Coagulation-related genes for thyroid cancer prognosis, immune infltration, staging, and drug sensitivity

Yuxiao Sun[1], Yifei Zhang[1], Yuchuan Yang[2], Weihao Liu[1] and Detao Yin[1,3,4]*

[1]Department of Thyroid Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [2]Department of Radiotherapy, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [3]Engineering Research Center of Multidisciplinary Diagnosis and Treatment of Thyroid Cancer of Henan Province, Zhengzhou, China, [4]Key Medicine Laboratory of Thyroid Cancer of Henan Province, Zhengzhou, China

Introduction: Thyroid cancer (THCA) is the most common endocrine tumor. Coagulation may be associated with the development of cancer, but its role in THCA patients is not yet clear.

Methods: In this study, we determined the predictive value of coagulation biomarker D-dimer for THCA patient lateral lymph node metastasis (LLNM) through receiver operating characteristics (ROC) analysis and logistic regression analysis. Subsequently, this study used the TCGA database to identify coagulation-related molecular subtypes through consensus clustering analysis and compared their prognosis. We identified coagulation-related genes (CRGs) associated with prognosis in thyroid cancer through gene expression data and clinical information, and constructed a prognostic model by selecting the prognostic CRGs using LASSO regression. Patients were divided into high-risk and low-risk groups based on the median score. Subsequently, prognosis, clinical characteristics, gene mutation occurrence, immune infiltration, function, and drug sensitivity of the two groups were analyzed. We also constructed a nomogram combining the model and clinical features. Finally, the expression of the prognostic CRGs was validated by RT-qPCR.

Results: D-dimers had better performance in predicting LLNM(the area under the curve was 0.656 (95% CI 0.580-0.733), with a cut-off value of 0.065 mg/l), and D-dimer>0.065mg/l was an independent predictor of LLNM. Then, we selected 8 prognostic CRGs to construct a predictive model. The prognosis of low-risk group patients was significantly better than that of high-risk group (P<0.001). The results showed significant differences in clinical characteristics, gene mutation occurrence, immune infiltration, function, and drug sensitivity between the high-risk and low-risk groups. We validated by qPCR that these 8 prognostic CRGs were overexpressed in THCA cell lines.

Discussion: Overall, this study provided an in-depth exploration of the potential role of the coagulation in thyroid cancer and its clinical significance, offering a new theoretical basis and research direction for personalized therapy and prognostic evaluation.

Introduction

Thyroid carcinoma (THCA) is the most prevalent endocrine tumor (1), and the incidence rate of THCA in China has significantly increased since 2000, with an estimated 22,000 new cases reported in 2022 (2). THCA is classified into three pathological types: differentiated thyroid cancer (DTC), medullary thyroid cancer (MTC), and anaplastic thyroid cancer (ATC) (3). Despite the generally good prognosis for most thyroid cancer patients, 10%-15% of patients experience recurrence, and 5% develop distant metastases (4). To further improve the prognosis of THCA patients, it is important to identify suitable prognostic biomarkers.

Coagulation, one of the hallmarks of tumor, could be a consequence of increasing plasma extravasation and vascular permeability which leads to extravascular coagulation, or be activated by disruption of vessels which leads to intravascular coagulation (5). A study has shown that tumor cells can release procoagulant factors, such as tissue factor, which may trigger the coagulation cascade both *in vitro* and *in vivo* (6). On the other hand, tumor coagulum, a cancer-driven network of molecular effectors favoring bleeding or thrombosis, could interact with the tumor microenvironment (TME) to orchestrate cancer inhibition or progression (7). Additionally, thyroid hormones directly regulate the transcription of genes encoding coagulation proteins in hepatocytes and endothelial cells (8). Therefore, the coagulation pathway may interact with the occurrence and development of THCA, but the role of coagulation in THCA patients remains unclear.

In this study, using The Cancer Genome Atlas (TCGA) cohort, we identified coagulation-related molecular subtypes via consensus clustering analysis and compared their prognoses. We used gene expression data and clinical information from TCGA's THCA to identify differentially expressed Coagulation-related genes (CRGs) in THCA and analyzed their correlation with prognosis. Through LASSO regression, we further filtered out 8 prognostic CRGs to construct a prognostic model and examined the relationships between risk scores and clinicopathological features, molecular functions, pathways, outcomes, immune infiltration, and immunotherapy. Additionally, we assessed the predictive performance of coagulation indicators for LLNM.

Methods

Patient selection and data collection

We gathered and screened data from patients admitted to the First Affiliated Hospital of Zhengzhou University from March 1, 2024, to May 1, 2024. The inclusion criteria were: (1) first-time thyroid surgery; (2) histologically confirmed papillary thyroid cancer (PTC); (3) age >18 years. The exclusion criteria were: (1) presence of other malignancies; (2) incomplete clinical data; (3) presence of diseases related to abnormal coagulation levels (including venous thromboembolism, disseminated intravascular coagulation, etc.). Finally, 408 patients were involved in our research. Age, gender, presence of Hashimoto's thyroiditis (HT), presence of extrathyroidal extension (ETE), multifocality of PTC, primary tumor size, and presence of lymph node metastasis (N stage, including central lymph node metastasis (CLNM) and lateral lymph node metastasis (LLNM)) were extracted from medical records. All patients underwent coagulation marker tests within two weeks before surgery. The antibody positivity often precedes clinical manifestations of thyroid dysfunction or sonographic changes in patients with HT. Studies have shown that elevated anti-TPO or anti-Tg antibodies can be present for years before the development of overt hypothyroidism or characteristic ultrasound changes, making antibody testing a valuable early diagnostic tool (9). HT was diagnosed by postoperative sectioning and examination of paraffin-embedded thyroid tissue specimens. Additionally, Serum antithyroglobulin and antithyroid peroxidase levels were measured within 30 days before surgery using the immuno-electrochemiluminescence method, and the patients were diagnosed with HT when these levels exceeded 115 IU/ml and 34 IU/ml, respectively. ETE referred to breaking through the thyroid capsule and invading adjacent soft tissues, muscles, trachea, oesophagus, nerves or blood vessels. Bilateral disease and multifocal disease were considered together for statistical analysis. Multifocality was defined as the presence of more than one lesion observed (10). Primary tumor size greater than 2 cm has been identified as an important risk factor for recurrence and lymph node metastasis in papillary thyroid carcinoma (3).

FIGURE 1
Flow chart of analysis.
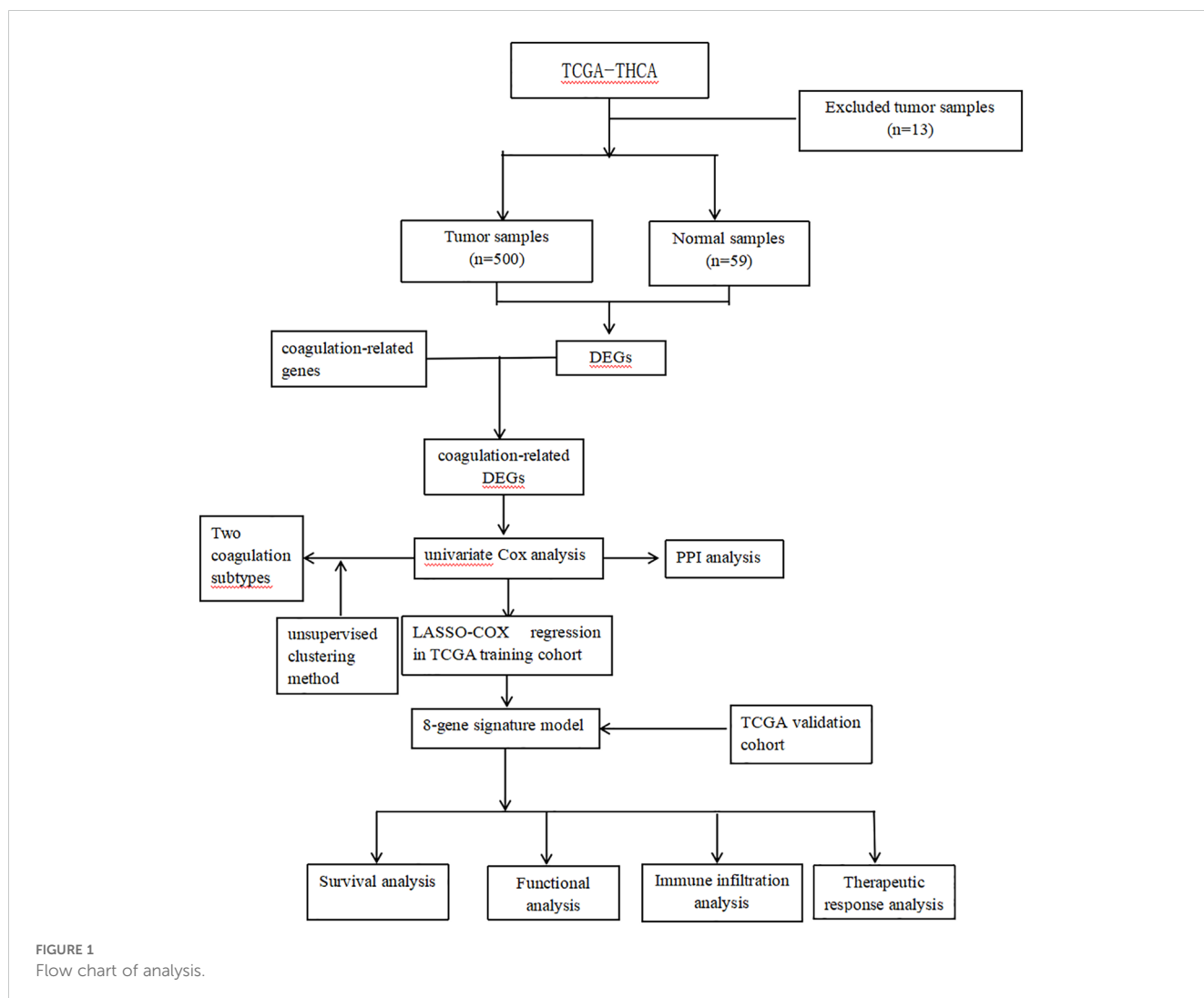
Data download and organization

The flow chart is shown in Figure 1. The RNA-seq data, clinical information and mutation data of THCA were acquired from the TCGA-THCA. After removing data with missing or less than 30 days of follow-up, missing outcome time occurrence status, and duplicates, the TCGA-THCA set remained 500 THCA samples. Based on the GeneCards website (https://www.genecards.org/), we retrieved 378 coagulation-related genes (CRGs, Relevance Score≥3) via searching the term "coagulation" (7, 11).

Collection of differentially expressed CRGs

The "limma" R package was conducted to figure out differentially expressed genes (DEGs) between normal and tumor tissues (FDR < 0.05 and log2FC >=1) in TCGA-THCA sets. Then, we took the intersection of these DEGs and CRGs to obtain differentially expressed CRGs (DECRGs). To explore the prognostic value of DECRGs in THCA, univariate Cox analysis was conducted in TCGA-THCA set (p < 0.05). Te co-expression

network of the prognostic DECRGs was explored by the "igraph" R package. Using the STRING (https://string-db.database, the protein-protein interaction (PPI) network org/) of proteins coded by the mitochondrial dynamic prognostic DECRGs was constructed and visualized. Finally, we validated these prognostic DECRGs using survival analyses and log-rank tests.

Identification of coagulation subtypes and survival analysis

Based on these prognostic DECRGs, we used the "ConsensusClusterPlus" package to perform consensus clustering on THCA samples from the TCGA-THCA dataset. To determine the optimal number of clusters, we used the Consensus Cumulative Distribution Function (CDF) to depict the CDF distribution patterns at different cluster numbers (k), and by plotting the Delta area plot to visually show the rate of change in the area under the CDF curve as the number of clusters increases from k to k+1, using this as a basis to select the optimal number of clusters and further subdivide TC patients into different subtypes. Then, we used the "survminer" package to plot

Kaplan-Meier survival curves to analyze the differences in progression-free interval (PFI) between different subtypes.

