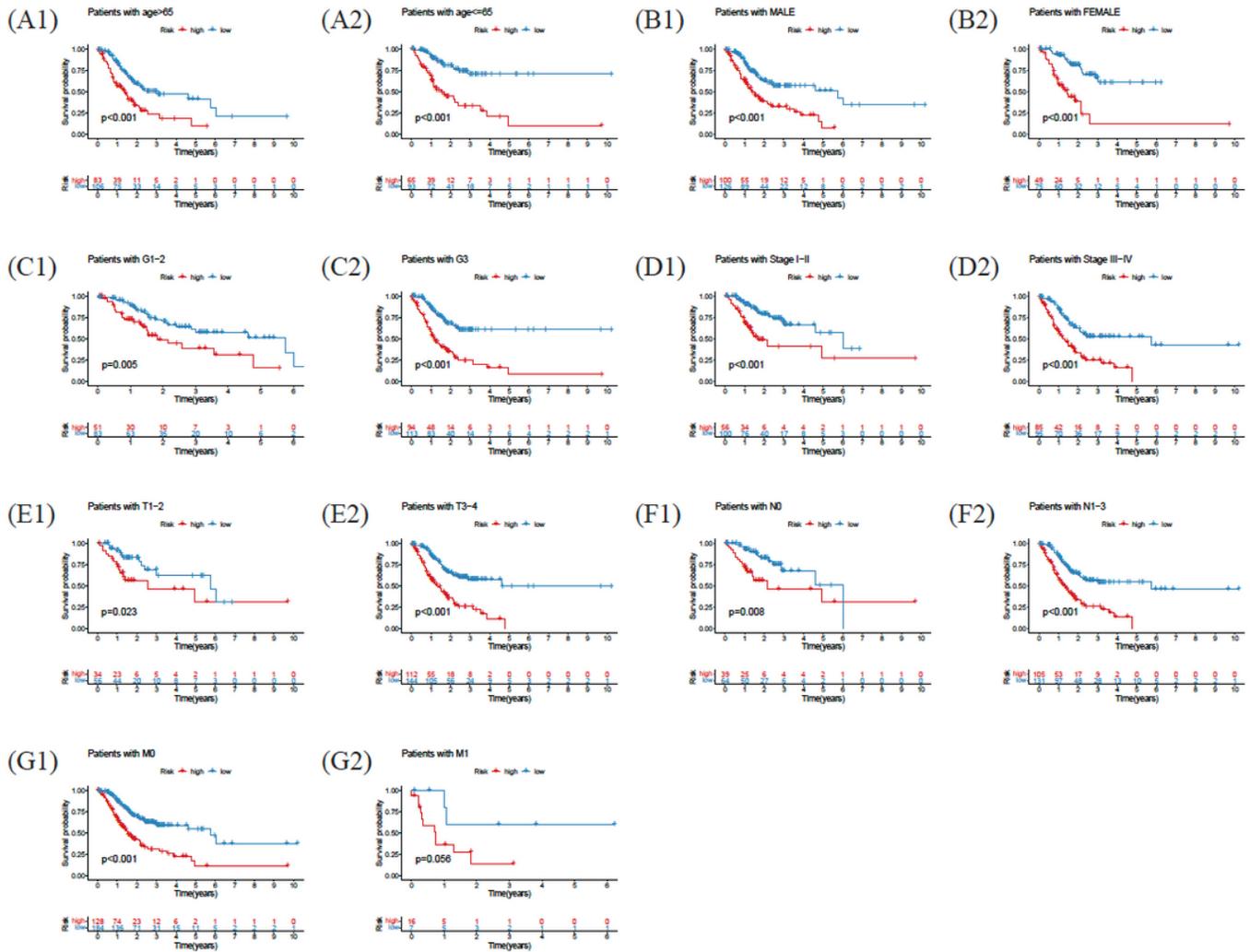


Figure 5

Kaplan-Meier (KM) survival analysis in different subgroups. (A1, A2) Patients with age >65 or ≤65 years old. (B1, B2) Male or female patients. (C1, C2) Patients with tumor grade of G1-2 or G3. (D1, D2) Patients with stage I-II or stage III-IV. (E1, E2) Patients with T1-2 or T3-4. (F1, F2) Patients with N0 and N1-3. (G1, G2) Patients with M0 and M1.

