

RESEARCH ARTICLE

A novel prognostic model for papillary thyroid cancer based on epithelial–mesenchymal transition-related genes

Rui Liu¹ | Zhen Cao¹ | Meng Pan² | Mengwei Wu¹ | Xiaobin Li¹ |
Hongwei Yuan¹ | Ziwen Liu¹ 

¹Department of General Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, People's Republic of China

²State Key Laboratory of Medical Molecular Biology & Department of Immunology, Institute of Basic Medical Sciences Chinese Academy of Medical Sciences, School of Basic Medicine Peking Union Medical College, Beijing, People's Republic of China

Correspondence

Ziwen Liu, Department of General Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, People's Republic of China.

Email: liuziwen@pumch.cn

Funding information

CAMS Innovation Fund for graduate students, Grant/Award Number: 2019-1002-44; CAMS Innovation Fund for Medical Sciences (CIFMS), Grant/Award Number: 2016-12M-3-005; National Natural Science Foundation of China, Grant/Award Number: 82172727; the Nature Science Foundation of Beijing, Grant/Award Number: 7202164

Abstract

Background: The frequent incidence of postsurgical recurrence issues in papillary thyroid cancer (PTC) patients is a primary concern considering the low cancer-related mortality. Previous studies have demonstrated that epithelial–mesenchymal transition (EMT) activation is closely related to PTC progression and invasion. In this study, we aimed to develop a novel EMT signature and ancillary nomogram to improve personalized prediction of progression-free interval (PFI).

Methods: First, we carried out a differential analysis of PTC samples and pairwise normal thyroid samples to explore the differentially expressed genes (DEGs). The intersection of the DEGs with EMT-related genes (ERGs) were identified as differentially expressed EMT-related genes (DE-ERGs). We determined PFI-related DE-ERGs by Cox regression analysis and then established a novel gene classifier by LASSO regression analysis. We validated the signature in external datasets and in multiple cell lines. Further, we used uni- and multivariate analyses to identify independent prognostic characters.

Results: We identified 244 prognosis-related DE-ERGs. The 244 DE-ERGs were associated with several pivotal oncogenic processes. We also constructed a novel 10-gene signature and relevant prognostic model for recurrence prediction of PTC. The 10-gene signature had a C-index of 0.723 and the relevant nomogram had a C-index of 0.776. The efficacy of the signature and nomogram was satisfying and closely correlated with relevant clinical parameters. Furthermore, the signature also had a unique potential in differentiating anaplastic thyroid cancer (ATC) samples.

Conclusions: The novel EMT signature and nomogram are useful and convenient for personalized management for thyroid cancer.

KEYWORDS

bioinformatics, epithelial–mesenchymal transition, nomogram, papillary thyroid cancer, predictive model, recurrence, The Cancer Genome Atlas

Rui Liu and Zhen Cao are contributed equally to this work.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Cancer Medicine* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

With a 3% annual increase of the incidence of the past recent decades in the United States, thyroid cancer (TC) has become the most common endocrine malignant tumor and a leading public health issue.¹ Although most cases (>90%) are papillary thyroid cancer (PTC) with a good prognosis, one of the main obstacles to the treatment is the continuing high rate of recurrence, distant metastasis, and even persistent disease.^{2,3} Compared with cancer-related morbidity of 13.5/100,000 in 2015, approximately 30% of PTC patients would suffer from postoperative recurrence, with the consequent requirement for reoperation.⁴ Postoperative recurrence not only results in a higher rate of surgical complications such as vocal cord paralysis or hypocalcemia but also with dedifferentiation events that coexist in metastatic lesions.⁵ In other words, anaplastic thyroid cancer (ATC), which accounts for 90% of TC-related deaths, may occur in recurrent or metastatic sites.⁶ Hence, individualized therapy and surveillance based on precise prediction of the risk of recurrence remains a key focus of clinical research on PTC.⁷

Epithelial-mesenchymal transition (EMT) is a morphological alteration in which cells develop a fibroblast-like phenotype following the initiation of mesenchymal cell-related transcriptive events.⁸ In the process of EMT, epithelial cells dedifferentiate with a reduction in the expression of adhesion molecules (e.g., E-cadherin), resulting in loss of the contact inhibition mediated by the tissue structure. Cells also undergo a change in morphology from a pebble to spindle shape caused by rearrangement of the cytoskeleton, loss of connection to the basement membrane, and acquisition of mesenchymal phenotypes.⁹ An increasing number of studies have demonstrated that EMT activation is closely related to multiple milestones in the development of cancer, including distant metastasis, extrathyroidal extension, and increased cancer cell stemness, and can be observed at the tumor-stromal boundary of TC.^{10,11} Due to the complicated process of EMT, which involves various transcriptional and translational factors associated with TC, ancillary diagnostic tools based on standardized genomic characteristics would offer better predictive value than risk stratification based on clinical parameters alone.^{12,13} Based on the pivotal role of EMT in PTC invasion and anaplastic transformation, we established a novel 10-gene signature derived from prognosis-related differentially expressed EMT-related genes (DE-ERGs) using Cox proportional hazard analysis. Clinical validation on external datasets and experimental validation on PTC cell lines were also conducted. Thereafter, we used uni- and multivariate analyses to identify independent prognostic characters. We then constructed a stepwise regression model

and a nomogram based on the risk score and other crucial clinical parameters to predict the prognosis of PTC patients. Subsequently, we verified the prognostic value of the novel ancillary model.

2 | MATERIALS AND METHODS

2.1 | Acquisition of TCGA-THCA transcriptional data

We downloaded level-3 RNA sequencing data in the form of fragments per kilobase of exon model per million reads mapped (FPKM) and counts from The Cancer Genome Atlas (TCGA) updated in July 2019.¹⁴ The TCGA, a landmark cancer genomics program, molecularly characterized over 20,000 primary cancer and matched normal samples spanning 33 cancer types. We applied a transcript per million (TPM) transformation and normalized the sequencing data with a base-2 logarithm. Cases with a follow-up period of less than a month were excluded. We obtained clinical data from the University of California, Santa Cruz (UCSC) Xena database, which provides interactive online visualization of seminal cancer genomics datasets, including data from TCGA and gene mutation status from the cBioPortal database.¹⁵

2.2 | Differential classification of DE-ERGs

We performed differential gene expression analysis of all the 502 PTC cases with 58 normal thyroid tissues from TCGA-THCA dataset using the R package “EdgeR.” We identified DEGs according to the criteria of false discovery rate (FDR) <0.05 and $|\text{Log}_2\text{FC}| > 1$.¹⁶ A list of EMT-related genes (ERGs) was obtained from the Epithelial-Mesenchymal Transition Gene Database (dbEMT2, <http://dbemt.bioinfo-minzhao.org/>).¹⁷ The DE-ERGs were defined as the DEGs that intersected with the ERGs.

2.3 | Functional annotation

The “clusterProfiler” package was used to annotate the enriched biological functions of the previously defined DE-ERGs.¹⁸ Gene ontology (GO) and Kyoto Genes and Genomes (KEGG) pathway analyses were applied, and the terms biological process (BP), cellular component (CC), and molecular function (MF) were explored. The Benjamini-Hochberg method was used for false FDR correction.

2.4 | Screening and assessment of the novel EMT signature

After differential gene expression analysis, all the 488 PTC cases from TCGA thyroid cancer (THCA) dataset were allocated randomly into the training and the testing datasets on a ratio of 0.8.¹⁹ The baseline characteristics of the 488 PTC patients in TCGA-THCA dataset are presented in Table 1. We then used a univariate Cox regression model to screen the genes that were significantly related to progression-free interval (PFI) ($P < 0.05$) in the training dataset from the DE-ERGs. Further, crucial genes and relevant coefficients were selected from the PFI-related DE-ERGs by Least absolute shrinkage and selection operator (LASSO) regression analysis with the “glmnet” R package.²⁰ To assess the efficacy of stratification of different risks of recurrence, we defined the PTC cases as “low risk” or “high risk” of recurrence according to the optimum cutoff defined by X-Tile.²¹ Finally, we evaluated of the predictive power of the novel EMT signature by calculation of the area under the curve (AUC) of the receiver operating characteristic (ROC) curve in training, validation, and total set,²² and Kaplan–Meier (K–M) analysis and determination of Harrell’s C-index using the “survcomp” and “survival” packages of R.²³ Next, we analyzed the correlation between the EMT signature and clinical characters including age, gender, BRAF mutation status, TNM stage, extrathyroidal extension, residual tumor and primary focality in TCGA-THCA cohort. Since the prognosis of PTC/ poorly differentiated thyroid cancer (PDTC)/ATC varies from relatively well to very poor, we were interested in the role of EMT signature in TC samples with different degree of differentiation. To partially address this issue, we implemented three datasets which contains PTC, PDTC, and ATC samples from GSE29265, GSE33630,²⁴ and GSE76039²⁵ to calculate risk score with EMT signature and immune signature previously reported by Lin et al.²⁶ In order to prove the predictive power of the signature, we compared with the predictive power of other clinical information including American Thyroid Association (ATA) stratification, American Joint Committee on Cancer (AJCC) stage, and (Metastasis, Age, Completeness of resection, Invasion, and Size) (MACIS) score. To validate the clinical value of the EMT signature, we also implemented GSE138042,²⁷ GSE82208, GSE58545,²⁸ and GSE60542²⁹ including TC samples of different stage and tissue types. All the GSE datasets were obtained from Gene Expression Omnibus (GEO); the information is listed in Table 3.