Construction of prognostic model

We randomly split the samples from the whole set into a train set and a test set according to a 6:4 ratio using the R package "caret". To further compress these prognostic DECRGs, we performed LASSO analysis using 10-fold cross-validation (12). Finally, we constructed a prognostic model using stepwise multivariate Cox regression analysis. The scores for each sample were calculated according to the following formula: Risk score = coefficient1 * gene 1 expression +…+ coefficientN * gene N expression, and the samples were classified into high or low risk groups based on the median score.

Validation of prognostic signature

Kaplan-Meier analysis with chi-squared test was applied to estimate survival differences between risk groups in the model. The test set, and whole set were applied to validate the internal stability of the model. The time-dependent receiver operating characteristic (ROC) curves were employed to analyze the predictive performance of the model and clinical characteristics through the R package "timeROC" (13). To assess the applicability of the model to patients with diverse clinical characteristics, we compared the survival differences between different risk score groups within each subgroup. To assess whether the model was an independent predictor for predicting patient prognosis, we performed univariate and multivariate Cox analyses including risk scores and clinical characteristics. To assess the clinical relevance of the model, the study examined the relationship between groups and clinical features. To better apply the model to clinical work, we constructed a nomogram combining the model and clinical features to predict the 3-, 5-, and 7-year survival probabilities of patients.

Enrichment analysis and somatic mutations

To explore the differences in somatic mutations between different risk groups, we performed analysis and visualization using the R package "maftools" (14). To find differences in molecular mechanisms and relevant pathways between risk groups, we recognized differentially expressed risk genes (DERGs) between risk groups (|logFC > 1| and FDR < 0.05) and conducted Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses (p < 0.05) (15–18).

Immune microenvironment and immunotherapy

Furthermore, we employed the EPIC, MCP-COUNTER, TIMER, XCELL, QUANTISEQ, CIBERSORT, and CIBERSORT-ABS algorithms to examine immune cell infiltration (19–25). Subsequently,

we scored the immune function of different risk groups using single-sample gene set enrichment analysis (ssGSEA) and compared the scores using the Wilcoxon test (26, 27). We also explored the differences in immune checkpoint gene expression levels between groups using the wilcox test. To predict the response to immunotherapy in each risk group, we calculated the tumor immune dysfunction and exclusion (TIDE) scores for each group and compared them (28). Finally, we used ESTIMATE to compare TME scores (StromalScore, ImmuneScore, and EstimateScore) across different risk groups.

Development of individualized anti-tumor treatment protocols

In this study, the calcPhenotype function in the "oncoPredict" R package was utilized to predict drug sensitivity in the Genomics of Drug Sensitivity in Cancer (GDSC) database (29).

Reverse transcription–qPCR (RT–qPCR)

Using qPCR technology to detect the differential expression of genes used to construct a prognostic model in human papillary thyroid carcinoma cell lines (IHH4, KTC-1, TPC-1) and normal thyroid cell lines (Nthy ori-3-1). First, TRIzol reagent (Invitrogen, USA) was used to extract total RNA from the cells. Subsequently, the extracted RNA was reverse-transcribed into cDNA using the Prime Script RT reagent kit (Takara, Japan). After obtaining the cDNA template, quantitative PCR was carried out according to the instructions of the SYBR Green Quantitative PCR Detection Kit (Takara, Japan). All the reactions were repeated for at least 3 times. GAPDH was used as an internal control gene, and the $2-\triangle\triangle Ct$ method was used for the relative quantification of the target genes. The primer sequences used in this study were: AZU1, Forward: 5′-AGAACCTGAACGACCTGATGC-3′ and Reverse: 5′-CCTGGGAAAACGGGAGAGA-3′; COL3A1, Forward: 5′-TTGCTGTGGTGGTGTTGGAG-3′ and Reverse: 5′-TTCTAGCGGGGTTTTTACGA-3′; CP, Forward: 5′-AGGAGATTCGGTCGTGTGGT-3′and Reverse: 5′-TTGAGGGAAGAGGTTTGCTG-3′; CSF2, Forward: 5′-AGAGACACTGCTGCTGAGATG-3′and Reverse: 5′-CAGGAAGTTTCCGGGGTT-3′; F12, Forward: 5′-AGGACCAGCGATGGGGATA-3′ and Reverse: 5′-TGTGGAAAAACCGGAGAAGC-3′; GNA14, Forward: 5′-TGTTACGACAGGAGGAGGGA-3′ and Reverse: 5′-CGAAGCACATCTTGTTGGGT-3′; IL1RN, 5′-AACAGAAAGCAGGACAAGCG-3′ and Reverse: 5′-CCTTCGTCAGGCATATTGGT-3′; SERPIND1, Forward: 5′-ATGGGTATGATTTCCTTAGGTCTG-3′ and Reverse: 5′-GGAAGAGATTATGAATGGTCGTG-3′; GAPDH, Forward: 5′-GGCAAATTCCATGGCACCG -3′ and Reverse: 5′-TCGCCCCACTTGATTTTGGA-3′.

Statistical analysis

All data analyses were completed using R software (version 4.3.9) and GraphPad Prism (version 8.0.2). ROC curves were

A

Features	N(%)
Total	408
Sex	
Female	333 (81.6)
Male	75 (18.4)
Age (years)	46.29 ±11.68
Hashimoto's thyroiditis	
Yes	102 (25)
No	306 (75)
Extrathyroidal extension	
Yes	85 (79.2)
No	323 (20.8)
Multifocality	
Yes	144 (35.3)
No	264 (64.7)
Primary tumor size (mm)	8.5 (6.5-12.5)
Primary tumor size>2cm	
Yes	38(9.3)
No	370(90.7)
N stage	
N0	221 (54.2)
N1a	143 (35.0)
N1b	44 (10.8)

B

Features	LLNM	non-LLNM	P
PT(s)	10.91±0.50	10.89±0.45	0.858
PTA(%)	101.95±137.56	102.92± 117.45	0.581
INR	0.99±0.01	0.987±0.01	0.849
APTT(s)	29.41±12.629	29.55±8.25	0.771
Fibrinogen(g/l)	2.74±0.35	2.97±10.23	0.642
TT(s)	15.84±1.93	15.51±1.92	0.136
D-dimer(mg/l)	0.30±0.34	0.145±0.03	<0.001

C

ROC Curves

D-dimer AUC = 0.656
TT AUC = 0.568
fibrinogen AUC = 0.575
APTT AUC = 0.524
INR AUC = 0.519
PTA AUC = 0.538
PT AUC = 0.517

(Sensitivity vs Specificity; legend: PT, PTA, INR, APTT, fibrinogen, TT, D-dimer)

D

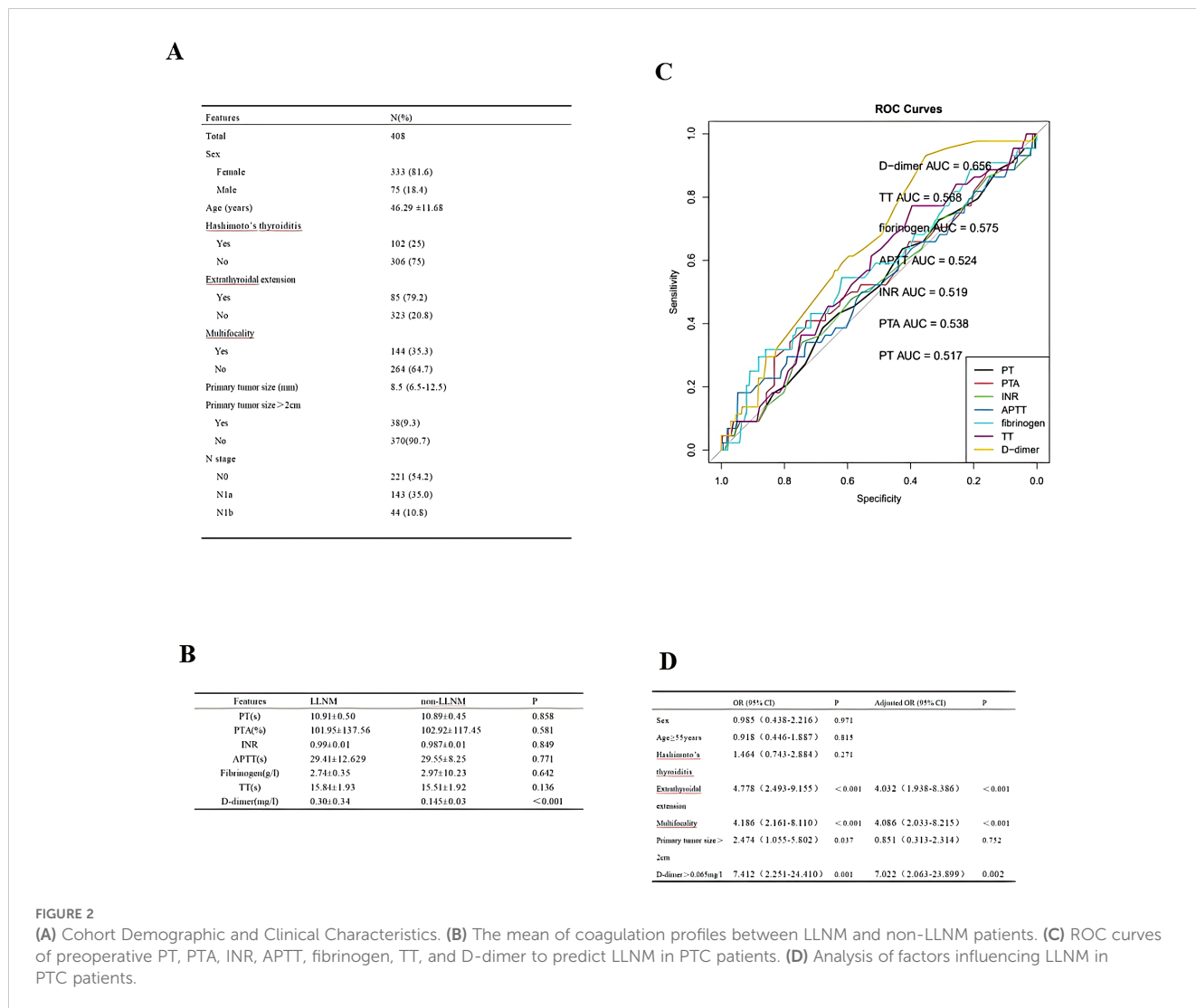|	OR (95% CI)	P	Adjusted OR (95% CI)	P
Sex	0.985 (0.438-2.216)	0.971		
Age≥55years	0.918 (0.446-1.887)	0.815		
Hashimoto's thyroiditis	1.464 (0.743-2.884)	0.271		
Extrathyroidal extension	4.778 (2.493-9.155)	<0.001	4.032 (1.938-8.386)	<0.001
Multifocality	4.186 (2.161-8.110)	<0.001	4.086 (2.033-8.215)	<0.001
Primary tumor size >2cm	2.474 (1.055-5.802)	0.037	0.851 (0.313-2.314)	0.752
D-dimer>0.065mg1	7.412 (2.251-24.410)	0.001	7.022 (2.063-23.899)	0.002

FIGURE 2

(A) Cohort Demographic and Clinical Characteristics. (B) The mean of coagulation profiles between LLNM and non-LLNM patients. (C) ROC curves of preoperative PT, PTA, INR, APTT, fibrinogen, TT, and D-dimer to predict LLNM in PTC patients. (D) Analysis of factors influencing LLNM in PTC patients.

applied to evaluate the area under the curve (AUC) for each coagulation index. We used univariate and multivariate logistic regression analyses to determine the association between clinical characteristics and LLNM. Spearman correlation analysis was used to evaluate the correlation between two continuous variables. The Wilcoxon rank-sum test or Student's T-test was used for the statistical analysis of two groups of continuous variables. Kaplan-Meier survival curves and log-rank tests were used to assess differences in DFS between different groups. P values less than 0.05 were considered to indicate statistical significance. ns indicates P>0.5, * indicates P<0.05, ** indicates P<0.01, *** indicates P<0.001.