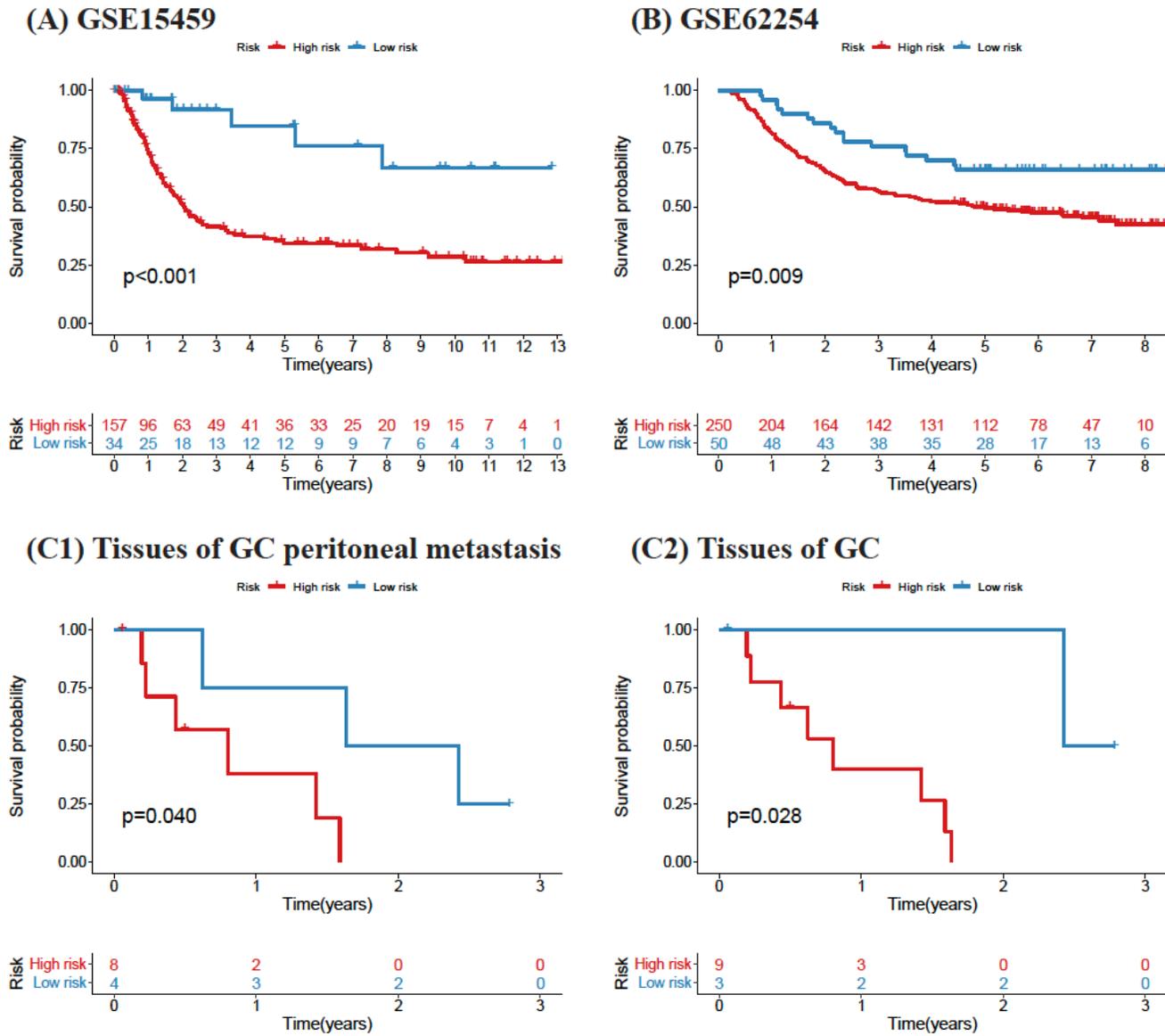


Figure 6

Kaplan-Meier (KM) survival analysis in independent datasets: (A) GSE15459, n=191; (B) GSE62254, n=300; (C1, C2) Cases in our center, n=12. (C1) Tissues of GC peritoneal metastasis; (C2) Tissues of GC.