2.5 | Gene set enrichment analysis (GSEA) of the ERG-based classifier

We explored the potential turbulence in the ERG-based signature at the transcriptional level by GSEA.³⁰

TABLE 1 Baseline of enrolled 488 PTC patients in the TCGA-THCA dataset

Characteristic	Training	Testing	<i>p</i>
<i>n</i>	380	108	
Progression, <i>n</i> (%)			0.901
Free	341 (69.9%)	98 (20.1%)	
Progression	39 (8%)	10 (2%)	
RAS_status, <i>n</i> (%)			0.602
Mutated	48 (9.8%)	11 (2.3%)	
Wild type	332 (68%)	97 (19.9%)	
BRAF_status, <i>n</i> (%)			0.482
Mutated	212 (43.4%)	65 (13.3%)	
Wild type	168 (34.4%)	43 (8.8%)	
Extrathyroid_extension, <i>n</i> (%)			0.756
Minimal (T3)	102 (21.7%)	28 (5.9%)	
Moderate/Advanced (T4)	13 (2.8%)	5 (1.1%)	
None	255 (54.1%)	68 (14.4%)	
Histological_type, <i>n</i> (%)			0.348
Classical/usual	268 (54.9%)	83 (17%)	
Follicular	84 (17.2%)	17 (3.5%)	
Tall Cell	28 (5.7%)	8 (1.6%)	
Neoplasm_focus_type, <i>n</i> (%)			0.101
Multifocal	178 (37.2%)	40 (8.4%)	
Unifocal	195 (40.8%)	65 (13.6%)	
Anatomic_site, <i>n</i> (%)			0.113
Bilateral	69 (14.3%)	12 (2.5%)	
Isthmus	19 (3.9%)	3 (0.6%)	
Unilateral	286 (59.3%)	93 (19.3%)	
Residual_tumor, <i>n</i> (%)			0.677
R0	291 (68.3%)	80 (18.8%)	
R1	38 (8.9%)	13 (3.1%)	
R2	3 (0.7%)	1 (0.2%)	
Ajcc_stage, <i>n</i> (%)			0.716
Stage I	215 (44.2%)	58 (11.9%)	
Stage II	37 (7.6%)	14 (2.9%)	
Stage III	84 (17.3%)	26 (5.3%)	
Stage IV	42 (8.6%)	10 (2.1%)	
M_stage, <i>n</i> (%)			0.691
M0	372 (76.4%)	107 (22%)	
M1	7 (1.4%)	1 (0.2%)	
N_stage, <i>n</i> (%)			0.378
N0	180 (41.1%)	45 (10.3%)	
N1	162 (37%)	51 (11.6%)	

(Continues)

TABLE 1 (Continued)

Characteristic	Training	Testing	<i>p</i>
T_stage, <i>n</i> (%)			0.960
T1	111 (22.8%)	30 (6.2%)	
T2	126 (25.9%)	35 (7.2%)	
T3	125 (25.7%)	38 (7.8%)	
T4	16 (3.3%)	5 (1%)	
Gender, <i>n</i> (%)			0.097
Female	286 (58.6%)	72 (14.8%)	
Male	94 (19.3%)	36 (7.4%)	
Age, <i>n</i> (%)			1.000
<55	253 (51.8%)	72 (14.8%)	
≥55	127 (26%)	36 (7.4%)	
Progression_free_interval, median (IQR)	888.5 (496.5, 1463.75)	927 (460, 1340.75)	0.671

All the 488 PTC cases from the THCA set were defined as low- or high-risk according to the optimum cutoff defined by X-Tile. We then explored the upregulated pathways in high-risk group using GSEA V4.1 software. C5: GO and C6: oncogenic signatures were used as background.

2.6 | Cell culture and lysis

Normal human thyroid follicular cell line Nthy-ori 3.1,³¹ and PTC cell lines B-CPAP (primary and well differentiated)³² and KTC-1 (metastatic and refractory to radio iodine)³³ were kindly provided by the National Collection of Authenticated Cell Cultures of the Chinese Academy of Sciences. B-CPAP and KTC-1 were cultured under the circumstance of 5% CO₂, 37°C, in RPMI Medium 1640 (Invitrogen) with 10% FBS (Gibco), non-essential amino acids (Invitrogen), Glutamax (Invitrogen), and Sodium Pyruvate (Invitrogen, 11,360,070) added. TRIzol (Lablead) was used to lysate and isolate RNA from cells according to the manufacturer's protocol.

2.7 | Quantitative real-time Polymerase Chain Reaction (RT-qPCR)

RT-qPCR was conducted after isolation of RNA and performed essentially as described previously with house-keeper (GAPDH) mRNA for normalization via the 2^{-ΔΔCt} method.³⁴ The sequences of primers are listed in Table 4.

2.8 | Identification of prognosis-associated factors

Univariate and multivariate Cox analyses were implemented to extract the correlating prognostic parameters. EMT risk score and clinical information included age, gender, BRAF mutation status, RAS mutation status, Neoplasm_size, TNM stage, extrathyroidal extension, residual tumor, primary focality, and anatomic site. Categorical variables including gender, age (over 55 or not), BRAFV600E, RAS, TNM, extrathyroidal extension, residual tumor, primary focality, anatomic site, and histological type were transformed into rank variables. We firstly conducted univariate analysis and excluded factors with *p* > 0.1 for further multivariate analysis.

2.9 | Implementation and validation of the novel predictive model

We constructed a stepwise Cox model after collinearity tests for predicting the PTC's PFI. Clinical factors in multivariate analysis and other previously reported significantly correlated ones were included. Thereafter, the model was visualized as a nomogram. The cutoff value to define high- or low-risk groups was generated by X-Tile based on the nomogram points for each case. We then assessed the predictive ability of the model based on the AUC of the ROC curve, the C-index, and the calibration curve, which was generated with 1000 times bootstrap resampling.

2.10 | Statistical analysis

Statistical analyses were conducted using R version 4.0.3 (Bunny-Wunnies Freak Out) and GraphPad Prism 8.4.3 (GraphPad Software, CA, USA). Codes used for analyses are listed in supplementary document S1. Continuous data were analyzed using unpaired *t*-tests. Categorical variables were analyzed using chi-squared tests. K-M survival curves were analyzed with the Gehan-Breslow-Wilcoxon test. Multiple testing correction were applied with Benjamini and Hochberg method for difference and enrichment analyses. *p*-values < 0.05 were considered to indicate statistical significance.

3 | RESULTS

3.1 | Determination of DE-ERGs

The workflow of our study is shown in Figure 1. In total, we enrolled 488 PTC cases with complete follow-up

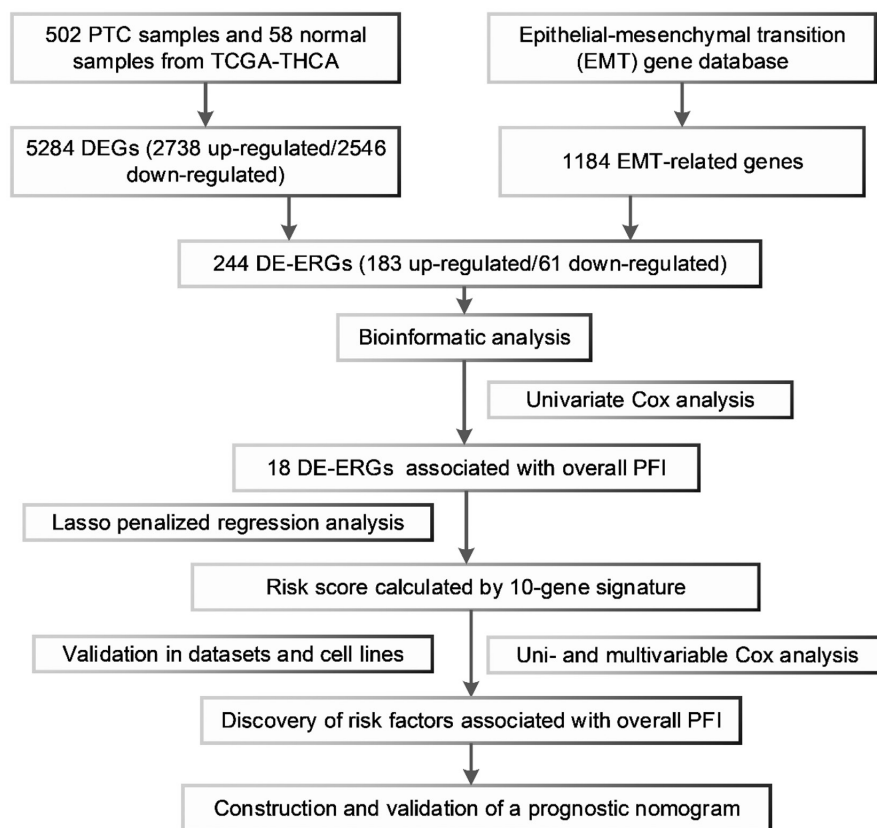


FIGURE 1 Flowchart showing the workflow of our analysis

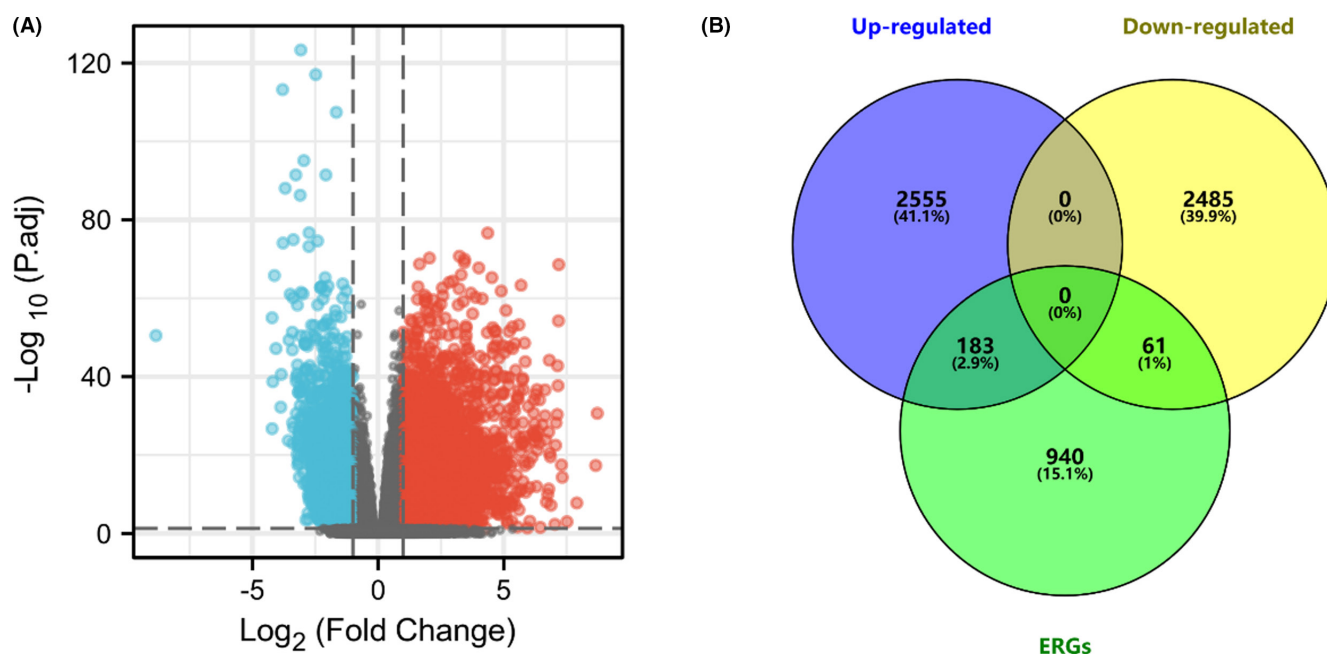


FIGURE 2 Differentially expressed gene (DEG) analysis and intersection showing the differentially expressed EMT-related genes (ERGs) (DE-ERGs). (A) Volcano plot showing the DEG analysis between the 502 PTC samples and 58 normal thyroid samples. (B) Venn diagram showing the intersection of 5284 DEGs with the list containing 1184 DE-ERGs derived from dbEMT 2.0.

information for our analysis. Through differential gene expression analysis of 502 PTC samples and 58 normal thyroid tissues from TCGA-THCA, we retrieved a total of 5284 DEGs (2738 up- and 2546 downregulated) (Figure 2A).

Furthermore, we downloaded the list of 1184 ERGs from dbEMT2. The intersection contained 244 DE-ERGs (183 up- and 61 downregulated) (Figure 2B). The full list of DEGs, ERGs, and 244 DE-ERGs is shown in Tables S1–S3.