Results

Predictive performance of coagulation indexes

We collected 408 cases of adult PTC, and their clinicopathological characteristics are shown in Figure 2A. The mean of coagulation profiles between LLNM and non-LLNM patients was compared in Figure 2B. The ROC curves revealed that D-dimer had superior predictive performance compared to prothrombin time (PT), prothrombin activity (PTA), international normalized ratio (INR), activated partial thromboplastin time (APTT), fibrinogen, and thrombin time (TT). The area under the curve (AUC) was 0.656 (95% CI 0.580-0.733), with a cut-off value of 0.065 mg/l (Figure 2C). The cut-off value for PT was 11.15s; for PTA, it was 130%; for INR, the cut-off value was 1.025; for APTT, it was 35.8s; for fibrinogen, the cut-off value was 3.175g/l; and for TT, it was 14.95s. Subsequently, through univariate and multivariate analyses, we identified the factors affecting LLNM, including extrathyroidal extension, multifocality, and D-dimer>0.065mg/l (Figure 2D).

Dysregulated CRGs in THCA and their prognosis

By comparing gene expression levels in tumor and normal tissue samples, we identified 1374 differentially expressed genes. The heatmap displaying differentially expressed genes is shown in
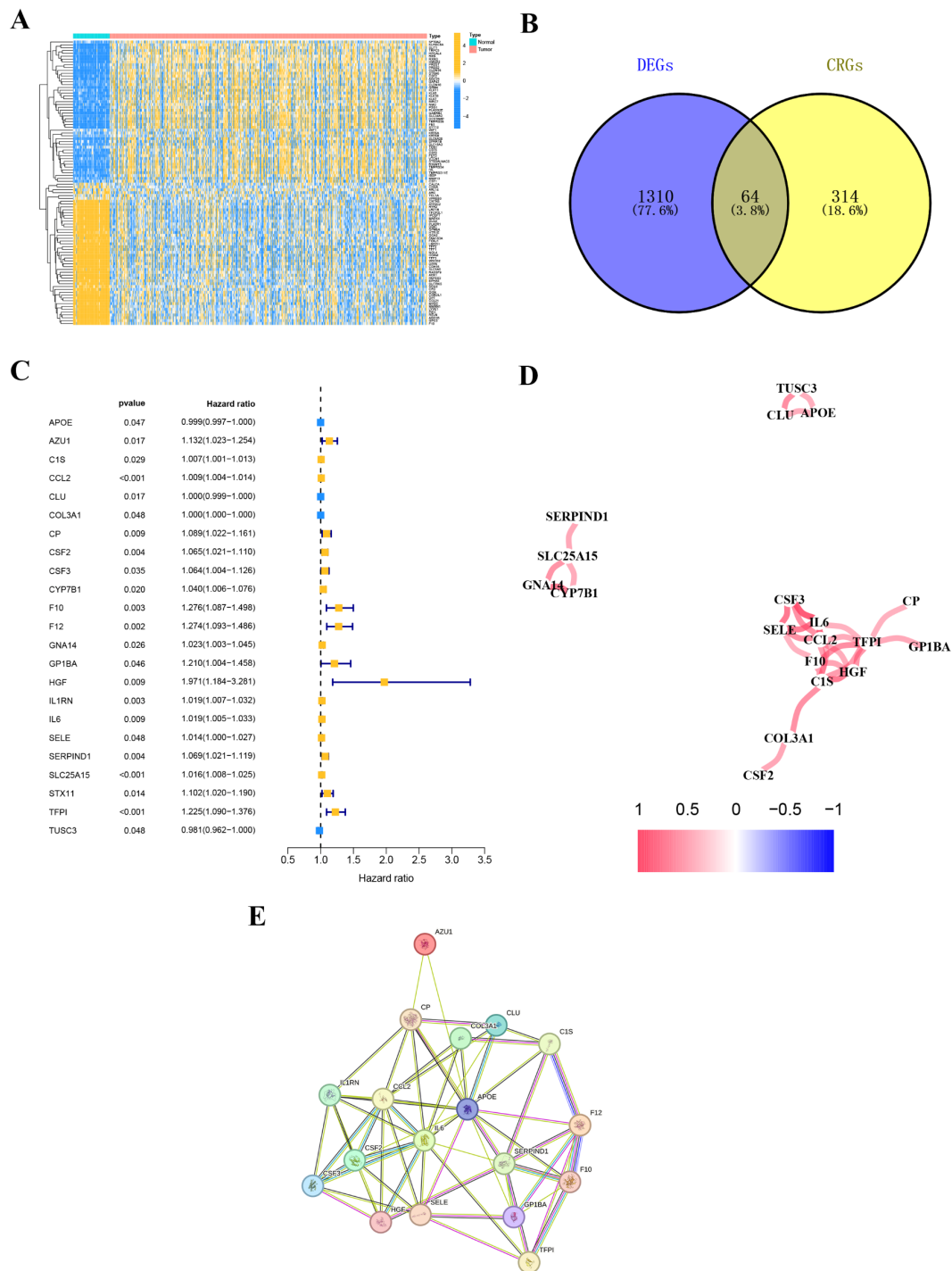
FIGURE 3
Coagulation-related differentially expressed genes. **(A)** The heatmap diagram for differential gene expression between THCA and normal tissues.
(B) The Venn plot displaying the overlap of differentially expressed coagulation-related genes (DE-CRGs). **(C)** Univariate Cox regression analysis.
(D) Correlation of the 23 prognostic crg. **(E)** A PPI network indicating the interactions among the 23 prognostic CRGs.

Figure 3A. For deeper analysis, we obtained 377 CRGs from the Genecards website (https://www.genecards.org), and 64 of these CRGs were differentially expressed between THCA tissues and normal controls as shown in the Venn plot (Figure 3B). We used univariate COX regression analysis to assess the relationship between the 64 CRGs and PFI, identifying 23 prognostic CRGs (Figure 3C). The correlations of these genes are shown in Figure 3D. Additionally, we constructed a PPI network of the 23 prognostic CRGs (Figure 3E). The Kaplan-Meier survival curves of 23 prognostic CRGs in THCA are shown in Figure 4.
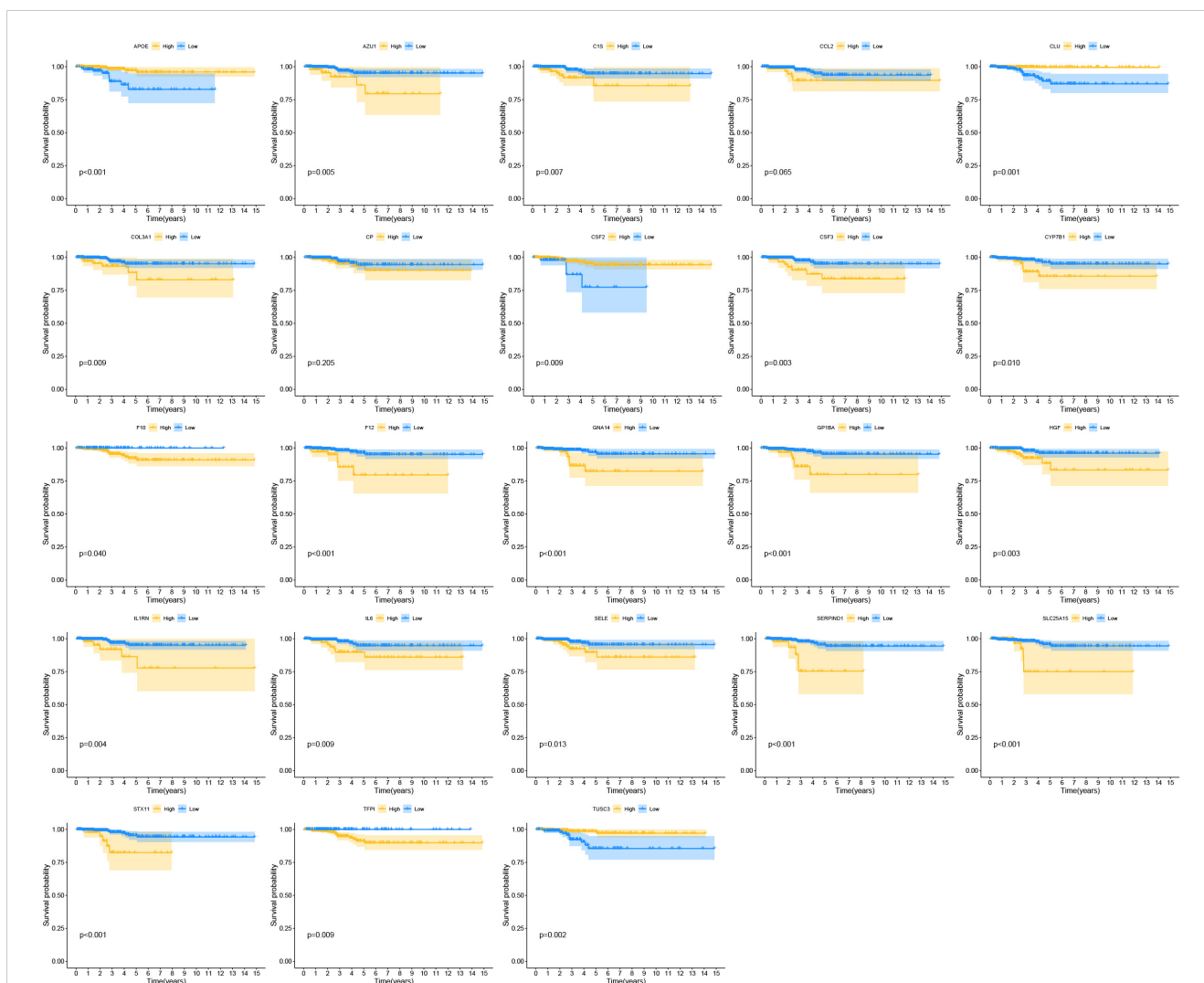
FIGURE 4
The Kaplan-Meier survival curves of 23 prognostic CRGs in THCA.

Identification of coagulation-related subtypes

We obtained transcriptomic data and corresponding clinical characteristics of patients from the TCGA-THCA cohort. Based on the 23 prognostic CRGs, we classified THCA patients into two subtypes using the unsupervised clustering method, including coagulation-related cluster 1 and cluster 2 (Figures 5A, B). We conducted K-M survival analysis on the TCGA-THCA cohort, and the results indicated that cluster 1 had a better prognosis than cluster 2 (p=0.014) (Figure 5C).

Prognostic model construction and validation

We divided the patients in the TCGA-THCA dataset into training and validation sets in a 6:4 ratio. In the training cohort of

TCGA-THCA, we conducted LASSO regression analysis (Figures 6A, B), screening out 8 key genes (AZU1, COL3A1, CP, CSF2, F12, GNA14, IL1RN, and SERPIND1). Using stepwise multivariate Cox regression analysis, we determined the coefficients associated with each gene (Figure 6C): Riskscore = 0.23947 * AZU1 + 0.00155 * COL3A1 + 0.08232 * CP + 0.03633 * CSF2 + 0.61566 * F12 + 0.09073 * GNA14 + 0.03401 * IL1RN + 0.03270 * SERPIND1.