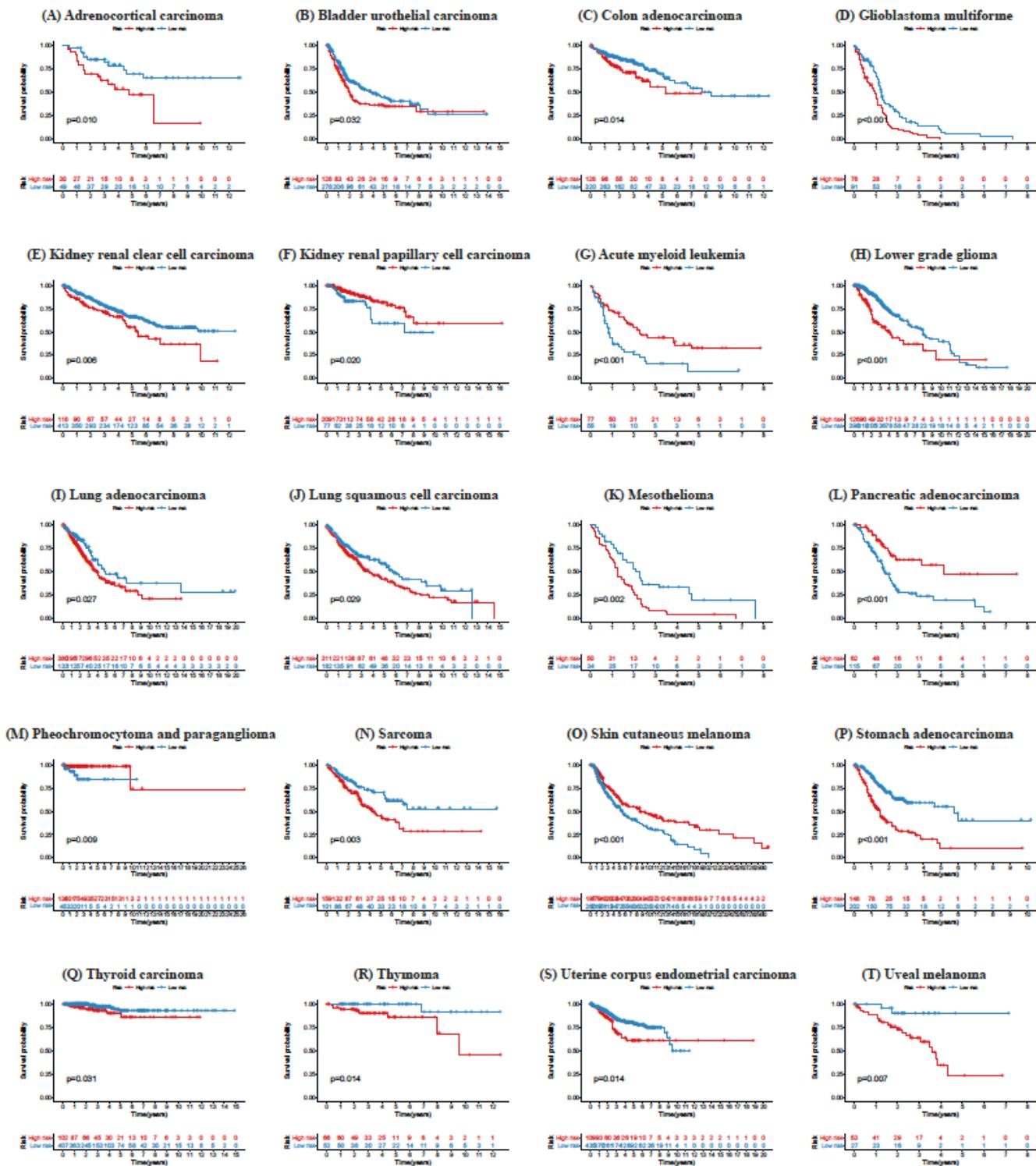


Figure 7

External validation in the independent datasets of 20 types of TCGA tumors. (A-T) Kaplan-Meier survival analysis between high and low risk groups in the 20 types of TCGA tumors.