3.2 | Functional annotation of the 244 DE-ERGs

Annotation of the 244 DE-ERGs by GO and KEGG pathway analyses is shown in Figure 3. For the biological process (BP) category, the DE-ERGs were mainly enriched in epithelial cell proliferation, extracellular structure organization, and regulation of epithelial cell proliferation (Figure 3A). In the terms of the cellular component (CC) category, the DE-ERGs were mainly enriched in collagen-containing extracellular matrix, cell-substrate junction, and focal adhesion (Figure 3B). In the terms of the molecular function (MF) category, the DE-ERGs were mainly enriched in receptor ligand activity, DNA-binding transcription activator activity, and polymerase II-specific (Figure 3C). In terms of KEGG pathways, the

DE-ERGs were mainly enriched in PI3K-Akt signaling pathway, microRNAs in cancer, and proteoglycans in cancer (Figure 3D).

3.3 | Screening of significant DE-ERGs and construction of the 10-gene signature

The baseline information of the training and testing sets allocated from the 488 PTC cases is shown in Table 1. A total of 18 PFI-related DE-ERGs were identified; full results are shown in Table S4. Forest plots of the logfc, *p*-value, and hazard ratio of each item are shown in Figure 4A. The potential protein-protein interaction network of the 18 items was explored using the STRING database.³⁵ In

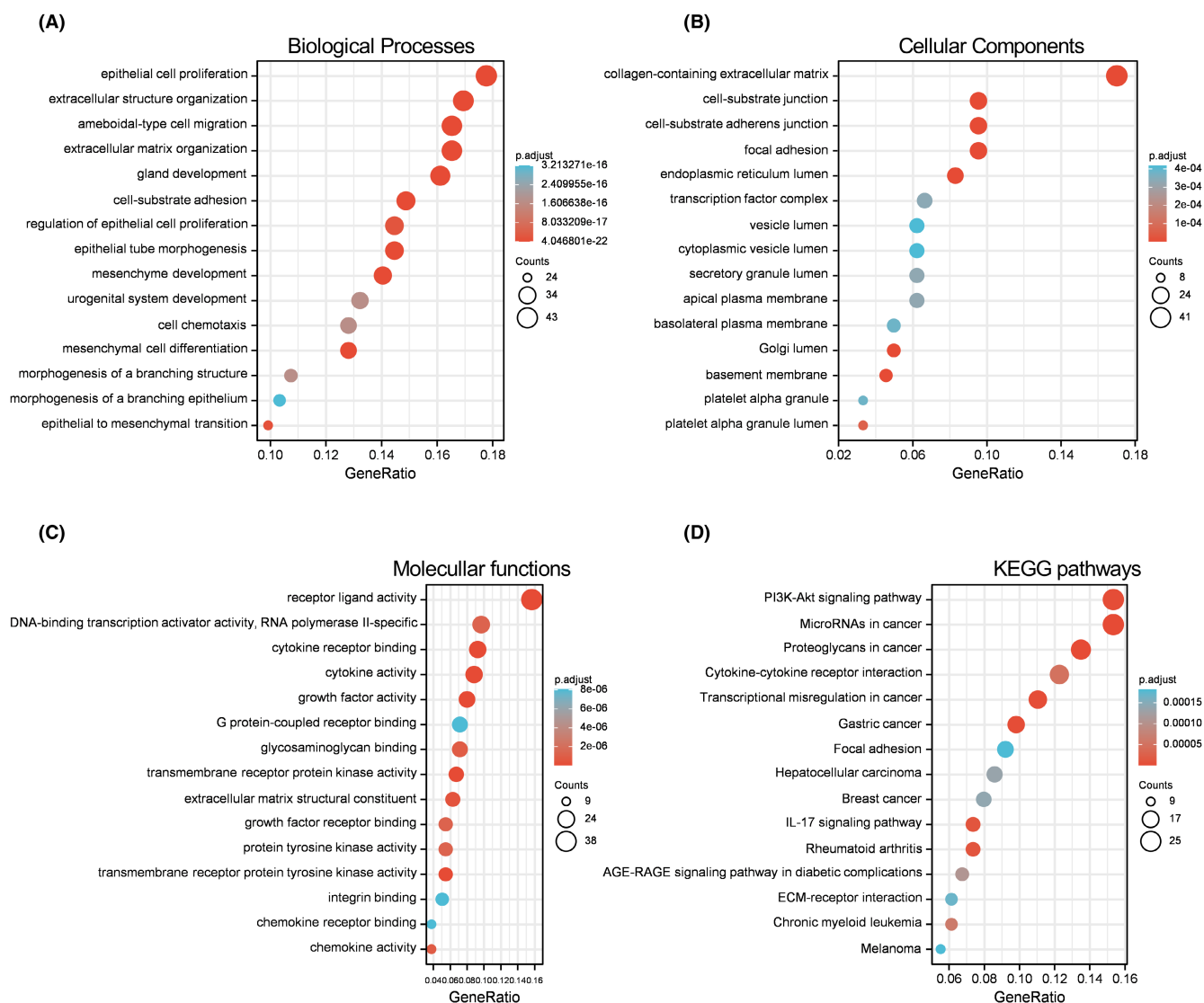


FIGURE 3 Functional annotations of the 244 DE-ERGs. Bar charts showing the annotation of 244 DE-ERGs by gene ontology (GO) analysis and KEGG pathway analysis. (A–C) The predominantly enriched terms of cellular component, biological process, and molecular function. (D). The predominantly enriched pathways of the 244 DE-ERGs.

the complex network, five proteins (EFEMP1, CTGF, EGR1, JUN, and FOXA2) were identified as key nodes (Figure 4B). LASSO penalty regression analysis reduced and constructed a novel 10-gene signature (Figure 4C,D). The risk score was calculated as follows: risk score = $\exp/\text{FHL1} \times (-0.04395) + \exp/\text{CTGF} \times (-0.13087) + \exp/\text{FOXP2} \times (-0.05892) + \exp/\text{FBLN5} \times (-0.19037) + \exp/\text{WT1} \times 0.560368 + \exp/\text{IL11} \times 0.448262 + \exp/\text{AQP9} \times 0.088904 + \exp/\text{TGFBR3} \times (-0.18178) + \exp/\text{WNT11} \times (-0.02575) + \exp/\text{UHRF1} \times 0.178616$. The coefficient of variation (CV) in training and testing sets was 26.57% and 26.95%.

3.4 | Verification of the discriminatory power of the 10-gene signature

According to K-M analysis, the 10-gene signature provided good discriminatory power for patients with high or low risk of recurrence (Figure 5A–C). PTC cases with higher risk score due to the optimum cutoff value of higher risk had a significantly poorer prognosis ($p < 0.0001$, Figure 5C). The relationships between risk score and recurrence events are shown in scatterplots (Figure 5D–F). In the training set, the AUCs for 1-year, 3-year, and 5-year PFI prediction based on the 10-gene

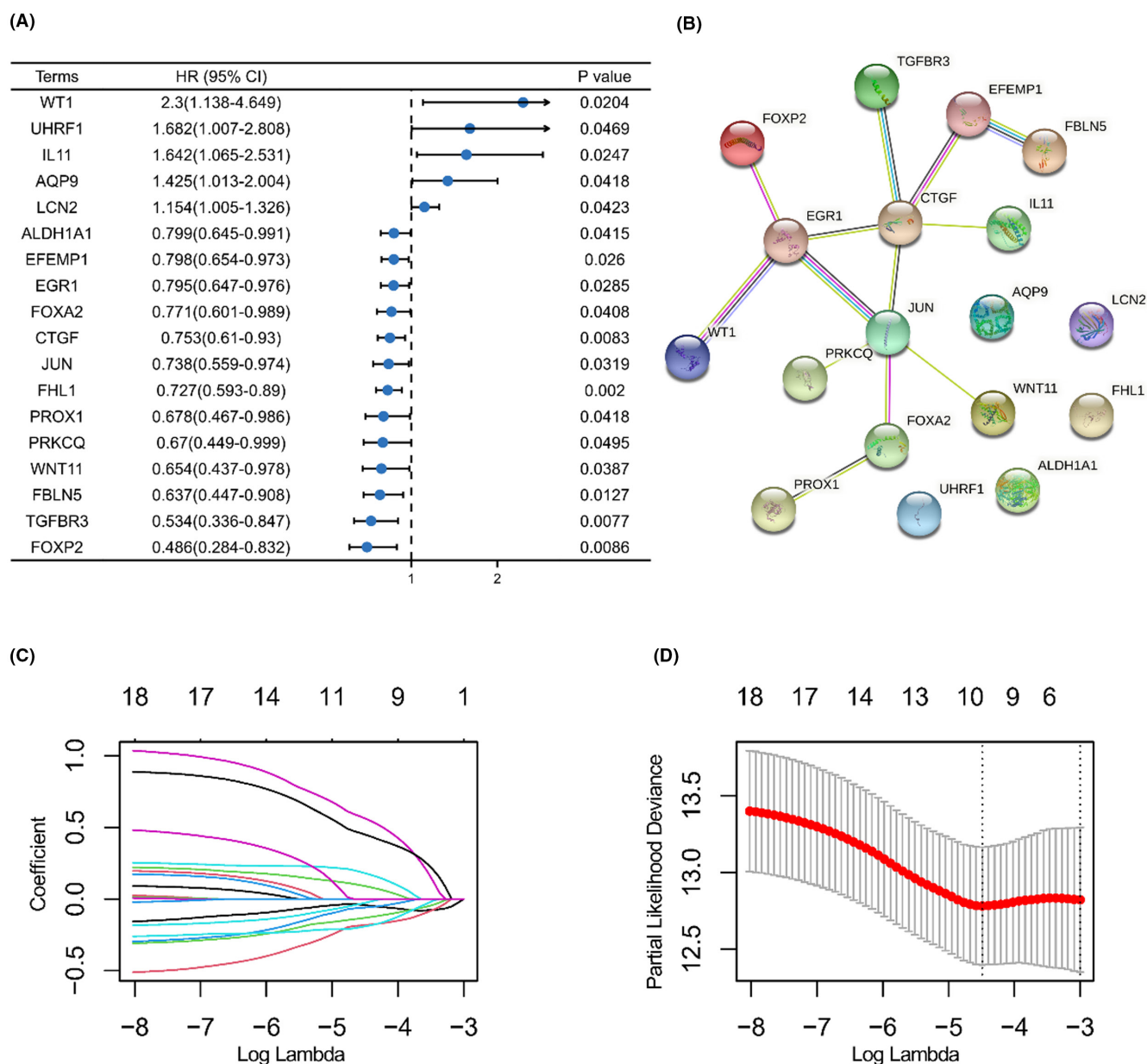


FIGURE 4 Screening of prognosis-related genes and identification of the 10-gene signature. (A) Forest plot showing the 18 prognosis-related genes filtered by Cox proportional-hazards analysis with criterion of a $p < 0.05$. (B) Potential PPI network of the 18 DE-ERGs. (C) LASSO coefficient profiles of the 18 DE-ERGs. (D) Lasso deviance profiles of the 18 DE-ERGs. The lambda selection criterion was based on the value of lambda giving a minimum mean cross-validation error; lambda min = 0.01124.

