



OPEN A Comprehensive multi-network analysis of ceRNAs and transcription factors for papillary thyroid carcinoma diagnosis and prognosis

Seyede Elmira Yazdi Rouholamini¹ , Seyed-Morteza Javadirad² & Mohsen Kolahdouzan³

Papillary thyroid carcinoma (PTC) is a malignancy with an ambiguous etiology. The competitive endogenous RNA (ceRNA) hypothesis provides a framework for clarifying the molecular mechanisms that drive carcinogenesis. In this study, we constructed a novel ceRNA network to identify reliable diagnostic and prognostic indicators applicable across all stages of PTC. Transcriptome analysis was performed to identify stage-specific hub genes using the MCC, IVI, and MCODE algorithms. A novel five-layer ceRNA network and its associated regulatory network (DE-TF) were constructed. Receiver operating characteristic curves were used to evaluate the diagnostic performance of elements within both networks. A risk assessment model was developed by identifying key genes from the ceRNA and DE-TF components through univariable Cox regression and LASSO regression analyses. RNA-seq findings were validated by RT-qPCR. The correlations between gene expression levels and blood calcium levels were examined. The ceRNA and DE-TF networks contained 33 and 21 components, respectively. Logistic regression analysis identified *PKMYT1*, *E2F1*, *NFATC1*, *STAT6*, *E2F3*, *LINC02910*, *GAS5*, and *TK1* as reliable diagnostic markers for PTC, achieving an AUC of 96.9%. Among these, *PKMYT1* and *GAS5* were stage-specific markers, showing significant upregulation in highly aggressive PTC tumors compared to less aggressive ones. Both genes demonstrated strong diagnostic value in differentiating high- from low-aggressive tumors, with AUCs of 0.81 and 0.87, respectively. *circMET*, which was overexpressed in both low- and high-aggressive tumors, showed diagnostic potential in distinguishing low-aggressive tumors from normal adjacent tissues (AUC=0.81). *GAS5* expression demonstrated an association with blood calcium levels. The SERN prognostic model, including *STAT6*, *E2F1*, *RMI2*, and *NR4A1*, illustrates the importance of these four genes as reliable prognostic markers for overall survival in PTC. Three components of the ceRNA network—*PKMYT1*, *GAS5*, and *circMET*—were significantly associated with PTC aggressiveness. *PKMYT1* and *GAS5* demonstrated strong diagnostic value in distinguishing high-aggressive from low-aggressive tumors, while *circMET* showed notable diagnostic efficacy in differentiating low-aggressive PTC tumors from adjacent normal tissues. Furthermore, *GAS5* expression levels were correlated with blood calcium levels.

Keyword Papillary thyroid carcinoma, ceRNA network, DE-TF network

Abbreviations

AR	Androgen receptor
ASF1B	Anti-silencing function 1B histone chaperone
CDT1	Chromatin licensing and DNA replication factor 1
ceRNA	Competing endogenous RNA
circRNA	Circular RNA

¹Department of Cell and Molecular Biology and Microbiology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan 81746-73441, Iran. ²Department of Cell and Molecular Biology and Microbiology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan 81746-73441, Iran. ³Department of Surgery, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran. ✉email: javadirad@yahoo.com

DEcircs	Differentially expressed circRNAs
DEGs	Differentially expressed genes
DElncs	Differentially expressed lncRNAs
DEmiRs	Differentially expressed microRNAs
DEmRNAs	Differentially expressed messenger RNAs
DE pseudogenes	Differentially expressed pseudogenes
DE-TF	Differentially expressed transcription factor
<i>E2F1</i>	E2F transcription factor 1
<i>E2F3</i>	E2F transcription factor 3
<i>E2F4</i>	E2F transcription factor 4
<i>FOLH1B</i>	Folate hydrolase 1B
<i>FOS</i>	Fos proto-oncogene, AP-1 transcription factor subunit
<i>FOXA1</i>	Forkhead box A1
<i>GAS5</i>	Growth arrest-specific 5
GEPIA	Gene expression profiling interactive analysis
GTEx	Genotype–tissue expression
<i>HES1</i>	Hes family bHLH transcription factor 1
IVI	Integrated Value of Influence
LASSO	Least absolute shrinkage and selection operator
<i>LEF1</i>	Lymphoid enhancer binding factor 1
<i>LINC02910</i>	Long intergenic nonprotein coding RNA
lncRNA	Long noncoding RNA
MCC	Maximal clique centrality
MCODE	Molecular complex detection
MRE	MiRNA response element
<i>NFATC1</i>	Nuclear factor of activated T cells 1
<i>NR4A1</i>	Nuclear receptor subfamily 4, group A member 1
<i>PKMYT1</i>	Protein kinase membrane associated tyrosine/threonine 1
PPI network	Protein–protein interaction network
PTC	Papillary thyroid carcinoma
ROC	Receiver operating characteristic
<i>RMI2</i>	RecQ-Mediated genome instability 2
<i>SP3</i>	Sp3 transcription factor
<i>STAT6</i>	Signal transducer and activator of transcription 6
THCA	Thyroid cancer
<i>THRB</i>	Thyroid hormone receptor beta
<i>TK1</i>	Thymidine kinase 1
<i>TP53</i>	Tumor protein P53

Papillary thyroid carcinoma (PTC) accounts for approximately 85% of differentiated thyroid tumors^{1,2}. Although PTC generally carries a favorable prognosis, with a five-year survival rate exceeding 90%, outcomes worsen significantly in aggressive cases. Recurrence occurs in nearly 30% of patients, and recurrent disease can result in mortality rates of up to 60%^{3–5}. Despite extensive research, the genetic basis of PTC remains incompletely understood, with only a limited number of gene mutations identified to date⁶.