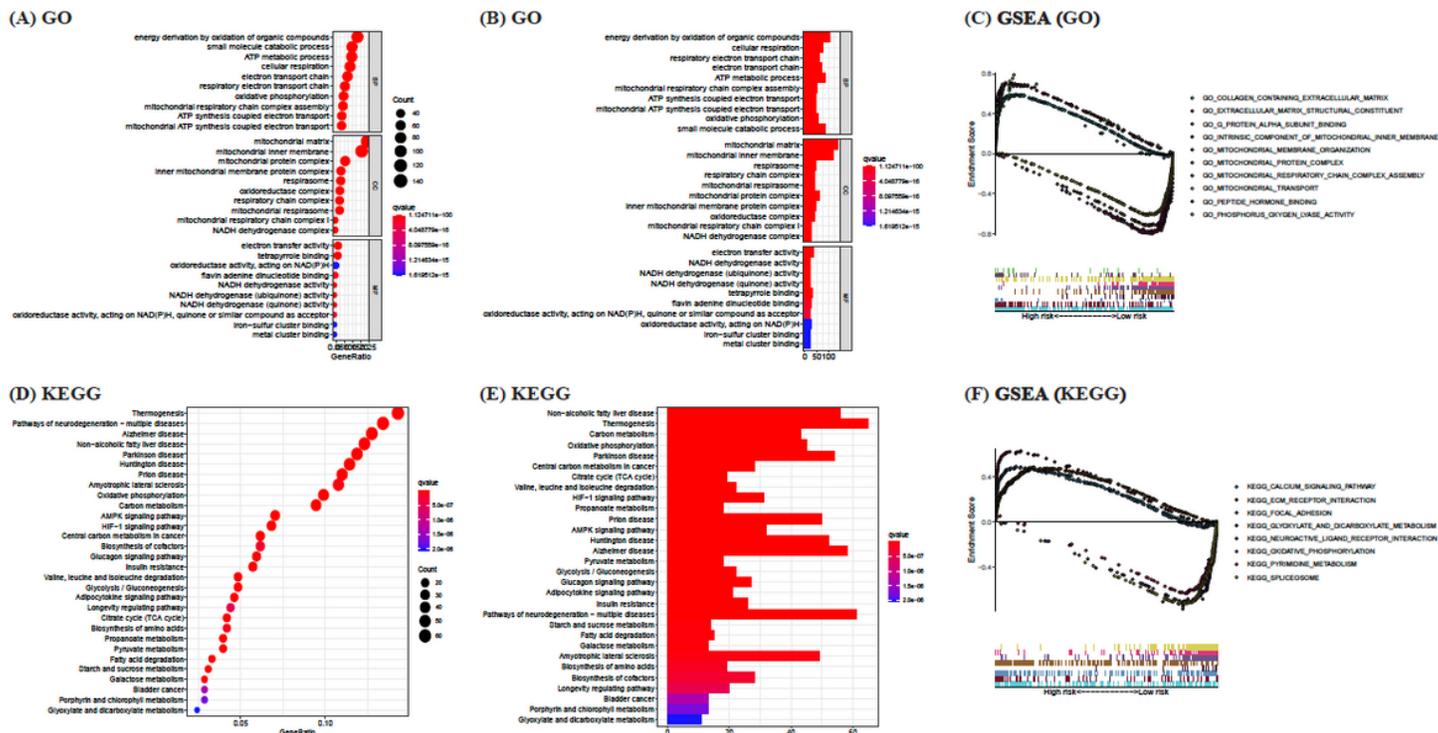


Figure 8

Gene function analysis and gene set enrichment analysis. (A, B) GO annotation analysis and (D, E) KEGG enrichment pathway analysis of LM-DEGs. (C, F) GSEA analysis between high and low risk groups. GO: Gene Ontology; KEGG: Kyoto encyclopedia of genes and genomes; GSEA: gene set enrichment analysis; BP: biological process; CC: cellular component; MF: molecular function.