signature were 0.721, 0.715, and 0.664, respectively, with a C-index of 0.714 (Figure 5G). In the testing set, the AUCs were 0.858, 0.718, and 0.810, respectively, with a C-index of 0.761 (Figure 5H). In the THCA total dataset, the AUCs were 0.756, 0.717, and 0.700, respectively, with a C-index of 0.723 (Figure 5I). The optimal cutoff value for discriminating high risk patients of PTC on Illumina Hiseq 2500 platform was -1.47 . Surprisingly, we observed statistically significant differences of undifferentiated TC to differentiated TC in all the three datasets (Figure S1A–C). However, for the risk score using the classifier reported by Lin et al, the difference was statistically significant only in GSE33630. We supposed that the optimal cutoff value for discriminating ATC from PTC on Affymetrix Human Genome U133 Plus 2.0 Array platform was 1.78. (Figure S1D–F). Furthermore, in GSE138042, we verified the classifier in samples with different disease stage (T1/T2 versus T3/T4) and found a statistically significant difference ($p < 0.05$). In GSE82208, the difference of risk score between follicular adenoma and follicular thyroid cancer (FTC) was statistically significant ($p < 0.0001$). Also, in GSE60542, GSE58545, GSE29265, and GSE33630, the difference of risk score between normal and PTC samples was statistically significant ($p < 0.0001$). And for discriminating adenoma from FTC, the optimal cutoff value would be -0.95 . (Figure S2A–F). Full results are also listed in Tables S5 and S6.

3.5 | GSEA

GSEA in the 488 PTC cases from THCA dataset showed the representative altered biological functions of the high-risk group (Figure 6A–C). In the terms of KEGG pathway, the transcriptional alterations were annotated to the processes such as cell adhesion molecules and chemokine signaling pathway. In the terms of gene ontology, the transcriptional alterations were annotated to the processes such as negative regulation of cell killing and keratinization. For the oncological signatures, multiple classical oncological pathways, including the p53, the KRAS pathways were altered in the high-risk group. Full results are shown in Table S7.

3.6 | RT-qPCR quantification and EMT signatures in PTC cell lines

The relative 10-gene expression level of EMT signature in B-CPAP and KTC-1 cell were generated through RT-qPCR quantification. In 10 genes of EMT signature, the expression level of FOXP2, WT1, IL11, and UHRF1 were higher in KTC-1 than B-CPAP, the differences were statistically significant ($p < 0.05$). The differences of FHL1, TGFBR3, CTGF, FBLN5, WNT11, and AQP9 were not statistically significant ($p > 0.05$), as shown in Figure 7A. The EMT risk score of KTC-1 was higher than B-CPAP, as shown in Figure 7B. The expression level of TGFBR3, CTGF, FOXP2, WT1, and WNT11 was higher in KTC-1 than normal thyroid cell line Nthy-ori 3.1, and the differences were statistically significant ($p < 0.05$), as shown in Figure 7C. The risk score of KTC-1 was also higher than Nthy-ori 3.1, and the difference was statistically significant, as shown in Figure 7D. The expression level of CTGF and WNT11 was higher in B-CPAP than Nthy-ori 3.1, and the differences were statistically significant ($p < 0.05$), as shown in Figure 7E. The risk score of B-CPAP was higher than Nthy-ori 3.1, but the difference was not statistically significant, as shown in Figure 7D.

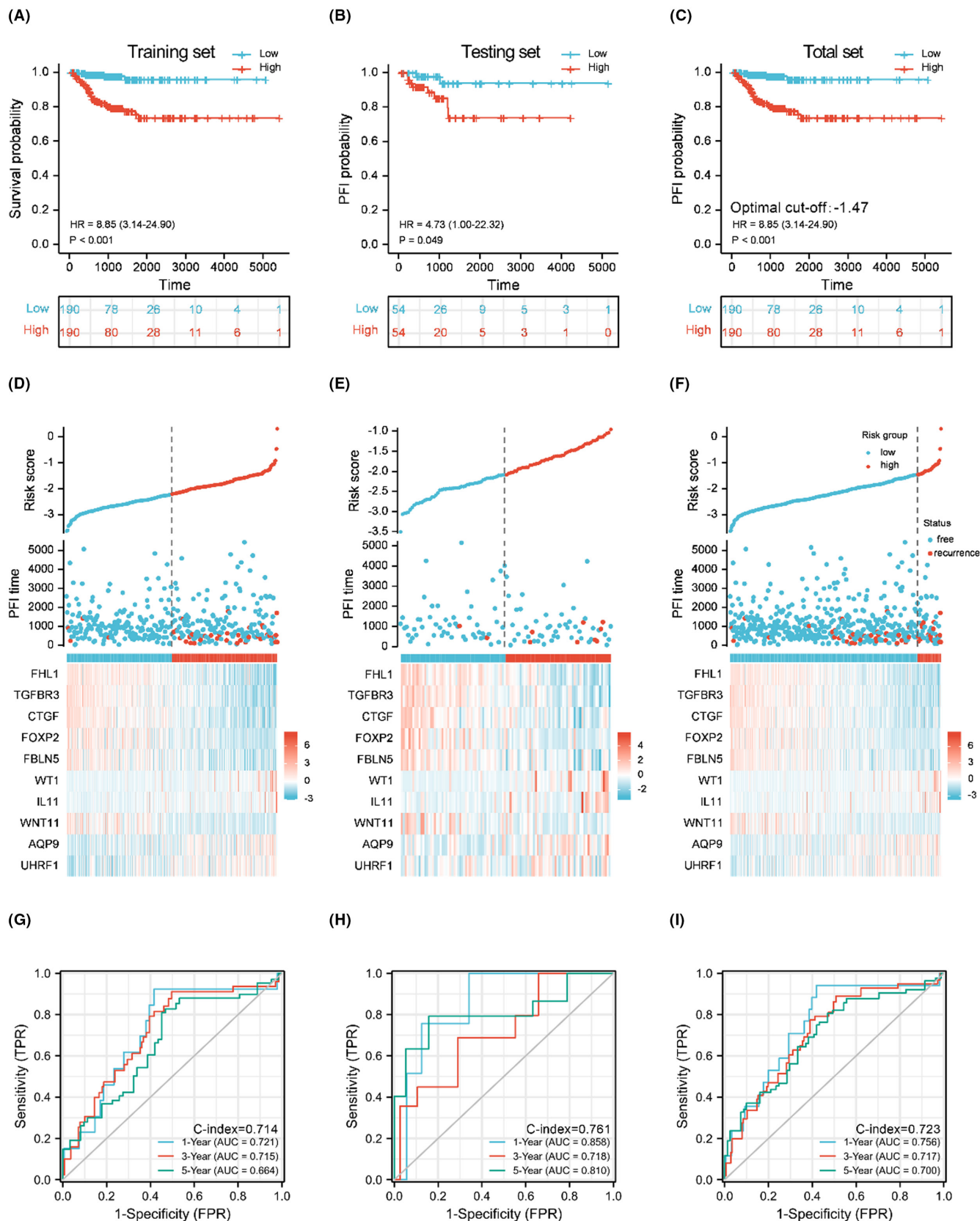
3.7 | Clinical correlation of the EMT signature

Next, we analyzed the correlation between the EMT signature and clinical characters. In groups divided by age, patients who were over 55 had higher risk score than patients who were younger than 55 (Figure 8A). In groups divided by BRAF status, wild-type samples had higher risk score than who were mutated (Figure 8C). In terms of extrathyroidal invasion, patients with no extra-invasion had lower risk score than those with extra-invasion existence (Figure 8D). Patients in T1/T2 had lower risk score than those in an advanced disease stage (T3/T4) (Figure 8E). Patients in N1 had higher risk score than those without lymph node metastases (N0) (Figure 8F). In groups divided by residual tumor, patients without residual tumor had lower risk score than

FIGURE 5 Verification of the discriminative power and predictive efficacy of the 10-gene signature. 80% of TCGA-THCA dataset cases ($n = 380$) were allocated into the training set (A, D, G) and the left 20% cases ($n = 188$) were allocated into the validation set (B, E, H); results for the total set are also presented (C, F, I). Cases were defined as “high risk” or “low risk” according to the median value (training and testing set) or optimal risk score cutoff determined by X-Tile (total set), and the PFI correlation was evaluated by K–M analysis. (A–C) The proportion of cases of PFI in the different risk category groups at different timepoints. (D–F) Scatterplots showing the distribution of recurrence (red dots) or progression-free events (blue dots) in the different risk groups. (G–I) The area under the curve (AUC) and C-index of the receiver operating characteristic (ROC) curve showing the predictive efficacy of 10-gene signature in 1-, 3-, and 5-year follow-up periods.

patients with residual tumor (Figure 8G). In groups divided by focality type, unifocal samples had higher risk score than multifocal samples (Figure 8H). Patients with aggressive histological (tall cell) type have higher risk

score than those with nonaggressive histological type (Figure 8I). Patients in stage 1/2 had lower risk score than those in stage 3/4 (Figure 8J). The differences were statistically significant ($p < 0.05$). For the terms divided