We classified THCA patients in the training and validation sets into high-risk and low-risk groups based on the median risk score. Figures 6G–I provide detailed information on the risk score distribution and survival status of each THCA patient. Kaplan-Meier survival analysis curves were drawn to compare PFI differences between high- and low-risk groups in the training, validation, and overall sets. The results indicated that PFI was shorter in high-risk patients compared to low-risk patients (all P<0.05, Figures 6D–F), suggesting that the prognostic model has clinical utility in predicting THCA prognosis. Our study found that AZU1, COL3A1, CP, CSF2, F12, GNA14, IL1RN, and SERPIND1 were overexpressed in the high-risk group (Figures 6J–L).
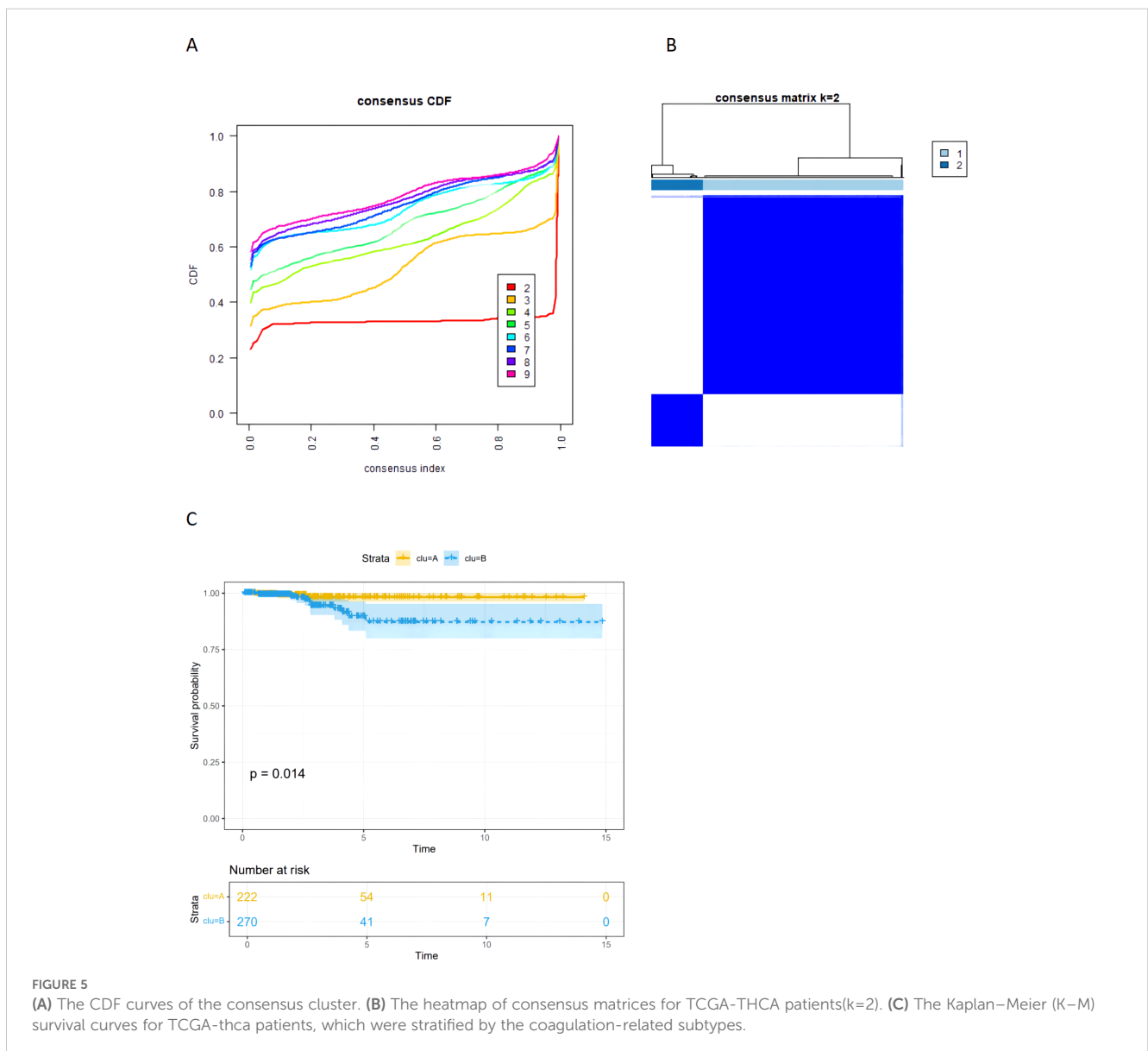
Validation of the 8-Gene signature

The ROC curves further showcased the excellent performance of the prognostic model in predicting PFI. We found that the AUC for the prognostic features in the training set reached 1.000, 0.901, and 0.924 at 1, 3, and 5 years, respectively. Likewise, in the test set, the AUC were 0.790, 0.777, and 0.828, respectively. For the entire cohort, the AUC reached 0.865, 0.860, and 0.877, respectively (Figures 7A, C, E). Figures 7B, D, F illustrate the ROC curves for the risk score and different clinical characteristics predicting 5-year PFI.

To assess the applicability of the model for patients with various clinical characteristics, we compared survival differences between different risk score groups within each subgroup (Figure 8). We found that the prognostic model exhibited good performance in predicting PFI across subgroups including male, female, N0, N1, Stage II-IV, M0, T1-2, and T3-4 (Figure 8).

A heatmap illustrated the distribution of clinical characteristics and gene expression in the high-risk and low-risk groups (Figure 9B). To assess the independence of the risk score as a prognostic indicator for THCA, univariate and multivariate Cox regression analyses were conducted in the TCGA-THCA cohort. The analyses consistently affirmed the risk score's status as an independent prognostic predictor (all p < 0.05) (Figure 9A). We created a nomogram that integrates the risk score and clinical characteristics to enhance the clinical application of our results (Figure 9C). Figures 10A, B depict the gene mutation occurrences in the low-risk and high-risk groups, respectively. In the low-risk group, the somatic mutation frequency rankings were BRAF (75%),
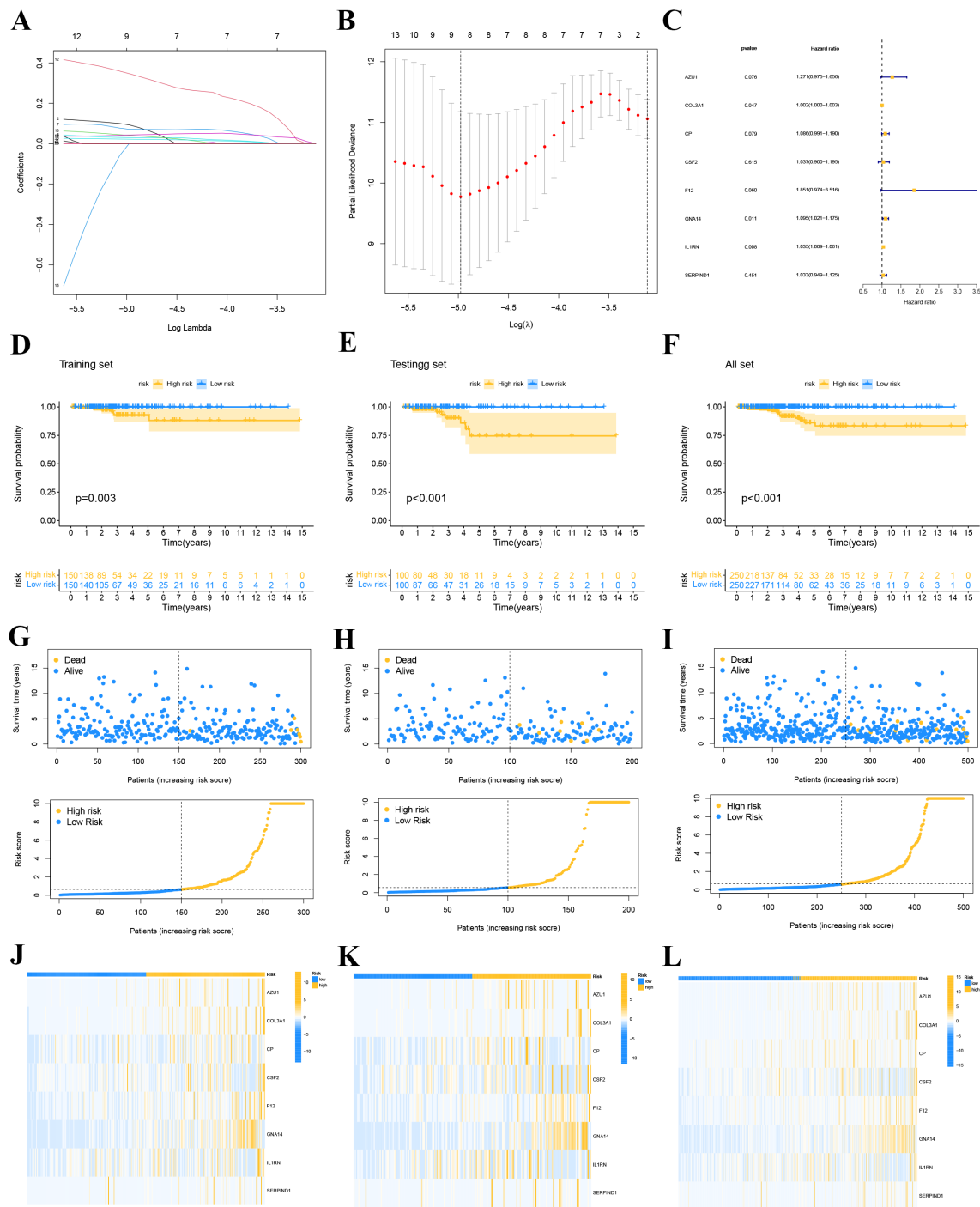

FIGURE 5
(A) The CDF curves of the consensus cluster. **(B)** The heatmap of consensus matrices for TCGA-THCA patients(k=2). **(C)** The Kaplan−Meier (K−M) survival curves for TCGA-thca patients, which were stratified by the coagulation-related subtypes.

FIGURE 6

LASSO regression model. **(A)** Coefficient screening was performed based to LASSO analysis. **(B)** Parameters were adjusted by ten cross-validation. **(C)** stepwise multivariate Cox regression analysis. **(D-F)** Survival curve for high and low risk groups in the training set **(D)**, the validation set **(E)** and whole set **(F)**. **(G-I)** Distribution of risk scores and survival status of each THCA patient in the training set **(G)**, validation set **(H)** and whole set **(I)**. **(J-L)** Expression levels of the gene of each THCA patient in the training set **(J)**, validation se t **(K)** and whole set **(L)**.