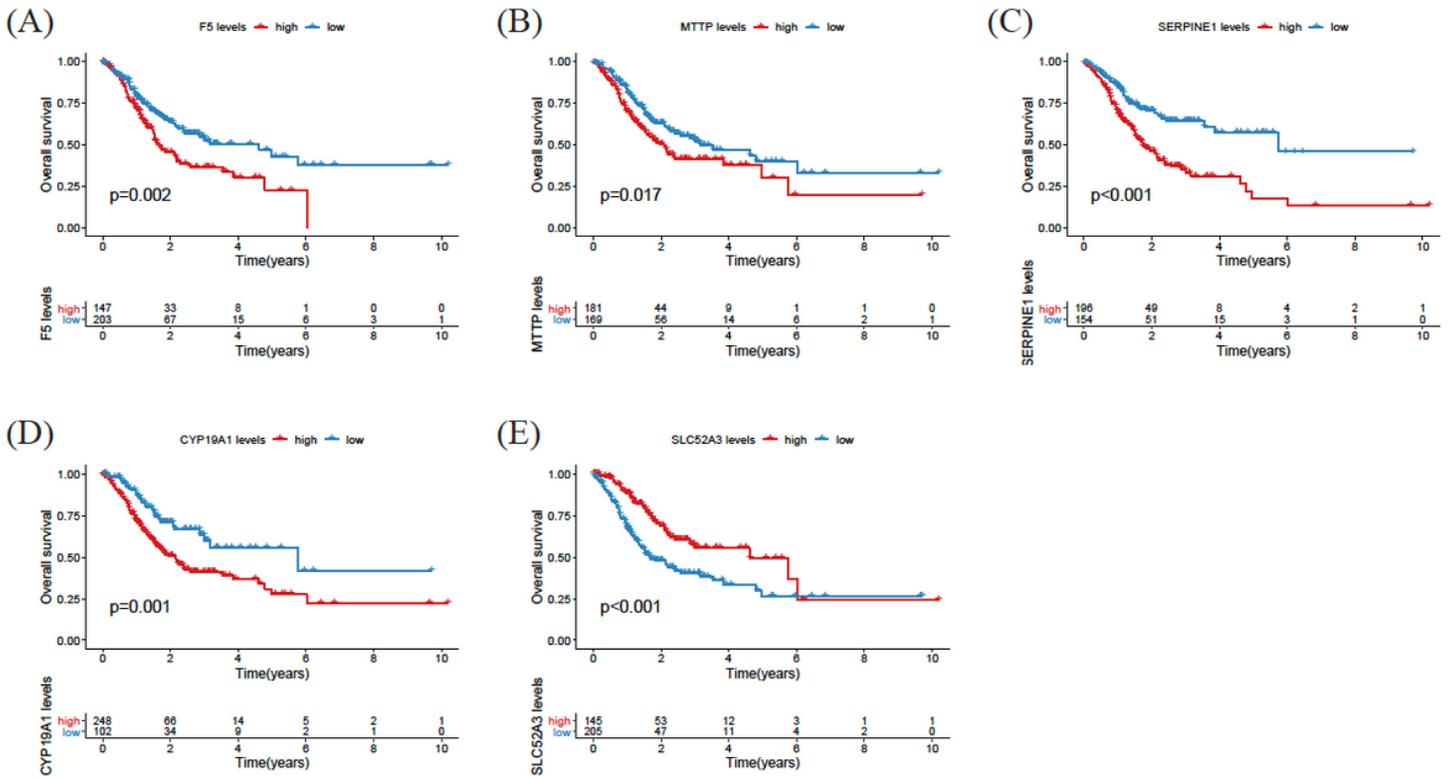


Figure 9

Impact of the five model genes on patients' OS. (A-E) Kaplan-Meier survival analysis of F5, MTTP, SERPINE1, CYP19A1 and SLC52A3, respectively.