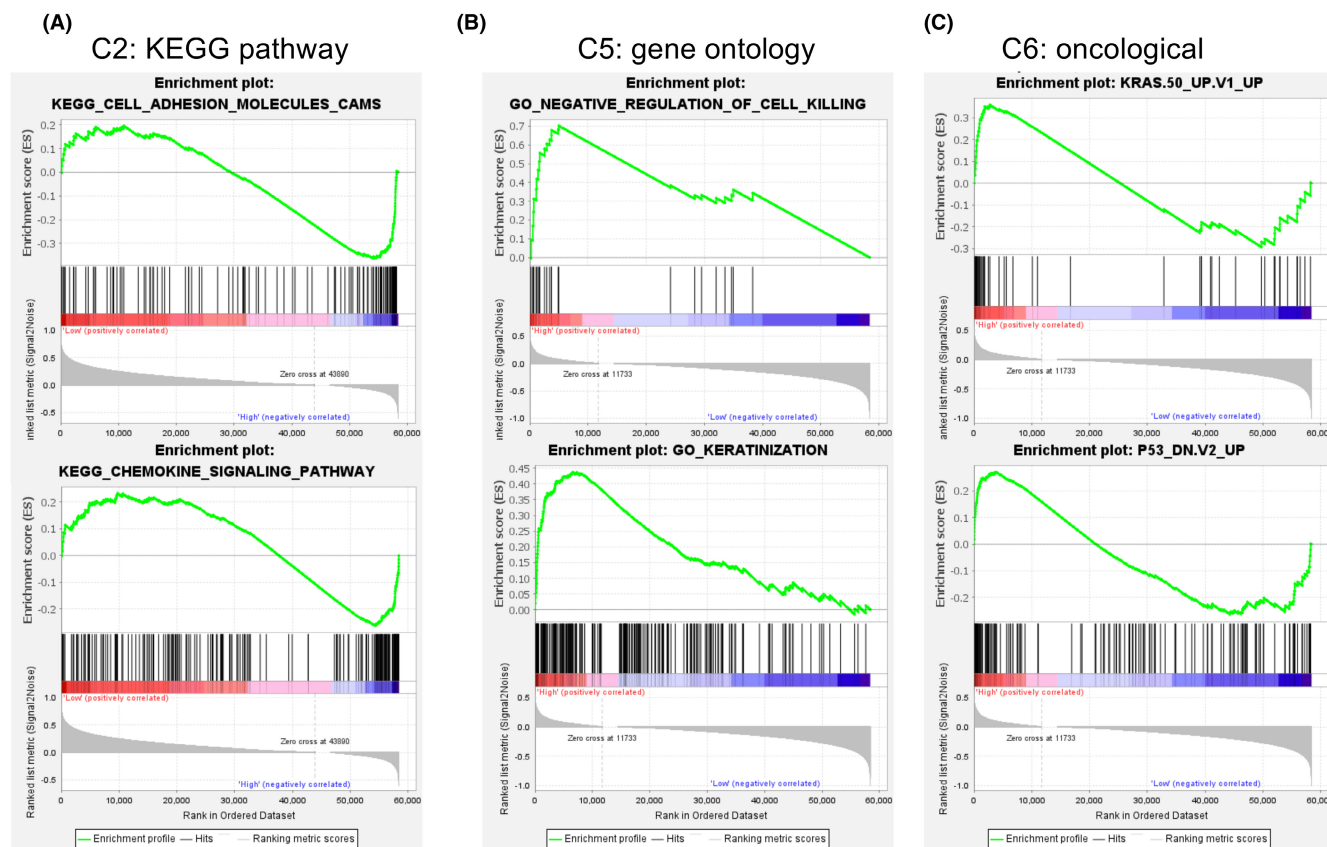


FIGURE 6 Gene set enrichment analysis of the turbulence at the molecular level caused by the signature. (A) The gene sets upregulated in the “high-risk” phenotype group of C2: KEGG pathway. (B) The gene sets upregulated in the “high-risk” phenotype group of C5: gene ontology. (C) The gene sets upregulated in the “high-risk” phenotype of C6: oncological signatures.

by gender, the difference was not statistically significant ($p > 0.05$), as shown in Figure 8B; full clinical characters are as shown in Table S8. We also compared AUCs of ATA stratification, MACIS score,³⁶ and risk score in 435 patients from TCGA-THCA with complete information. For ATA stratification, the AUCs for 1-year, 3-year, and 5-year PFI prediction were 0.731, 0.608, and 0.631, respectively, with a C-index of 0.667 (Figure 8K). For MACIS score, the AUCs for 1-year, 3-year, and 5-year PFI prediction were 0.773, 0.665, and 0.691, respectively, with a C-index of 0.664 (Figure 8L). For EMT signature risk score, the AUCs for 1-year, 3-year, and 5-year PFI prediction were 0.760, 0.726, and 0.682, respectively, with a C-index of 0.731 (Figure 8M), as shown in Table S9.

3.8 | Identification of independent prognosis-related parameters

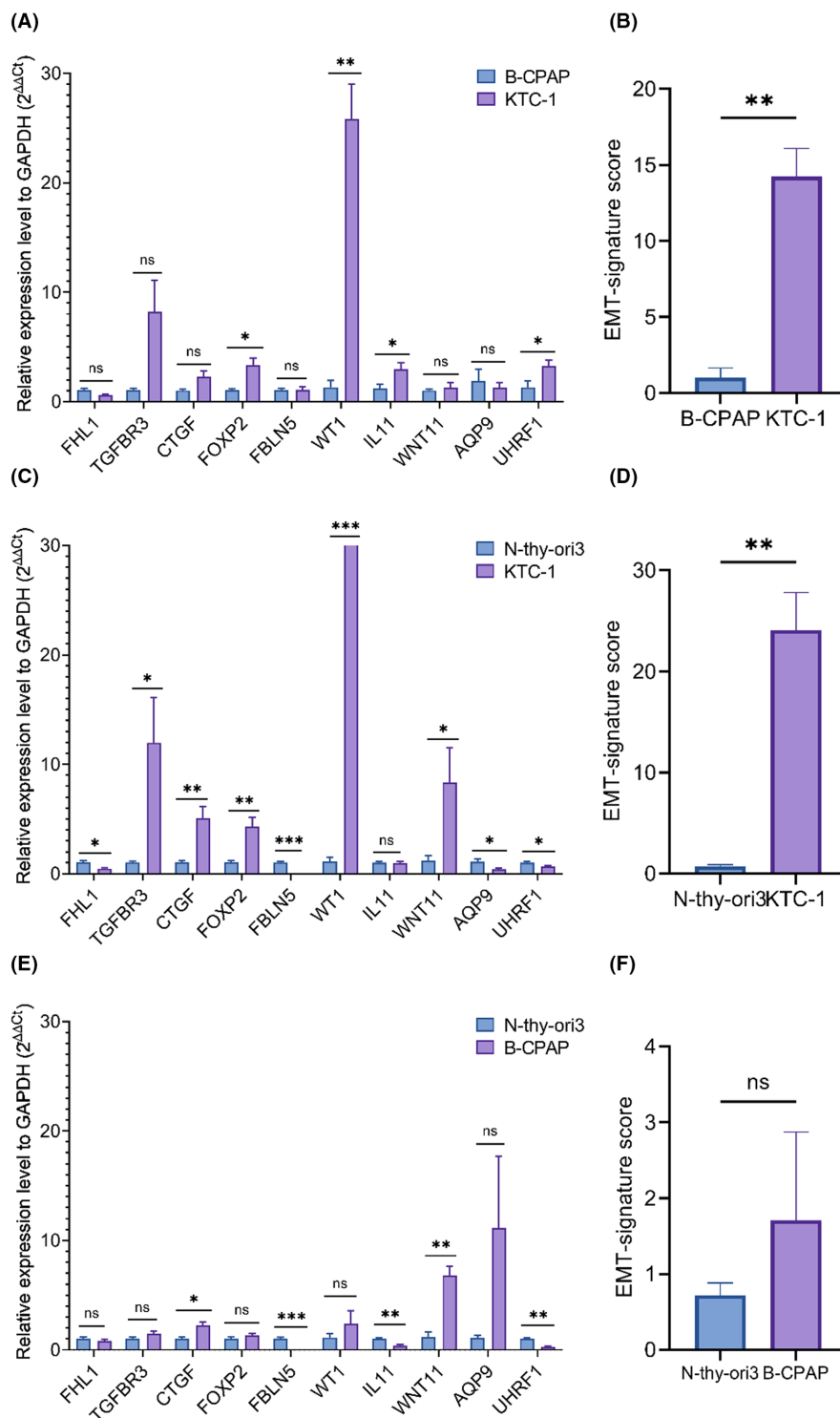
The univariate Cox analysis revealed that parameters including age (≥ 55), neoplasm_size (≥ 2 cm), histological_type (tall cell), T stage (T3/T4), M stage (M1), extrathyroidal extension, and risk score were significantly associated with prognosis ($p < 0.05$, Table 2). After

exclusion of samples with incomplete information and $p < 0.1$ in univariate analysis, a total of 422 patients were enrolled in multivariate analysis. In multivariate analysis, neoplasm_size (≥ 2 cm) and risk score were significantly associated with prognosis ($p < 0.05$, Table 2). Neoplasm_size (≥ 2 cm) and risk score were identified as independent prognosis-related factors in both the uni- and multivariate analysis.

3.9 | Establishment and validation of the novel signature-based nomogram

We constructed a stepwise Cox regression model including the parameters of risk score, age, TNM stage, neoplasm size, residual tumor, histological type, and RAS status. The model was visualized in a predictive nomogram as shown in Figure 9A. The AUCs for 1-year, 3-year, and 5-year PFI prediction were 0.889, 0.765, and 0.721, respectively, with a C-index of 0.776 (Figure 9B). The evaluation of the predictive nomogram using a calibration curve revealed the efficacy and robustness of the model for prediction of

FIGURE 7 RT-qPCR quantification and EMT signatures in PTC cell lines. (A, C, E) Relative expression level of 10 genes to GAPDH ($2^{-\Delta\Delta CT}$) in Nthy-ori 3.1, B-CPAP and KTC-1 cell line. (B, D, F) Risk score of PTC cell lines calculated by EMT signature. Data are presented as interleaved bar plot. Unpaired *t*-test with Welch's correction ($n = 4$), * $p < 0.05$, ** $p < 0.01$.



the prognosis of PTC patients (Figure 9C). For the purposes of better understanding and use of this predictive nomogram, we uploaded a program that facilitates visualization of the prediction results using the “DynNom” package of R (https://liuruiseurgeon.shinyapps.io/EMTbased_nomogram_PTC/).³⁷ The progression-free predictions are obtained by clicking the “predict” button after determination of the parameters (Figure S3).

4 | DISCUSSION

As previously described, the progression of PTC, lymph node metastasis, stemness, and anaplastic transformation are closely related to EMT³⁸; therefore, we hypothesized that EMT-based signature would perform well in risk prediction and stratification of PTC patients. In our study, we extracted the EMT-related genes from the dbEMT2, which curated over 1000 genes reported in more than 2000