NRAS (4%), TG (4%), with mutation frequencies of other genes below 4%. In the high-risk group, the somatic mutation frequency rankings were BRAF (43%), NRAS (13%), HRAS (6%), TG (4%), with mutation frequencies of other genes below 4%. It is noteworthy

that the primary type of somatic mutation in both high- and low-risk groups was missense mutation, indicating that missense mutations may play a crucial role in the common mechanisms of tumorigenesis across different risk levels.
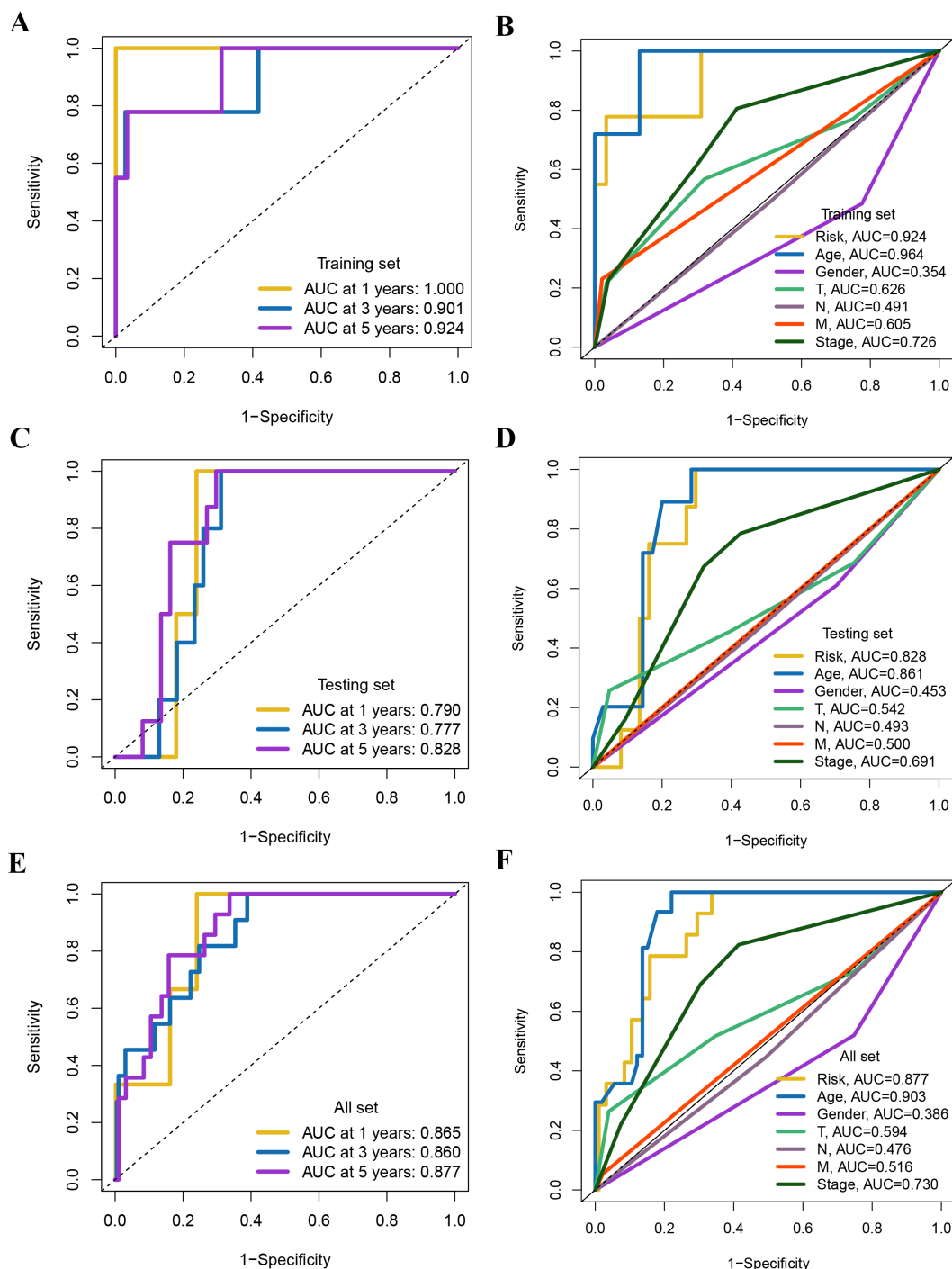
FIGURE 7
(A, C, E) ROC curves for the training set, validation set, and total set based on risk score versus survival state. **(B, D, F)** ROC curves of risk score and each clinical feature for predicting 5-year PFI in the training set, the validation set and whole set.

Functional, immune infiltration analyses and drug sensitivity analysis in different risk groups

We examined the biological functions and pathways of the high-risk and low-risk groups using GO and KEGG pathway analyses. GO analysis indicated that the Biological process, cellular component, and

Molecular Function affected in the low-risk group mainly included antimicrobial peptide production, extracellular matrix disassembly, lamellar body, neuron to neuron synapse, hippocampal mossy fiber to CA3 synapse, serine-type endopeptidase activity, and serine-type peptidase activity (Figure 10C). KEGG analysis revealed that the significantly enriched pathways in the low-risk group were the Ras signaling pathway, Taste transduction, Pancreatic secretion, and
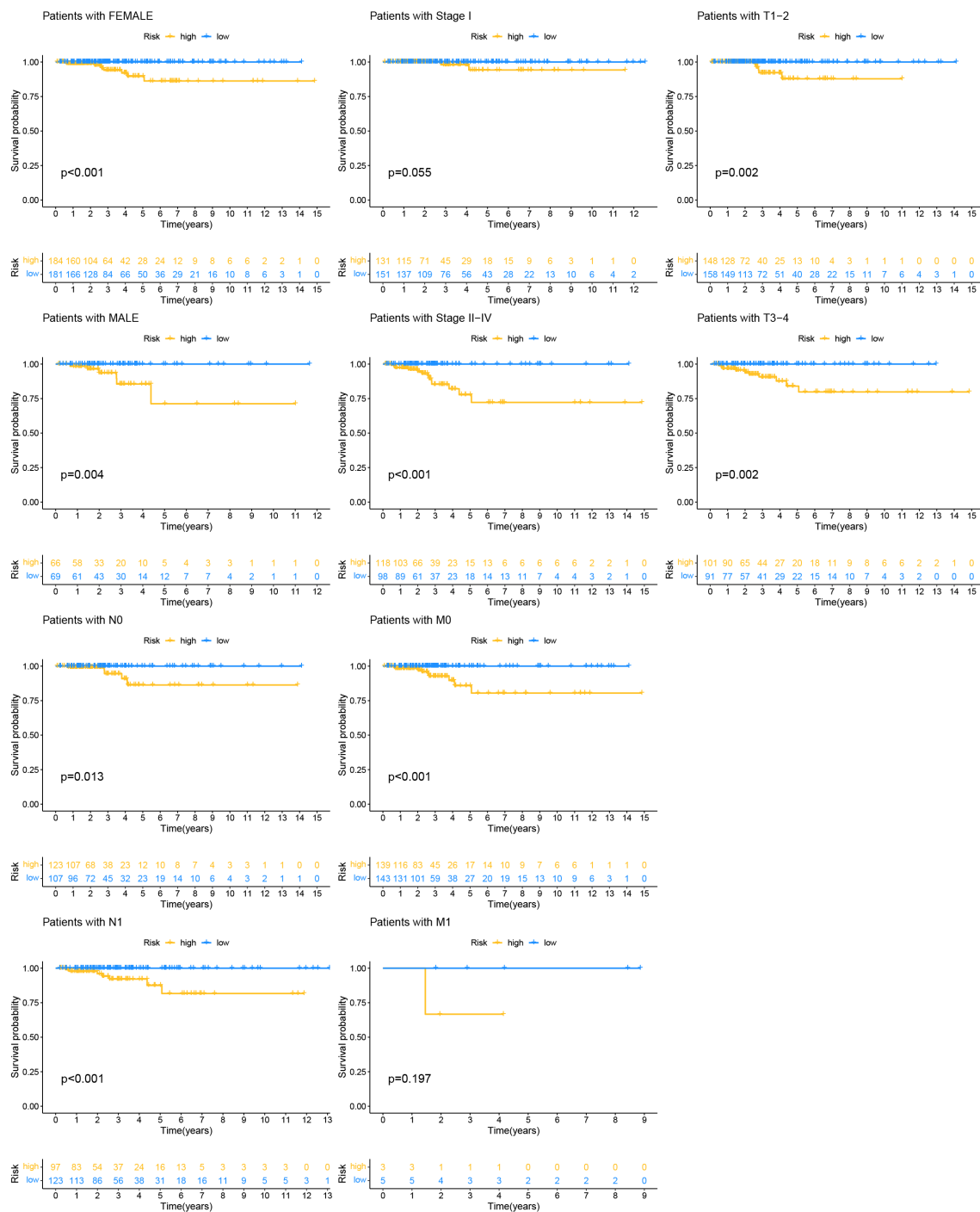
FIGURE 8
Kaplan-Meier survival analysis based on the grouping of each clinical feature and risk score.

Neuroactive ligand-receptor interaction (Figure 10D). GO analysis indicated that the Biological process, cellular component, and Molecular Function affected in the high-risk group mainly included ossification, osteoblast differentiation, collagen-containing extracellular matrix, collagen trimer, extracellular matrix structural constituent, and extracellular matrix structural constituent conferring tensile strength (Figure 10E). KEGG analysis revealed that the significantly enriched pathways in the high-risk group were Protein

digestion and absorption, Cytoskeleton in muscle cells, and ECM-receptor interaction (Figure 10F).

We further analyzed the differences in immune infiltration levels between the high-risk and low-risk groups. The results indicated that the risk score was negatively correlated with the infiltration levels of macrophage M1 and M2 subtypes, CD8+ T cells, and NK cells, with correlation coefficients all below -0.2 (Figure 11A), suggesting that the role of these cells may weaken as the risk score increases. The
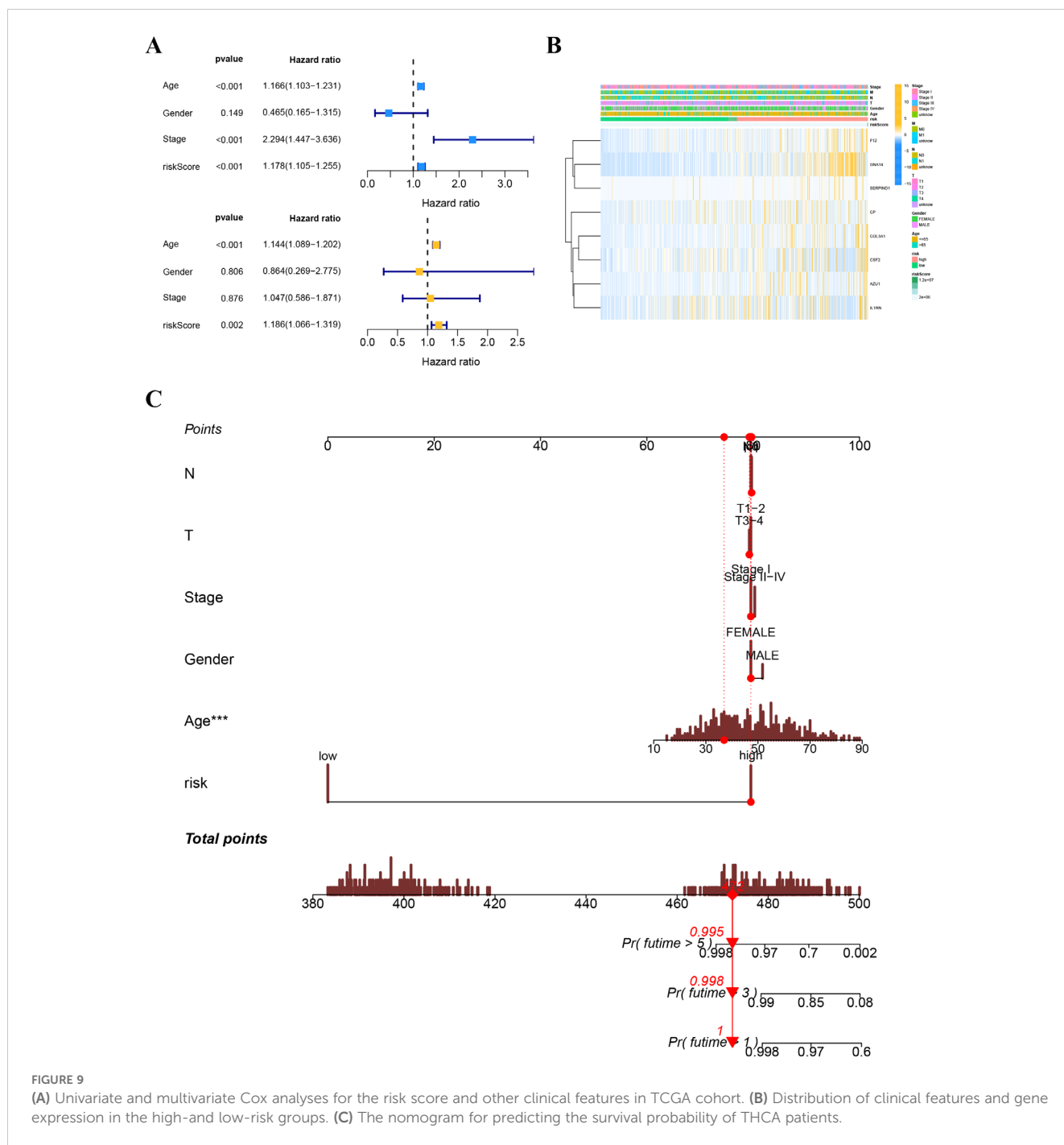
FIGURE 9
(A) Univariate and multivariate Cox analyses for the risk score and other clinical features in TCGA cohort. **(B)** Distribution of clinical features and gene expression in the high-and low-risk groups. **(C)** The nomogram for predicting the survival probability of THCA patients.

immune function analysis of different risk groups is depicted in Figure 11B. The differential expression analysis of immune checkpoints between different risk groups showed that immune checkpoints significantly expressed in the low-risk group compared to the high-risk group included CD274, TMIGD2, and TNFRSF18, while immune checkpoints significantly expressed in the high-risk group compared to the low-risk group included CD27, IDO2, and NRP1 (Figure 12A). We also calculated the Tumor Immune Dysfunction and Exclusion (TIDE) score for each group and compared them. The results indicated that the high-risk group had

higher TIDE scores, and higher TIDE scores were associated with poorer prognosis. The group with high TIDE scores and high-risk scores had a relatively worse prognosis compared to other groups (Figures 12B–D). TME scores also suggested better prognosis for the low-risk group (Figure 12E).