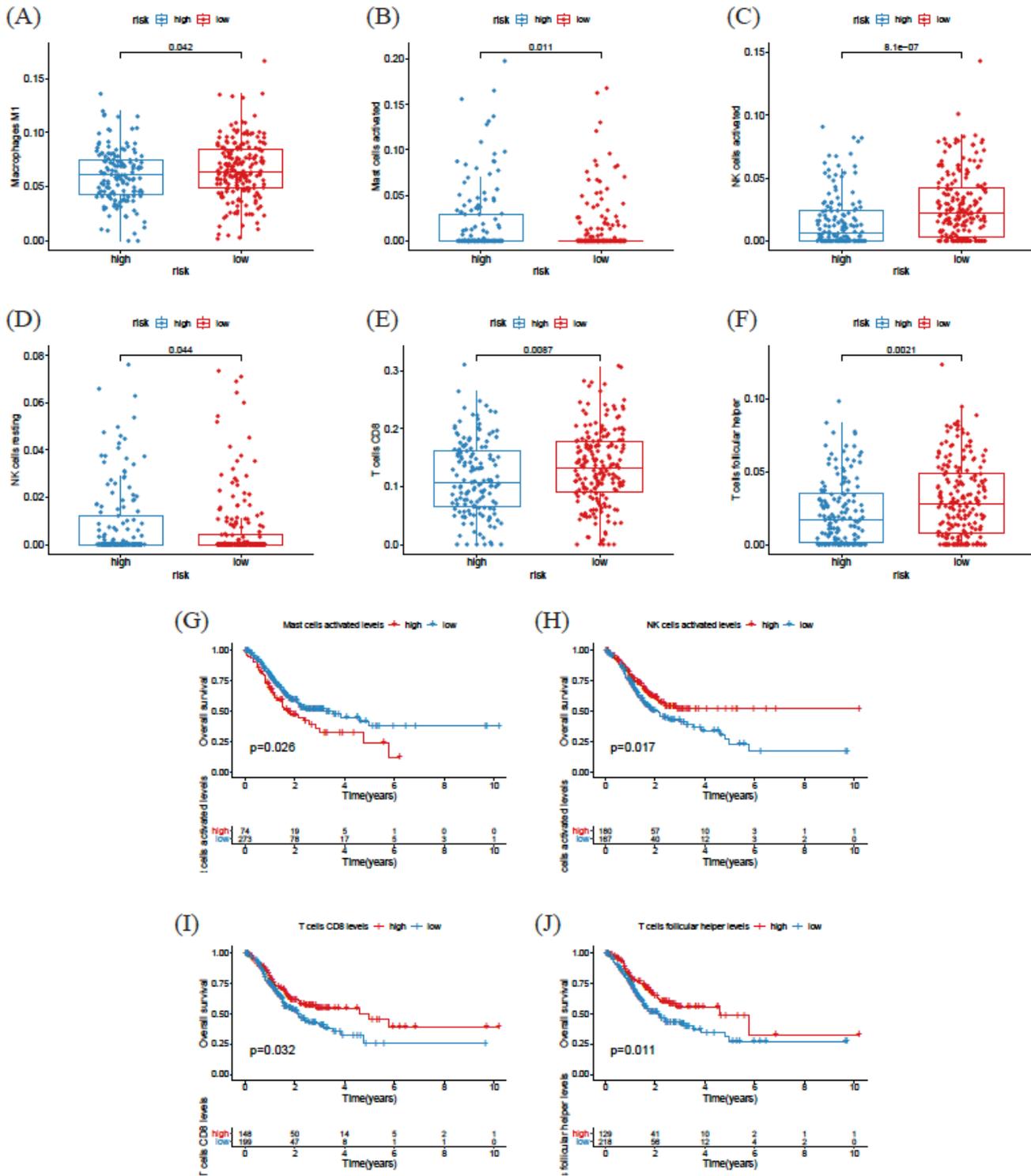


Figure 10

Investigation of tumor-infiltrating immune cells. (A-F) Immune infiltration levels of Macrophages M1, Mast cells activated, NK cells activated, NK cells resting, T cells CD8 and T cells follicular helper between high and risk groups, respectively. (G-J) Correlation between patients' OS and the levels of Mast cells activated, NK cells activated, T cells CD8 and T cells follicular helper, respectively.