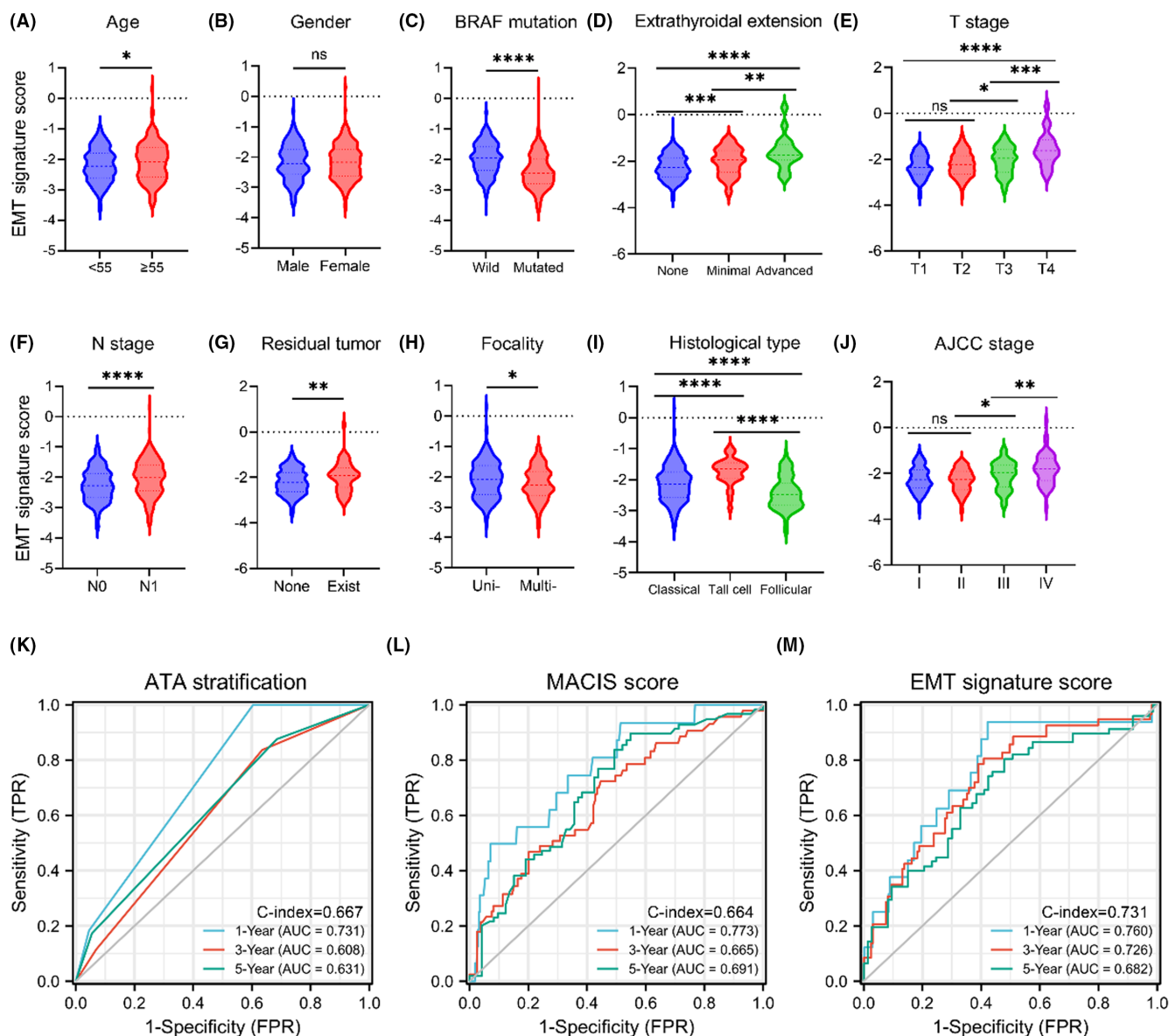


FIGURE 8 Clinical correlation of the EMT signature in TCGA-THCA cohort. (A-E) The distribution of the risk score according to different status of age, gender, BRAF mutation, extrathyroidal invasion, and T stage in the TCGA-THCA dataset. (F-J) The distribution of the risk score according to different status of N stage, residual tumor, focality type, histological type, and AJCC stage in the TCGA-THCA dataset. (K-M) The AUC and C-index of the ROC curve showing the predictive efficacy of ATA stratification, MACIS score and 10-gene signature in 1-, 3-, and 5-year follow-up periods. Data are presented as violin plot. Unpaired *t*-test with Welch's correction, **p* < 0.05, *****p* < 0.0001.

published studies.¹⁷ Using integrated bioinformatics analysis, we established a novel EMT-based signature for PFI prediction and recurrent risk stratification. Nomograms, which allow visualization of regression models, are widely used to provide an intuitive understanding of specific events based on enrolling meaningful characteristics.³⁹ Traditional approach to staging, including the ATA recurrence stratification system, are widely used and practicable, but cannot be used for individualized and prospective risk evaluation prior to radical resection. The novel EMT-related classifier and the relevant nomogram established

in this study is economically feasible and simple to apply. Based on droplet digital PCR (ddPCR) technique with fine-needle aspiration biopsy samples, the risk score could be obtained before surgery, making this a convenient approach to for decision-making in relation to future therapy.⁴⁰ For example, a more extensive neck dissection and close follow-up after surgery would be taken if the PTC patient scored as high risk but with a relatively low risk in imaging or pathology examination. Another scenario is that indetermined FNA cases for PTC but with a high risk on the classifier, decision of active surveillance

TABLE 2 Univariate and multivariate analyses of prognosis-related characters

Characteristics	Total (N)	Univariate analysis		Multivariate analysis	
		Hazard ratio (95% CI)	P-value	Hazard ratio (95% CI)	p-value
Gender	488				
Male	130	Reference			
Female	358	0.572 (0.320–1.024)	0.060	0.749 (0.391–1.435)	0.384
Age	488				
<55	325	Reference			
≥55	163	2.259 (1.289–3.961)	0.004	1.468 (0.659–3.268)	0.347
Risk score	488	4.335 (2.641–7.116)	<0.001	2.685 (1.462–4.933)	0.001
BRAF_status	488				
Wild type	211	Reference			
Mutated	277	1.455 (0.801–2.645)	0.218		
RAS_status	488				
Wild type	429	Reference			
Mutated	59	1.640 (0.768–3.504)	0.201		
Extrathyroid_extension	471				
None/Minimal	453	Reference			
Moderate/Advanced	18	2.100 (0.754–5.847)	0.156		
Neoplasm_size	474				
<2 cm	153	Reference			
≥2 cm	321	3.914 (1.547–9.901)	0.004	2.675 (1.022–7.001)	0.045
Histological_type	488				
Classical/Follicular	452	Reference			
Tall Cell	36	2.417 (1.084–5.389)	0.031	1.293 (0.500–3.345)	0.596
Anatomic_site	482				
Unilateral	379	Reference			
Bilateral	81	1.101 (0.513–2.365)	0.805		
Isthmus	22	0.413 (0.057–3.004)	0.382		
M_stage	487				
M0	479	Reference			
M1	8	5.630 (2.021–15.687)	<0.001	1.618 (0.434–6.031)	0.473
Ajcc_stage	486				
Stage I/II	324	Reference			
Stage III/IV	162	2.753 (1.567–4.839)	<0.001	1.386 (0.543–3.532)	0.495
N_stage	438				
N0	225	Reference			
N1	213	1.736 (0.950–3.172)	0.073	1.055 (0.526–2.114)	0.881
T_stage	486				
T1/T2	302	Reference			
T3/T4	184	2.806 (1.569–5.018)	<0.001	1.221 (0.577–2.581)	0.602

Bold values in Table 2 indicated the statistically significant ($p < 0.05$).

would be deliberated over. Since one of the optimum options for newly diagnosed low-risk PTC is active surveillance, the risk score would provide accessible information that helps the patient to understand the decision to adopt

this approach.^{41,42} Moreover, we analyzed the correlations between EMT signature with relevant clinical characters, and found significant differences in terms divided by age, gender, BRAF mutation, residual tumor, focality,

TABLE 3 External GEO datasets for validation of EMT signature

Datasets	References	Platform	Samples	Usage
GSE82208	Contributed by Wojtas B, et, al.	Affymetrix Human Genome U133 Plus 2.0 Array	25 Adenomas, 27 FTCs	External Validation
GSE29265	Contributed by Tomas G, et, al.	Affymetrix Human Genome U133 Plus 2.0 Array	20 PTCs, 20 Normals, 9 ATCs	External Validation
GSE33630	²⁴	Affymetrix Human Genome U133 Plus 2.0 Array	49 PTCs, 45 Normals, 11 ATCs	External Validation
GSE60542	²⁹	Affymetrix Human Genome U133 Plus 2.0 Array	33 PTCs, 30 Normals	External Validation
GSE76039	²⁵	Affymetrix Human Genome U133 Plus 2.0 Array	20 ATCs, 17 PDTCs	External Validation
GSE58545	²⁸	Affymetrix Human Genome U133A Array	27 PTCs, 18 Normals	External Validation
GSE138042	²⁷	Illumina HiSeq 3000 (Homo sapiens)	50 PTCs (T1/T2), 24 PTCs (T3/T4)	External Validation

TABLE 4 Sequences of primers for RT-qPCR analyses

Target gene	Primer	Sequence (5'-3')
GAPDH	Forward	AGTCCCTGCCACACTCAG
	Reverse	TACTTTATTGATGGTACATGACAAGG
FHL1	Forward	GACTGCTTCACCTGTAGTAACT
	Reverse	AGGTAACACACACAAAGCAATC
CTGF	Forward	ATTCTGTGGAGTATGTACCGAC
	Reverse	GTCTCCGTACATCTTCCTGTAG
FOXP2	Forward	AGCTCTGAAGTAAGCACAGTAG
	Reverse	TGCTGCTGTAAAAGAAGTTGTC
FBLN5	Forward	CTGTGACCCAGGATATGAACTT
	Reverse	TTGTAAATTGTAGCACGTCTGC
WT1	Forward	CCAATACAGAATACACACGCAC
	Reverse	CATCTGTAAGTGGGACAGCTTA
IL11	Forward	GTGGCCAGATACAGCTGTC
	Reverse	GAATTTGTCCCTCAGCTGTG
WNT11	Forward	TCTGCATGAAGAATGAGAAGGT
	Reverse	CACCAGTGGTACTTACAGTGG
AQP9	Forward	ATACCCAGCTCCGTATCTATCT
	Reverse	GTCAAAGATGGCAAAGACGATT
UHRF1	Forward	AATGTCAAGGGTGGCAAGAATA
	Reverse	GCCAGTATTTCACAACCTTGTA
TGFB3	Forward	TCCTCTGAATGGCTGCGGTACTC
	Reverse	GGCTGGAACCTGTATCACAATGGAG

extrathyroidal invasion, histological type, and TNM stage. To the best of our knowledge, PTC patients with elder age, female gender, residual tumor, tall cell type, and advanced disease stage had significantly higher rates of all-cause mortality and recurrence,^{43–46} the results enhanced the potential clinical significance of EMT signature. Another interesting finding is the value of EMT signature in discovering possible dedifferentiation of thyroid cancer, which would partially address the mechanism of the ability of EMT signature in progression prediction, since anaplastic transformation are closely related to EMT.³⁸ Further, the differences of risk score were statistically significant between follicular adenoma samples from follicular cancer, revealing a significant clinical value in FNA diagnosis, which remains challenging in differentiating FTC with benign follicular thyroid lesions.⁴⁷ To validate the signature in clinical practice, we implemented multiple GEO datasets and confirmed the robustness of the signature in discriminating normal thyroid sample from PTCs. Also, we applied qPCR experiment of 10 genes on two PTC cell lines. According to background literatures, the B-CPAP was originated from primary low-risk PTC and with a relatively low invasiveness.⁴⁸ The KTC-1, which originated from advanced metastatic PTC and refractory