Drug sensitivity analysis results indicated that drugs such as Cisplatin, Dactinomycin.1, Docetaxel, Erlotinib, and Fludarabine had higher sensitivity in the high-risk group, implying they might have greater potential in the treatment of patients in the high-risk group (Figure 13).
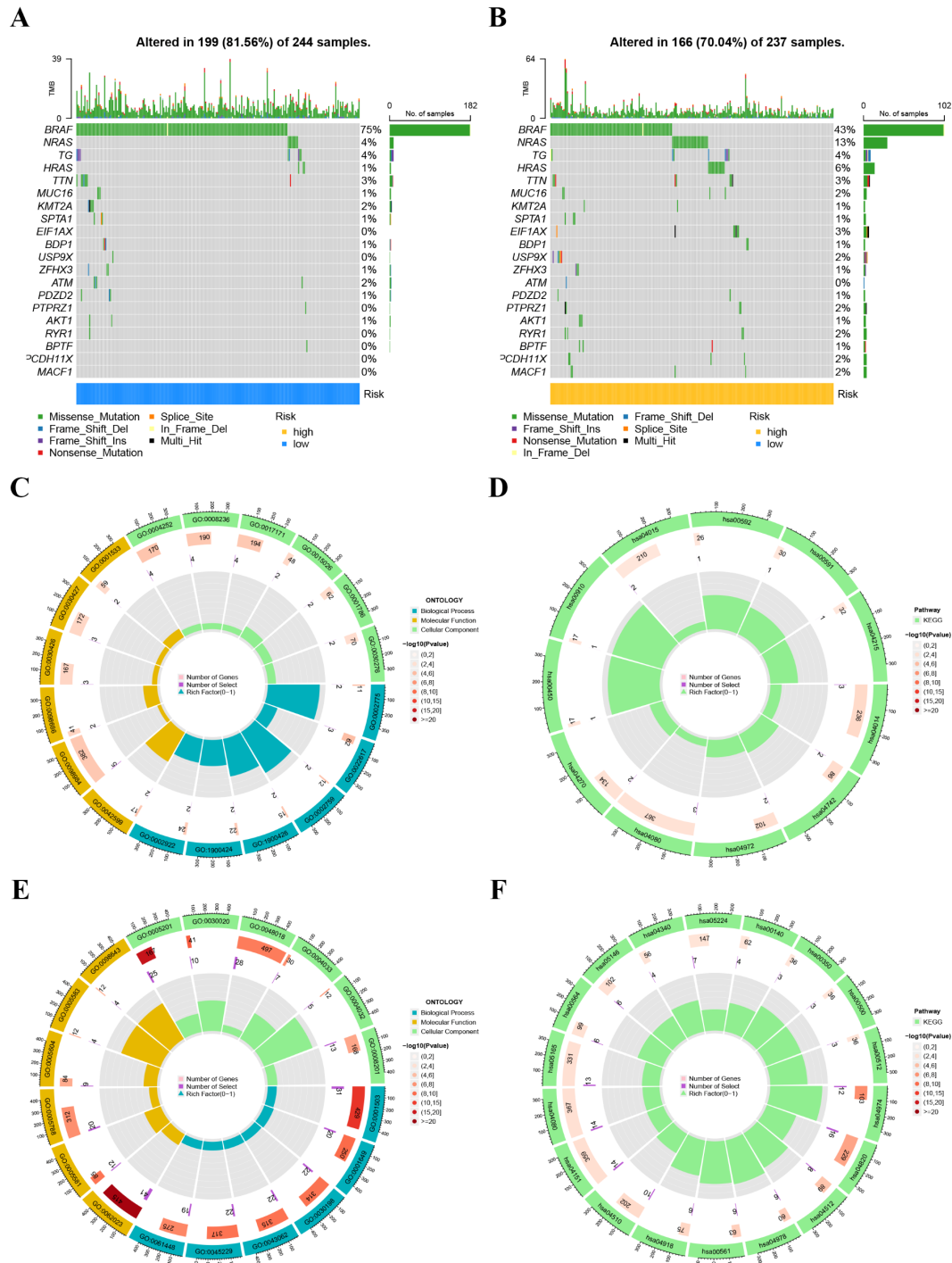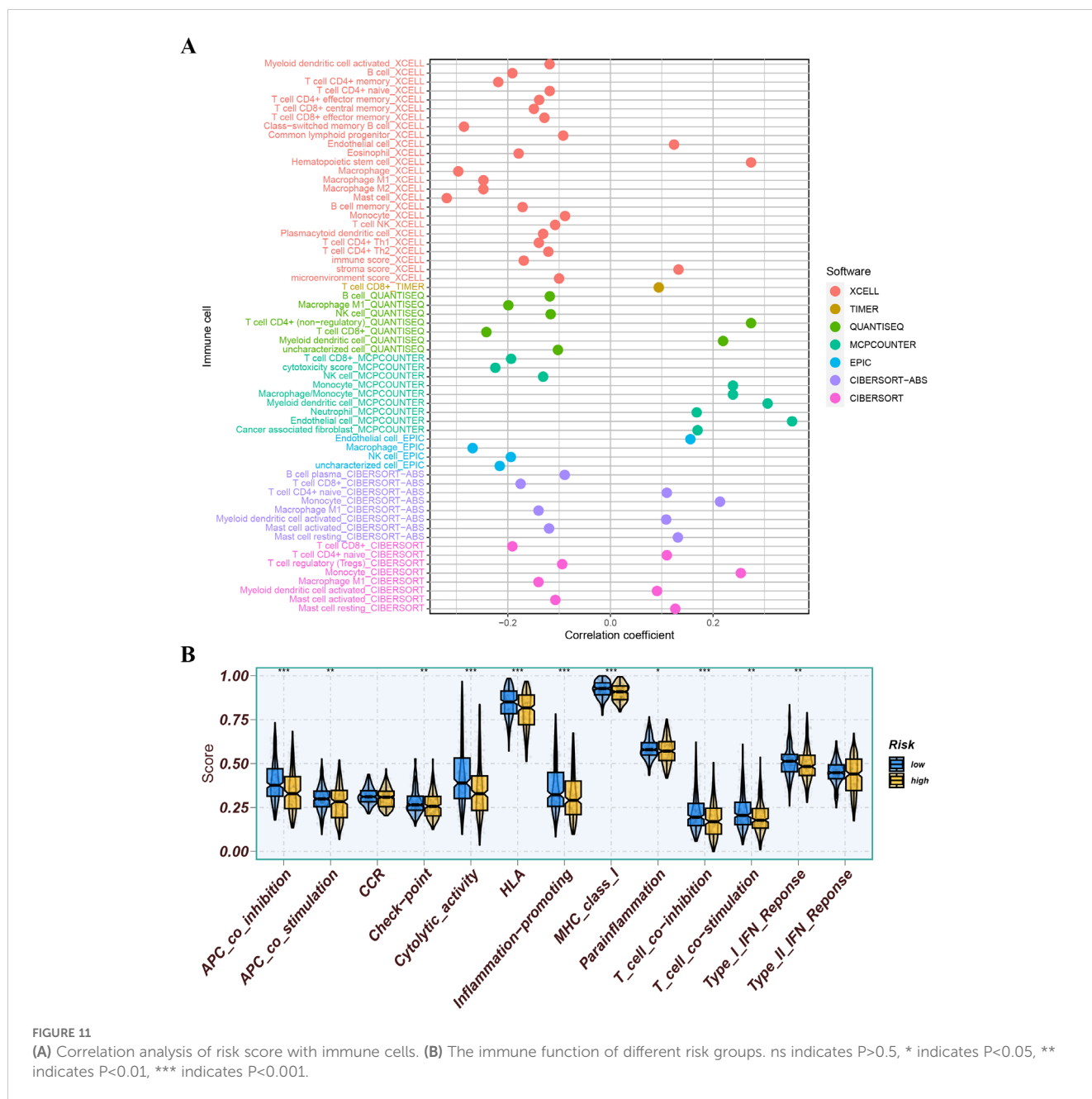
FIGURE 10
(A, B) Tumor mutation profile in the high-risk **(B)** and low-risk **(A)** groups. **(C, D)** Gene Ontology (GO) **(C)** and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways **(D)** enrichment analysis in the low-risk group. **(E, F)** Gene Ontology (GO) **(E)** and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways **(F)** enrichment analysis in the high-risk group.

Validation of the genes expression

The expression levels of 10 SHMRGs were detected in human PTC cell lines (IHH4, KTC-1, TPC-1) and human normal thyroid cell line (Nthy ori-3-1) using qPCR technology (Figure 14).

Discussion

THCA accounts for 3% of all cancers. While it is mostly regarded as an indolent disease, its recurrence rate remains at 10%-15% (4). Generally, cancer is associated with an imbalance in the hemostatic

FIGURE 11
(A) Correlation analysis of risk score with immune cells. **(B)** The immune function of different risk groups. ns indicates P>0.5, * indicates P<0.05, ** indicates P<0.01, *** indicates P<0.001.

system, but the pathogenesis of cancer-related coagulation disorders is complex (30). However, there is limited research on the relationship between THCA and coagulation. This study aims to clarify the correlation between coagulation and prognosis in THCA patients.

Several studies have indicated that PTC patients with LLNM have a higher risk of recurrence compared to those without LLNM (31, 32). Generally, prophylactic central lymph node dissection is deemed necessary. However, prophylactic dissection of negative lateral lymph nodes is not recommended because it may lead to numerous surgical complications. Therefore, it is important to determine the presence of LLNM preoperatively. This study analyzed the influence of coagulation markers (PT, PTA, INR, APTT, fibrinogen, TT, and D-dimer) on LLNM. The ROC curve showed that D-dimer has a good predictive performance for LLNM. Multivariate analysis indicated that

extrathyroidal extension, multifocality, and D-dimer >0.065mg/l are independent predictors of LLNM. The effects of extrathyroidal extension and multifocality on LLNM have been confirmed in numerous studies (33, 34). The research has shown that elevated D-dimer is significantly associated with cancer recurrence, metastasis, and worse survival outcomes (35). This study suggests that D-dimer >0.065mg/l is significantly associated with LLNM in PTC.