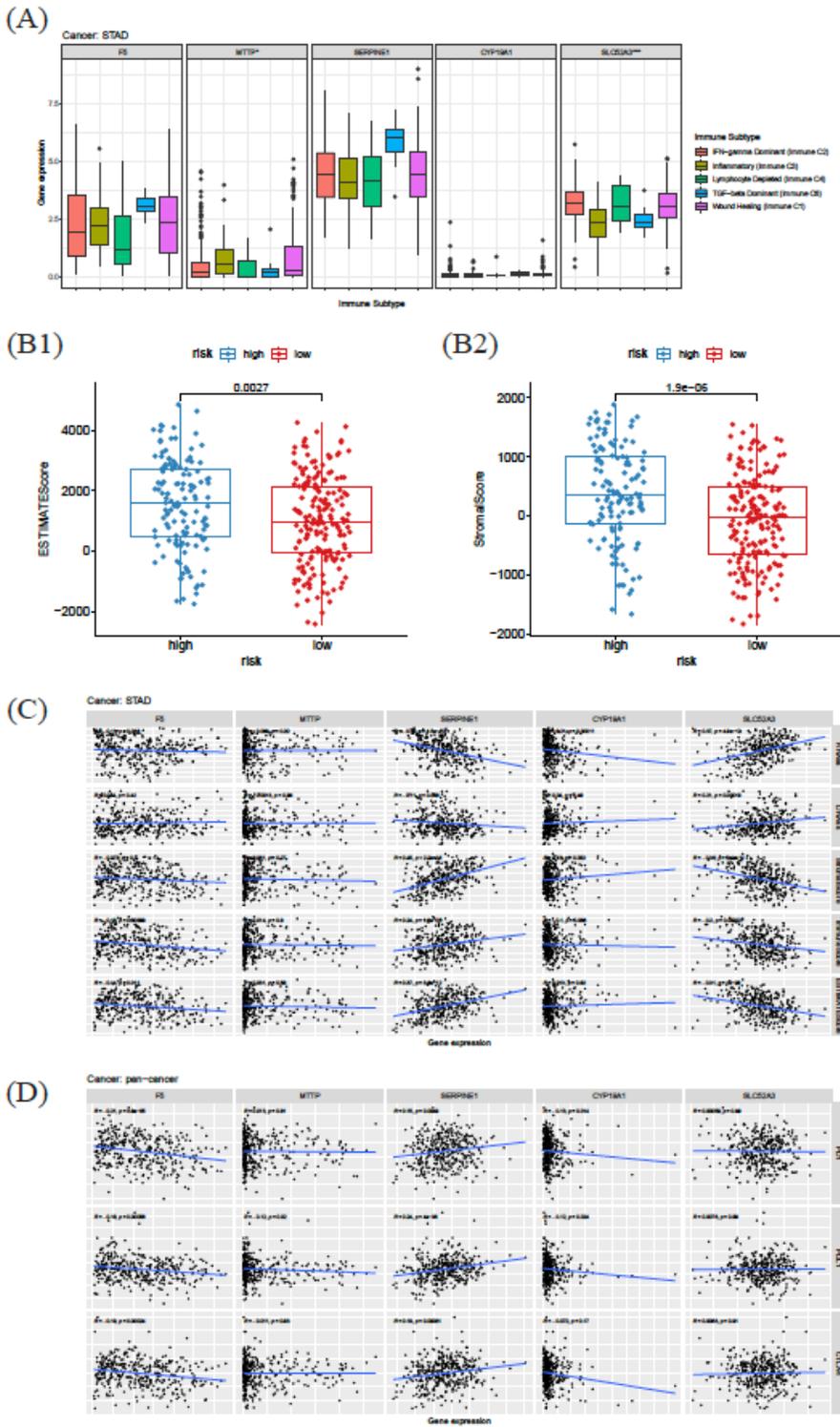


Figure 11

Tumor microenvironment analysis. (A) Association of the 5 model gene expressions with immune infiltrate subtypes. (B) The Estimate Score and Stromal Score between high and low risk groups. (C) The correlation between the expressions of 5 model genes and RNAss, DNAss, Stromal Score, Immune Score or Estimate Score. (D) The correlation between the expressions of the 5 model genes and immune checkpoint molecules (PD-1, PD-L1 and CTLA4).