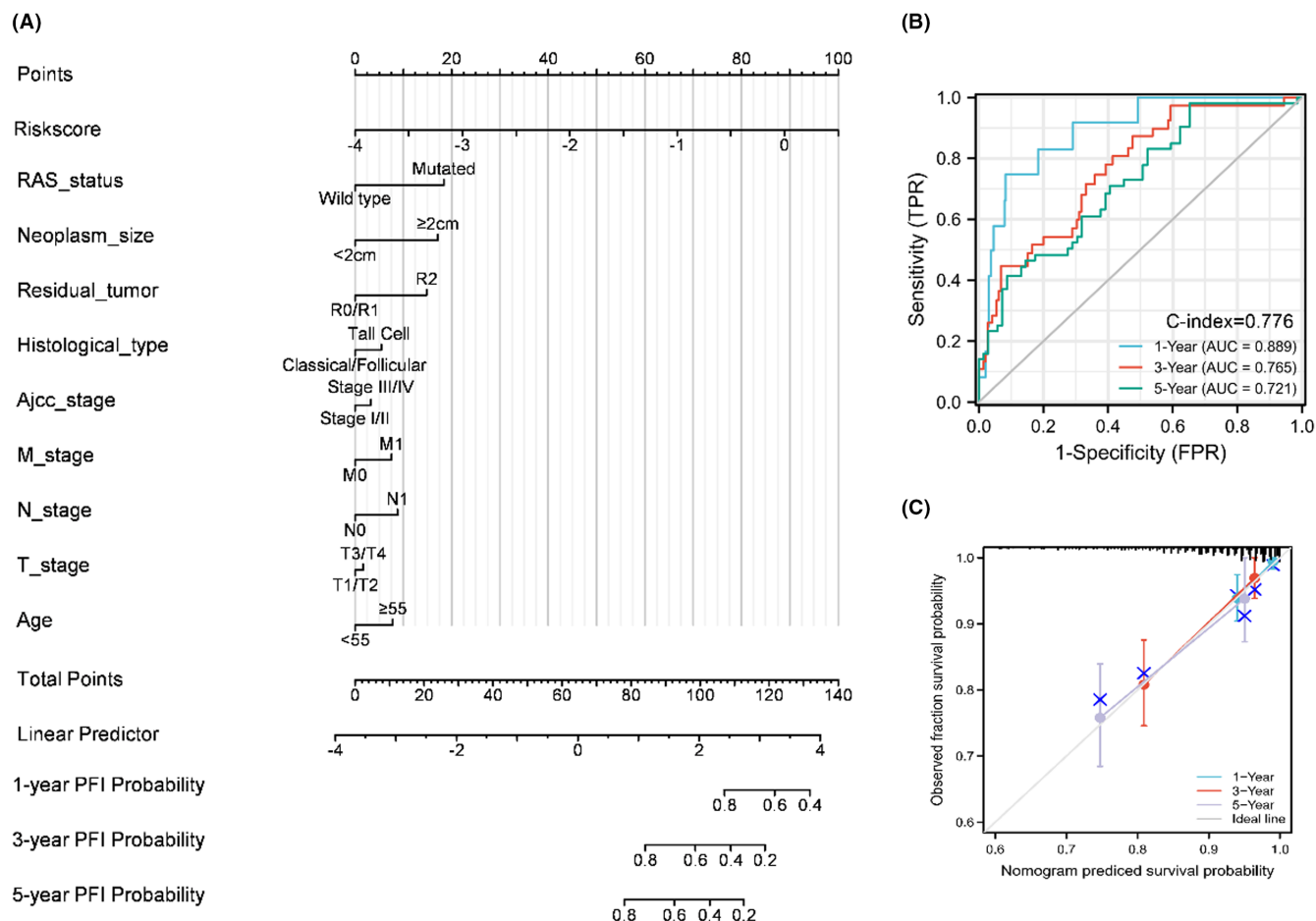


FIGURE 9 Establishment and assessment of the EMT-based nomogram. (A) Visualization of the adjusted Cox proportional regression model of the risk score and relevant clinical parameters. (B) 1-, 3-, and 5-year AUC and C-index of the nomogram. (C) Calibration curve of the nomogram for the predicted and actual 1-, 3-, and 5-year PFI.

to radio iodine therapy, was with lower degree of differentiation and invasiveness.³³ Several significantly highly expressed genes and the higher risk score in KTC-1 than in B-CPAP also confirmed the connection between EMT signature with PTC invasion and dedifferentiation. The results also confirmed the intrinsic connection between EMT signature with dedifferentiation and invasiveness. Furthermore, the risk score was also lower in immortalized thyroid follicular cell line than in B-CPAP. However, the difference was not statistically significant, indicating the mild invasiveness and progression for well differentiated cell line and circumstances for most original PTC cases.

Among our novel panel, half of them were previously identified in association with TC invasion and progression. For example, downregulated expression of FHL1 was associated with invasion of PTC.⁴⁹ CTGF plays an essential role in PTC cell spheroids formation.⁵⁰ LncRNA WT1 downregulates survivin by upregulating miR-203 to inhibit PTC cell proliferation.⁵¹ Moreover, UHRF1 was highly expressed in human ATC cells, knockdown of

UHRF1 inhibited proliferation and downregulated the expression of dedifferentiation marker CD97.⁵² All this evidence confirmed the intrinsic relationship between the novel 10-gene signature and EMT-induced aggressiveness of TC, although the roles of TGFBR3, FOXP2, FBLN5, IL11, WNT11, and AQP9 in TC have not been reported previously. Thus, fundamental research focused on these genes may lead to a progress in the development of novel therapeutic targets. GO analysis indicated that DE-ERGs were enriched in basement membrane, focal adhesion epithelial cell proliferation, extracellular matrix organization, and mesenchymal cell differentiation, which is consistent with the definition of ERGs. We assessed the potential turbulence caused by the novel ERG-based signature by GSEA, which can help reflect the disturbance of the whole genome at the translational level better than general overrepresentation analysis (ORA).⁵³ Multiple biological functions involved in several oncogenic and EMT-related processes were found to be upregulated in the high-risk group. These functions included process such as the cell adhesion molecules, chemokine signaling

pathway, and keratinization, as well as the p53 and KRAS pathways. These findings partially explain the mechanism underlying the shorter PFI of cases with the higher risk score, although further studies are required for confirmation. During the establishment of stepwise regression model, we incorporated the risk score, clinical characters enrolled in multivariate analysis, and relevant characters previously reported significantly correlated with prognosis of PTC such RAS mutation status.⁵⁴

Compared with the relatively low risk of cancer-related death, recurrence or persistent disease remains the main issue that influence the prognosis of PTC patients.⁵⁵ Reoperation results in a higher morbidity related to injury to the laryngeal nerve followed by vocal paralysis as the most serious complication.⁵⁶ Personalized follow-up based on precise prediction of recurrence risk after surgery would achieve a balance between excessive and insufficient management. The American Thyroid Association (ATA) published a risk stratification system (RSS) and response to therapy re-classification (RTR), which has been widely employed for recurrence risk prediction.⁵⁷ However, current prediction of recurrence risk is only possible after surgery or pathological examination.⁵⁸ Furthermore, the complexity and diversity of the biological behavior of PTC are not taken into consideration, thus impeding early prognosis. Massive amounts of information obtained from high-throughput methods, such as the multi-omics technologies and next-generating sequencing, have contributed tremendously to elucidation of the mechanism of TC.^{59,60} Additionally, multiple novel classifiers have been identified for prognosis prediction from different gene sets focused on specified biological functions, which innovated us for the development of the novel classifier based on EMT database.^{26,61}

Some limitations of our study should be noted. First, due to the lack of array-based independent gene expression profiling or sequence datasets with complete follow-up information, we had to apply internal verification of our results. Thus, our findings require further validation using an external independent dataset. In addition, as compared to the 26.57% to 26.95% of CV in TCGA-THCA cohort, a 10–20% CV for regular qRT-PCR is considered more acceptable. At last, the primary source of our study was from TCGA program, in which the samples were collected from North American people. Possible biases would occur when applying the classifier in different countries or regions, which should be addressed in further clinical study.

5 | CONCLUSION

In this study, we successfully identified a novel 10-gene classifier and established a classifier-based regression

model combining relevant clinical characteristics for prediction of prognosis in PTC patients. The novel gene classifier and nomogram may be useful and convenient for personalized management for thyroid cancer.

AUTHOR CONTRIBUTIONS

ZC and RL designed the study, carried out experiments, and obtained the data; RL, ZC, MP, and WM analyzed the data and conducted statistical analyses; RL, ZC, MP, and XL wrote the manuscript; HY and ZL revised and approved the manuscript.

ACKNOWLEDGMENT

The authors thank Dr. Wei Ge for her helpful suggestions on the study.

FUNDING INFORMATION

This research was supported by the National Natural Science Foundation of China (grant number: 82172727), the Nature Science Foundation of Beijing [grant number: 7202164], CAMS Innovation Fund for Medical Sciences (CIFMS) [grant number: 2016-12M-3-005], and CAMS Innovation Fund for graduate students [grant number: 2019-1002-44].

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

ETHICS STATEMENT

Not applicable

DATA AVAILABILITY STATEMENT

Data Availability Statement All the sequencing datas were obtained from the TCGA (<https://portal.gdc.cancer.gov/>) and the UCSC Xena database (<https://xenabrowser.net/datapages/>), the mutational datas were obtained from cBioPortal database (<http://www.cbioportal.org/>), the microarray datasets were obtained from GEO database (<https://www.ncbi.nlm.nih.gov/geo/>), other data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Ziwen Liu  <https://orcid.org/0000-0003-1574-8460>

REFERENCES

1. Lim H, Devesa SS, Sosa JA, Check D, Kitahara CM. Trends in thyroid cancer incidence and mortality in the United States, 1974–2013. *JAMA*. 2017;317(13):1338–1348.
2. Su D-H, Chang S-H, Chang T-C. The impact of locoregional recurrences and distant metastases on the survival of patients with papillary thyroid carcinoma. *Clin Endocrinol*. 2015;82(2):286–294.