In this study, we divided THCA patients into two distinct subtypes based on the expression of prognostically valuable CRGs. K-M survival analysis indicated that cluster1 had better outcomes than cluster2 (p=0.014), offering a reference for refining the classification and management of THCA patients in clinical practice. Subsequently, we used LASSO regression to screen 8 genes (AZU1, COL3A1, CP, CSF2, F12, GNA14, IL1RN, and SERPIND1) to construct the
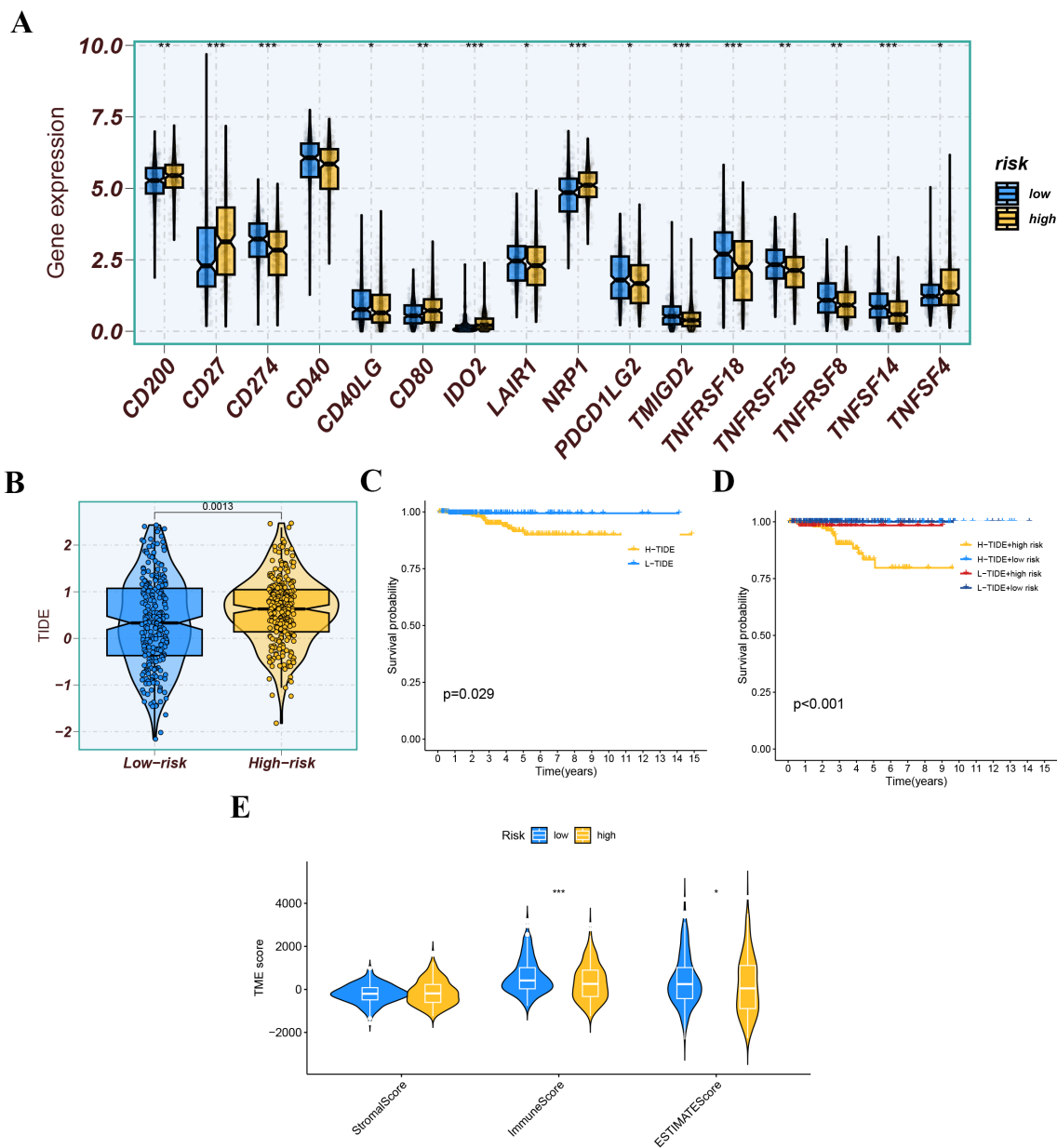
FIGURE 12
(A) Expression levels of immune checkpoint molecules between different risk groups. **(B)** The tumor immune dysfunction and exclusion (TIDE) score for each group. **(C, D)** Kaplan-Meier survival analysis based on the TIDE score and risk score. **(E)** TME score (StromalScore, ImmuneScore, and EstimateScore) across different risk groups. ns indicates P>0.5, * indicates P<0.05, ** indicates P<0.01, *** indicates P<0.001.

prognostic model. ROC results from both the training and validation sets suggested that the constructed model had high predictive efficacy, and the high-risk group had poorer outcomes. We further used univariate and multivariate Cox regression analyses to validate the effectiveness of the prognostic model. We also developed a nomogram integrating risk scores and clinical features to facilitate the clinical application of our research findings.

The tumor immune microenvironment (TME) plays an important role in tumor development. Recently, a pan-cancer analysis of the human tumor coagulome revealed that coagulation links to the TME (36). In our study results, the high-risk group had lower infiltration levels of macrophage M1 and M2 subtypes, CD8+ T cells, and NK cells, which may be related to poorer prognosis. Macrophages may combat cancer progression by directly engulfing cancer cells or activating antitumor immune responses, and CD8+ T cells and NK cells also have strong antitumor functions (37). To explore the biological functional differences between high and low-risk groups, we conducted GO and KEGG analyses. The results indicated significant differences in biological functions between the two groups, but the underlying mechanisms remain to be clarified. Finally, we investigated the intrinsic connection between drug sensitivity and risk groups by correlating THCA patients' prognostic risk scores with the half-maximal inhibitory concentration (IC50) values of chemotherapeutic drugs.
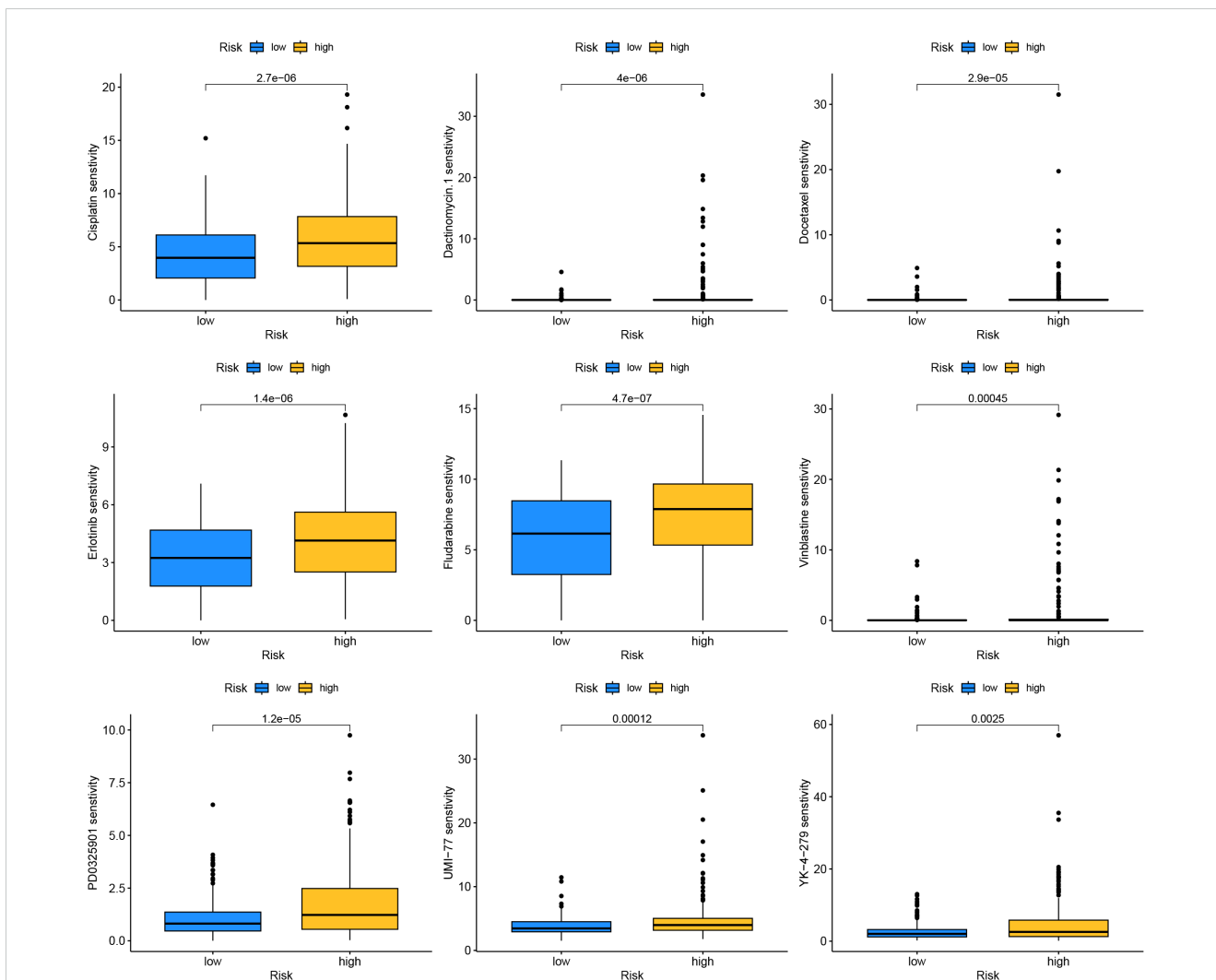
FIGURE 13

Drug sensitivity analysis for the different risk groups.
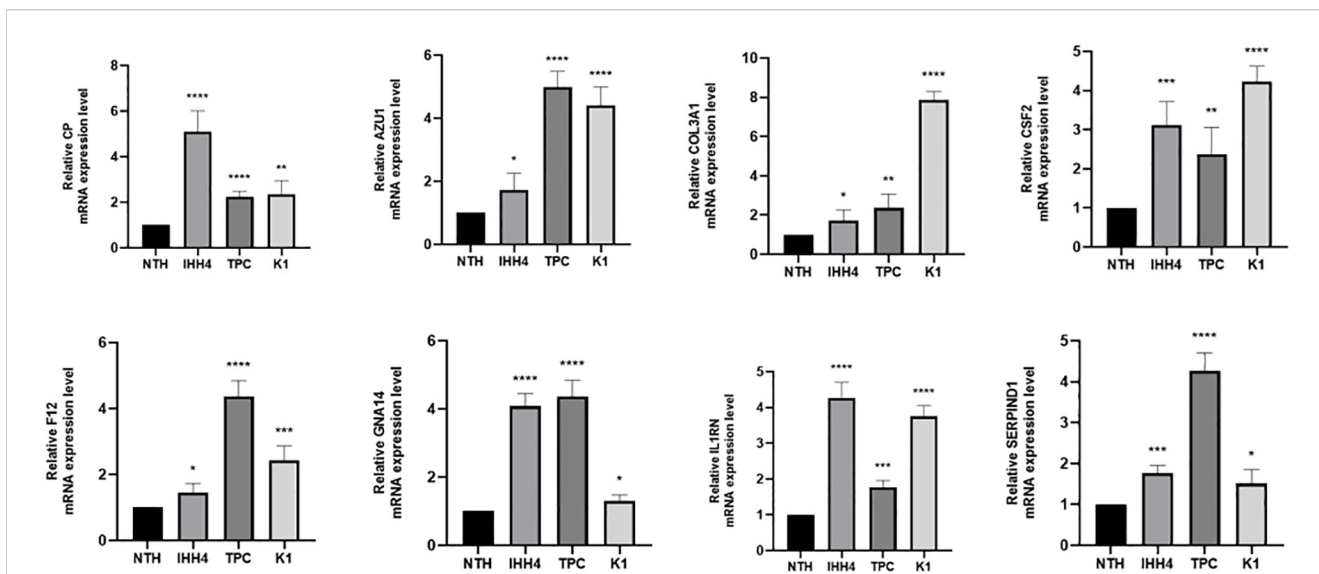

FIGURE 14

Expression of 8 genes in IHH 4, KTC-1, TPC-1 human papillary thyroid carcinoma cell line and human normal thyroid cell line Nthy ori-3-1. ns indicates P>0.5, * indicates P<0.05, ** indicates P<0.01, *** indicates P<0.001.