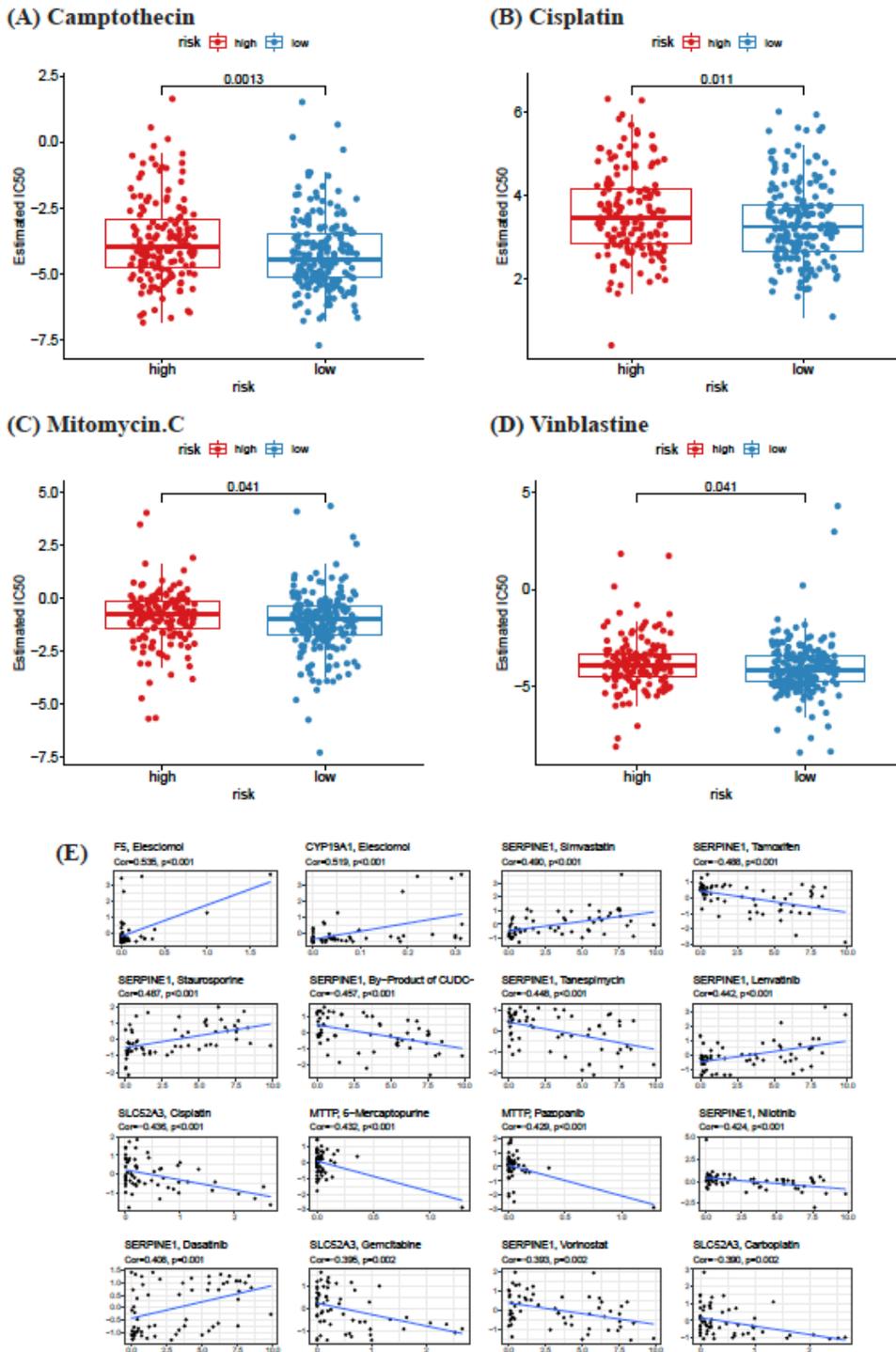


Figure 12

Drug sensitivity analysis. (A-D) The estimated IC50 of Camptothecin, Cisplatin, Mitomycin C and Vinblastine between high and low risk groups, respectively. (E) Correlation between the expressions of the 5 model genes and drug sensitivity.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Fig.S12.pdf](#)