3. Wang X, Wang X-L. Prognostic analysis of recurrence in children and adolescents with differentiated thyroid cancer. *Chin Med J (Engl)*. 2020;133(19):2281-2286.
4. Orosco RK, Hussain T, Brumund KT, Oh DK, Chang DC, Bouvet M. Analysis of age and disease status as predictors of thyroid cancer-specific mortality using the surveillance, epidemiology, and end results database. *Thyroid*. 2015;25(1):125-132.
5. Nakayama R, Horiuchi K, Susa M, et al. Anaplastic transformation of follicular thyroid carcinoma in a metastatic skeletal lesion presenting with paraneoplastic leukocytosis. *Thyroid*. 2012;22(2):200-204.
6. Ragazzi M, Ciarrocchi A, Sancisi V, Gandolfi G, Bisagni A, Piana S. Update on anaplastic thyroid carcinoma: morphological, molecular, and genetic features of the most aggressive thyroid cancer. *Int J Endocrinol* 2014;2014:790834, 1, 13.
7. Sorrenti S, Carbotta G, Di Matteo FM, et al. Evaluation of clinicopathological and molecular parameters on disease recurrence of papillary thyroid cancer patient: a retrospective observational study. *Cancers (Basel)*. 2020;12(12):3637.
8. Visciano C, Liotti F, Prevete N, Cali G, Franco R, Collina F, et al. Mast cells induce epithelial-to-mesenchymal transition and stem cell features in human thyroid cancer cells through an IL-8-Akt-slug pathway. *Oncogene*. 2015;34(40):5175-86.
9. Meng X, Kong D-H, Li N, et al. Knockdown of BAG3 induces epithelial-mesenchymal transition in thyroid cancer cells through ZEB1 activation. *Cell Death Dis*. 2014;5(2):e1092.
10. Liao Y, Hua Y, Li Y, et al. CRSP8 promotes thyroid cancer progression by antagonizing IKK α -induced cell differentiation. *Cell Death Differ*. 2020;08:1347-1363.
11. Zhou Q, Chen J, Feng J, Xu Y, Zheng W, Wang J. SOSTDC1 inhibits follicular thyroid cancer cell proliferation, migration, and EMT via suppressing PI3K/Akt and MAPK/Erk signaling pathways. *Mol Cell Biochem*. 2017;435(1-2):87-95.
12. Ruiz EML, Niu T, Zerfaoui M, et al. A novel gene panel for prediction of lymph-node metastasis and recurrence in patients with thyroid cancer. *Surgery*. 2020;167(1):73-79.
13. Chen H, Ma X, Yang M, Wang M, Li L, Huang T. A methylomics-associated nomogram predicts recurrence-free survival of thyroid papillary carcinoma. *Cancer Med*. 2020;9:7183-7193.
14. Agrawal N, Akbani R, Aksoy BA, et al. Integrated genomic characterization of papillary thyroid carcinoma. *Cell*. 2014;159(3):676-690.
15. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):pl1.
16. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139-140.
17. Zhao M, Liu Y, Zheng C, Qu H. dbEMT 2.0: an updated database for epithelial-mesenchymal transition genes with experimentally verified information and precalculated regulation information for cancer metastasis. *J Genet Genomics*. 2019;46(12):595-597.
18. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS Mei*. 2012;16(5):284-287.
19. Wu M, Yuan H, Li X, Liao Q, Liu Z. Identification of a five-gene signature and establishment of a prognostic nomogram to predict progression-free interval of papillary thyroid carcinoma. *Front Endocrinol*. 2019;10:790.
20. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1-22.
21. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res*. 2004;10(21):7252-7259.
22. Wu M, Li X, Zhang T, Liu Z, Zhao Y. Identification of a nine-gene signature and establishment of a prognostic nomogram predicting overall survival of pancreatic cancer. *Front Oncol*. 2019;9:996.
23. Schröder MS, Culhane AC, Quackenbush J, Haibe-Kains B. Survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics*. 2011;27(22):3206-3208.
24. Dom G, Tarabichi M, Unger K, et al. A gene expression signature distinguishes normal tissues of sporadic and radiation-induced papillary thyroid carcinomas. *Br J Cancer*. 2012;107(6):994-1000.
25. Landa I, Ibrahimipasic T, Boucai L, et al. Genomic and transcriptomic hallmarks of poorly differentiated and anaplastic thyroid cancers. *J Clin Invest*. 2016;126(3):1052-1066.
26. Lin P, Guo Y, Shi L, et al. Development of a prognostic index based on an immunogenomic landscape analysis of papillary thyroid cancer. *Aging (Albany NY)*. 2019;11(2):480-500.
27. Vladimirova U, Rumiantsev P, Zolotovskaia M, et al. DNA repair pathway activation features in follicular and papillary thyroid tumors, interrogated using 95 experimental RNA sequencing profiles. *Heliyon Maart*. 2021;7(3):e06408.
28. Rusinek D, Swierniak M, Chmielik E, et al. BRAFV600E-associated gene expression profile: early changes in the transcriptome, based on a transgenic mouse model of papillary thyroid carcinoma. *PLoS ONE*. 2015;10(12):e0143688.
29. Tarabichi M, Saiselet M, Trésallet C, et al. Revisiting the transcriptional analysis of primary tumours and associated nodal metastases with enhanced biological and statistical controls: application to thyroid cancer. *Br J Cancer*. 2015;112(10):1665-1674.
30. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-15550.
31. Kim M, Kim S-J, Xu Z, Ha SY, Byeon JH, Kang EJ, Shin SH, Yoo SK, Jee HG, Yoon SG, Yi JW, Bae JM, Yu HW, Chai YJ, Cho SW, Choi JY, Lee KE, Han W. BRAFV600E transduction of an SV40-immortalized Normal human thyroid cell line induces dedifferentiated thyroid carcinogenesis in a mouse xenograft model. *Thyroid* April 2020;30(4):487-500.
32. Fabien N, Fusco A, Santoro M, Barbier Y, Dubois PM, Paulin C. Description of a human papillary thyroid carcinoma cell line. Morphologic Study and Expression of Tumoral Markers. *Cancer*. 1994;73(8):2206-2212.
33. Kurebayashi J, Tanaka K, Otsuki T, et al. All-trans-retinoic acid modulates expression levels of thyroglobulin and cytokines in a new human poorly differentiated papillary thyroid carcinoma cell line, KTC-1. *J Clin Endocrinol Metab*. 2000;85(8):2889-2896.
34. Esmaeili M, Jennek S, Ludwig S, et al. The tumor suppressor ING1b is a novel corepressor for the androgen receptor and

- induces cellular senescence in prostate cancer cells. *J Mol Cell Biol*. 2016;8(3):207-220.
35. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607-D613.
 36. Jang HW, Lee JI, Kim HK, et al. Identification of a cut-off for the MACIS score to predict the prognosis of differentiated thyroid carcinoma in children and young adults. *Head Neck Mei*. 2012;34(5):696-701.
 37. Jalali A, Alvarez-Iglesias A, Roshan D, Newell J. Visualising statistical models using dynamic nomograms. *PLoS One*. 2019;14(11):e0225253.
 38. Shakib H, Rajabi S, Dehghan MH, Mashayekhi FJ, Safari-Alighiarloo N, Hedayati M. Epithelial-to-mesenchymal transition in thyroid cancer: a comprehensive review. *Endocrine*. 2019;66(3):435-455.
 39. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology – more than meets the eye. *Lancet Oncol*. 2015;16(4):e173-e180.
 40. Cazacu IM, Semaan A, Stephens B, et al. Diagnostic value of digital droplet polymerase chain reaction and digital multiplexed detection of single-nucleotide variants in pancreatic cytology specimens collected by EUS-guided FNA. *Gastrointest Endosc*. 2021;93(5):1142-1151.e2.
 41. Ito Y, Miyauchi A, Oda H. Low-risk papillary microcarcinoma of the thyroid: a review of active surveillance trials. *Eur J Surg Oncol*. 2018;44(3):307-315.
 42. Saravana-Bawan B, Bajwa A, Paterson J, McMullen T. Active surveillance of low-risk papillary thyroid cancer: a meta-analysis. *Surgery*. 2020;167(1):46-55.
 43. Nixon IJ, Shaha AR. Implications of extra nodal extension in differentiated thyroid cancer. *J Surg Oncol*. 2017;116(4):459-460.
 44. Veronese N, Luchini C, Nottegar A, et al. Prognostic impact of extra-nodal extension in thyroid cancer: a meta-analysis. *J Surg Oncol*. 2015;112(8):828-833.
 45. Kazaure HS, Roman SA, Sosa JA. The impact of age on thyroid cancer staging. *Curr Opin Endocrinol Diabetes Obes*. 2018;25(5):330-334.
 46. Cartwright S, Fingeret A. Contemporary evaluation and management of tall cell variant of papillary thyroid carcinoma. *Curr Opin Endocrinol Diabetes Obes*. 2020;27(5):351-357.
 47. Cipriani NA, Nagar S, Kaplan SP, et al. Follicular thyroid carcinoma: how have histologic diagnoses changed in the last half-century and what are the prognostic implications? *Thyroid*. 2015;25(11):1209-1216.
 48. Arturi F, Russo D, Bidart JM, Scarpelli D, Schlumberger M, Filetti S. Expression pattern of the pendrin and sodium/iodide symporter genes in human thyroid carcinoma cell lines and human thyroid tumors. *Eur J Endocrinol*. 2001;145(2):129-135.
 49. Dai J, Yu X, Han Y, et al. TMT-labeling proteomics of papillary thyroid carcinoma reveal invasive biomarkers. *J Cancer*. 2020;11(20):6122-6132.
 50. Warnke E, Pietsch J, Wehland M, et al. Spheroid formation of human thyroid cancer cells under simulated microgravity: a possible role of CTGF and CAV1. *Cell Commun Signal*. 2014;12:32.
 51. Le F, Luo P, Ouyang Q, Zhong X. LncRNA WT1-AS downregulates Survivin by upregulating miR-203 in papillary thyroid carcinoma. *Cancer Manag Res*. 2020;12:443-449.
 52. Wang B-C, Lin G-H, Wang B, et al. UHRF1 suppression promotes cell differentiation and reduces inflammatory reaction in anaplastic thyroid cancer. *Oncotarget*. 2018;9(62):31945-31957.
 53. Bild A, Febbo PG. Application of a priori established gene sets to discover biologically important differential expression in microarray data. *Proc Natl Acad Sci U S A*. 2005;102(43):15278.
 54. Cabanillas ME, Dadu R, Iyer P, Wanland KB, Busaidy NL, Ying A, Gule-Monroe M, Wang JR, Zafereo M, Hofmann MC. Acquired secondary RAS mutation in BRAFV600E-mutated thyroid cancer patients treated with BRAF inhibitors. *Thyroid*. 2020;30(9):1288-96.
 55. Calò PG, Lombardi CP, Podda F, Sessa L, Santini L, Conzo G. Role of prophylactic central neck dissection in clinically node-negative differentiated thyroid cancer: assessment of the risk of regional recurrence. *Updates Surg*. 2017;69(2):241-248.
 56. Wong H, Wong KP, Yau T, et al. Is there a role for unstimulated thyroglobulin velocity in predicting recurrence in papillary thyroid carcinoma patients with detectable thyroglobulin after radioiodine ablation? *Ann Surg Oncol*. 2012;19(11):3479-3485.
 57. Pandya A, Caoili EM, Jawad-Makki F, et al. Limitations of the 2015 ATA guidelines for prediction of thyroid cancer: a review of 1947 consecutive aspirations. *J Clin Endocrinol Metab*. 2018;103(9):3496-3502.
 58. Haugen BR, Alexander EK, Bible KC, 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):1-133.
 59. Luo D, Zhan S, Xia W, Huang L, Ge W, Wang T. Proteomics study of serum exosomes from papillary thyroid cancer patients. *Endocr Relat Cancer*. 2018;25(10):879-891.
 60. Luo Y-H, Liang L, He R-Q, et al. RNA-sequencing investigation identifies an effective risk score generated by three novel lncRNAs for the survival of papillary thyroid cancer patients. *Oncotarget*. 2017;8(43):74139-74158.
 61. Hu G, Feng H-F, Zhan H. Identification of an autophagy-related signature predicting overall survival for papillary thyroid carcinoma. *Dose Response Maart*. 2020;18(1):1559325819899265.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Liu R, Cao Z, Pan M, et al. A novel prognostic model for papillary thyroid cancer based on epithelial-mesenchymal transition-related genes. *Cancer Med*. 2022;11:4703-4720. doi: [10.1002/cam4.4836](https://doi.org/10.1002/cam4.4836)