Research has indicated that these genes are involved in cancer development. In gastric cancer, AZU1 was upregulated (38). COL3A1 could be an oncogene and promote drug resistance in lung cancer (39). The product expressed by CP, cancer procoagulant, is a biomarker for cancer. Elevated CP activity has been detected in pancreatic, breast, lung, digestive system, and urinary system cancers (40). In breast cancer, CSF2 could activate the Stat3 pathway in CAA via paracrine or autocrine mechanisms, leading to increased expression and secretion of CXCL3. CXCL3 binds to CXCR2, a receptor on breast cancer cells, and activates FAK, which promotes a mesenchymal phenotype, invasion, and metastasis of breast cancer cells (41) F12 expression might affect the OS of PTC patients by regulating metabolic pathways (42). GNA14 might accelerate colorectal cancer cell proliferation and malignant tumor progression through ERK and β-catenin pathways (43). IL1RN was a good prognostic and diagnostic biomarker for PTC and might promote thyroid cancer progression through immune-related pathways (44). SERPIND1 promoted the proliferation, migration, invasion, G1-to-S phase transition, and epithelial-mesenchymal transition of ovarian cancer cells and inhibited their apoptosis by promoting phosphorylation in the phosphoinositide 3-kinase/protein kinase B (PI3K/AKT) pathway (36). We also validated the high expression of these 8 genes in PTC by qRT-PCR.

There are some limitations in this study. Firstly, the TCGA-THCA cohort has a limited number of cases, and more datasets are needed to validate these findings. Otherwise, the selected genes were only validated by qRT-PCR in PTC, lacking comprehensive experimental validation. Moreover, although we found that the coagulation pathway affects the prognosis and immune microenvironment of THCA patients, its underlying mechanisms need further investigation.

Conclusion

Our study demonstrated that coagulation is related to immune infiltration and prognosis in THCA. Our study suggested the possibility of D-dimer predicting LLNM. Subsequently, we used the TCGA cohort to construct a new coagulation-related THCA risk scoring model for prognosis assessment and risk stratification in THCA patients. This risk model could provide a robust prognostic tool and promote clinical guidance for THCA patients. We also constructed a nomogram combining the model and clinical features to predict the 3-, 5-, and 7-year survival probabilities of patients. We analyzed and compared high- and low-risk groups regarding immune infiltration, somatic mutations, and other aspects, offering some guidance for treatment strategy selection. In summary, our research will aid in understanding the role and significance of the coagulation in THCA.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Materials. Further inquiries can be directed to the corresponding author/s.

Ethics statement

The studies involving humans were approved by The First Affiliated Hospital of Zhengzhou University Ethics Review Committee. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

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.

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/fimmu.2024.1462755/full#supplementary-material

References

1. Vaccarella S, Dal Maso L, Laversanne M, Bray F, Plummer M, Franceschi S. The impact of diagnostic changes on the rise in thyroid cancer incidence: A population-based study in selected high-resource countries. *Thyroid*. (2015) 25:1127–36. doi: 10.1089/thy.2015.0116

2. Xia C, Dong X, Li H, Cao M, Sun D, He S, et al. Cancer statistics in China and United States, 2022: profiles, trends, and determinants. *Chin Med J (Engl)*. (2022) 135:584–90. doi: 10.1097/cm9.0000000000002108

3. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the american thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid*. (2016) 26:1–133. doi: 10.1089/thy.2015.0020

4. Laha D, Nilubol N, Boufraqech M. New therapies for advanced thyroid cancer. *Front Endocrinol (Lausanne)*. (2020) 11:82. doi: 10.3389/fendo.2020.00082

5. Furie B, Furie BC. Mechanisms of thrombus formation. *N Engl J Med*. (2008) 359:938–49. doi: 10.1056/NEJMra0801082

6. Kocatürk B, Versteeg HH. Tissue factor isoforms in cancer and coagulation: may the best isoform win. *Thromb Res*. (2012) 129 Suppl 1:S69–75. doi: 10.1016/s0049-3848(12)70020-8

7. Yang J, Wang C, Zhang Y, Cheng S, Wu M, Gu S, et al. Clinical significance and immune infiltration analyses of a novel coagulation-related signature in ovarian cancer. *Cancer Cell Int*. (2023) 23:232. doi: 10.1186/s12935-023-03040-3

8. Bano A, Chaker L, de Maat MPM, Atiq F, Kavousi M, Franco OH, et al. Thyroid function and cardiovascular disease: the mediating role of coagulation factors. *J Clin Endocrinol Metab*. (2019) 104:3203–12. doi: 10.1210/jc.2019-00072

9. Vanderpump MP, Tunbridge WM, French JM, Appleton D, Bates D, Clark F, et al. The incidence of thyroid disorders in the community: a twenty-year follow-up of the Whickham Survey. *Clin Endocrinol (Oxf)*. (1995) 43:55–68. doi: 10.1111/j.1365-2265.1995.tb01894.x

10. Sun Y, Liu Y, Li H, Tang Y, Liu W, Zhang Y, et al. The significance and prognostic value of multifocal papillary thyroid carcinoma in children and adolescents. *BMC Cancer*. (2024) 24:690. doi: 10.1186/s12885-024-12403-6

11. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The geneCards suite: from gene data mining to disease genome sequence analyses. *Curr Protoc Bioinf*. (2016) 54:1. doi: 10.1002/cpbi.5

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

13. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med*. (2013) 32:5381–97. doi: 10.1002/sim.5958

14. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. (2018) 28:1747–56. doi: 10.1101/gr.239244.118

15. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. (2015) 43:e47. doi: 10.1093/nar/gkv007

16. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)*. (2021) 2:100141. doi: 10.1016/j.xinn.2021.100141

17. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular visualization in R. *Bioinformatics*. (2014) 30:2811–2. doi: 10.1093/bioinformatics/btu393

18. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. (2016) 32:2847–9. doi: 10.1093/bioinformatics/btw313

19. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. (2015) 12:453–7. doi: 10.1038/nmeth.3337

20. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res*. (2017) 77:e108–10. doi: 10.1158/0008-5472.Can-17-0307

21. Dienstmann R, Villacampa G, Sveen A, Mason MJ, Niedzwiecki D, Nesbakken A, et al. Relative contribution of clinicopathological variables, genomic markers, transcriptomic subtyping and microenvironment features for outcome prediction in stage II/III colorectal cancer. *Ann Oncol*. (2019) 30:1622–9. doi: 10.1093/annonc/mdz287

22. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol*. (2017) 18:220. doi: 10.1186/s13059-017-1349-1

23. Finotello F, Mayer C, Plattner C, Laschober G, Rieder D, Hackl H, et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med*. (2019) 11:34. doi: 10.1186/s13073-019-0638-6

24. Schenk E, Boland J, Mansfield A, Aubry MC, Dietz A. Local and systemic immunity predict survival in patients with pulmonary sarcomatoid carcinoma. *Med Oncol*. (2017) 34:140. doi: 10.1007/s12032-017-1000-8

25. Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife*. (2017) 6:1–25. doi: 10.7554/eLife.26476

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

27. Zhang Z. Reshaping and aggregating data: an introduction to reshape package. *Ann Transl Med*. (2016) 4:78. doi: 10.3978/j.issn.2305-5839.2016.01.33

28. Fu J, Li K, Zhang W, Wan C, Zhang J, Jiang P, et al. Large-scale public data reuse to model immunotherapy response and resistance. *Genome Med*. (2020) 12:21. doi: 10.1186/s13073-020-0721-z

29. Maeser D, Gruener RF, Huang RS. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Brief Bioinform*. (2021) 22:1943–54. doi: 10.1093/bib/bbab260

30. Ordookhani A, Motazedi A, Burman KD. Thrombosis in thyroid cancer. *Int J Endocrinol Metab*. (2018) 16:e57897. doi: 10.5812/ijem.57897

31. Kim SK, Park I, Hur N, Rayzah M, Lee JH, Choe JH, et al. Patterns, predictive factors, and prognostic impact of contralateral lateral lymph node metastasis in N1b papillary thyroid carcinoma. *Ann Surg Oncol*. (2017) 24:1943–50. doi: 10.1245/s10434-016-5761-7

32. Chéreau N, Buffet C, Trésallet C, Tissier F, Leenhardt L, Menegaux F. Recurrence of papillary thyroid carcinoma with lateral cervical node metastases: Predictive factors and operative management. *Surgery*. (2016) 159:755–62. doi: 10.1016/j.surg.2015.08.033

33. So YK, Kim MJ, Kim S, Son YI. Lateral lymph node metastasis in papillary thyroid carcinoma: A systematic review and meta-analysis for prevalence, risk factors, and location. *Int J Surg*. (2018) 50:94–103. doi: 10.1016/j.ijsu.2017.12.029

34. Kim HJ, Sohn SY, Jang HW, Kim SW, Chung JH. Multifocality, but not bilaterality, is a predictor of disease recurrence/persistence of papillary thyroid carcinoma. *World J Surg*. (2013) 37:376–84. doi: 10.1007/s00268-012-1835-2

35. Kong W, Zhang L, An R, Yang M, Wang H. Diagnostic value of serum D-dimer for detection of gallbladder carcinoma. *Cancer Manag Res*. (2021) 13:2549–56. doi: 10.2147/cmar.S272116

36. Saidak Z, Soudet S, Lottin M, Salle V, Sevestre MA, Clatot F, et al. A pan-cancer analysis of the human tumor coagulome and its link to the tumor immune microenvironment. *Cancer Immunol Immunother*. (2021) 70:923–33. doi: 10.1007/s00262-020-02739-w

37. de Visser KE, Joyce JA. The evolving tumor microenvironment: From cancer initiation to metastatic outgrowth. *Cancer Cell*. (2023) 41:374–403. doi: 10.1016/j.ccell.2023.02.016

38. Ran X, Xu X, Yang Y, She S, Yang M, Li S, et al. A quantitative proteomics study on olfactomedin 4 in the development of gastric cancer. *Int J Oncol*. (2015) 47:1932–44. doi: 10.3892/ijo.2015.3168

39. Wang L, Sun Y, Guo Z, Liu H. COL3A1 overexpression associates with poor prognosis and cisplatin resistance in lung cancer. *Balkan Med J*. (2022) 39:393–400. doi: 10.4274/balkanmedj.galenos.2022.2022-6-16

40. Szajda SD, Darewicz B, Kudelski J, Werel T, Zwierz K, Skrzydlewski Z, et al. Assessment of the efficacy of cancer procoagulant (CP) activity evaluation in the oncological diagnosis of the urinary system. *Pol Merkur Lekarski*. (2005) 19:596–9.

41. He X, Wang L, Li H, Liu Y, Tong C, Xie C, et al. CSF2 upregulates CXCL3 expression in adipocytes to promote metastasis of breast cancer via the FAK signaling pathway. *J Mol Cell Biol*. (2023) 15:1–14. doi: 10.1093/jmcb/mjad025

42. Luo JH, Zhang XX, Sun WH. F12 as a reliable diagnostic and prognostic biomarker associated with immune infiltration in papillary thyroid cancer. *Aging (Albany NY)*. (2022) 14:3687–704. doi: 10.18632/aging.204037

43. Park R, Lee S, Chin H, Nguyen AT, Lee D. Tumor-promoting role of GNA14 in colon cancer development. *Cancers (Basel)*. (2023) 15(18). doi: 10.3390/cancers15184572

44. Xie Z, Li X, He Y, Wu S, Wang S, Sun J, et al. Analysis of the expression and potential molecular mechanism of interleukin-1 receptor antagonist (IL1RN) in papillary thyroid cancer via bioinformatics methods. *BMC Cancer*. (2020) 20:1143. doi: 10.1186/s12885-020-07620-8