The competitive endogenous RNA (ceRNA) hypothesis provides a framework to explain tumorigenesis, based on the observation that many cellular transcripts contain microRNA (miRNA) response element (MRE) sequences^{7,8}. Within this system, ceRNA components—including miRNAs, pseudogenes, long noncoding RNAs (lncRNAs), circular RNAs (circRNAs), and mRNAs^{9,10}—compete for binding to shared MREs, thereby regulating the expression of corresponding mRNAs. Recent studies have shown that pseudogenes, lncRNAs, and circRNAs contribute to cancer development through miRNA sponging mechanisms^{11–15}. Thus, maintaining balance among ceRNA transcripts is essential, and their dysregulation can drive pathological processes, including oncogenesis. Moreover, transcriptional regulation of ceRNA network components by specific transcription factors (TFs) may further illuminate mechanisms of cancer progression.

However, there remains a lack of research integrating ceRNA and TF networks to identify key causal genes in PTC. To address this gap, we performed RNA-seq transcriptome analysis to identify overlapping differentially expressed genes (DEGs) across various PTC stages and subsequently highlighted hub genes for further validation. Next, we constructed ceRNA and DE-TF networks and validated candidate genes using RT-qPCR. Finally, we performed downstream analyses, including receiver operating characteristic (ROC) curve evaluation and survival analysis, to assess the diagnostic and prognostic significance of candidate genes within these networks. A schematic representation of the entire investigation is shown in Fig. 1.

Materials and methods

Expression data

RNA-seq expression data for thyroid cancer (THCA) were obtained from The Cancer Genome Atlas (TCGA) portal (<https://portal.gdc.cancer.gov/>). Data from the TCGA-THCA project were retrieved using the *TCGAbiolinks* package¹⁶, yielding 571 thyroid tissue samples, including both PTC and normal tissues. After excluding eight metastatic samples and two with ambiguous pathological staging, 561 samples remained, comprising 502 PTC tissues and 59 normal tissues. Among the PTC cohort, 283 cases were stage I, 52 stage II, 112 stage III, and 55 stage IV. Because circRNA expression data are not available in TCGA, the GSE171011

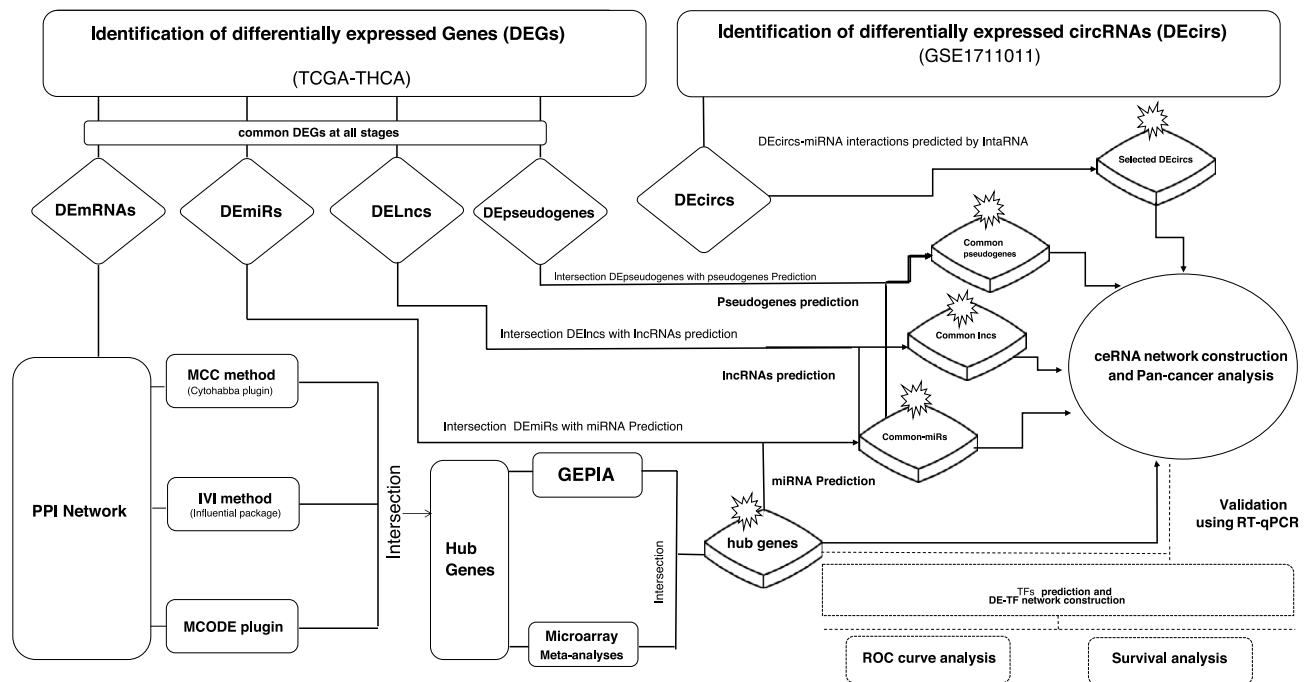


Fig. 1. Schematic representation of multilayer gene analysis in papillary thyroid carcinoma (PTC).

dataset was used to evaluate circRNA expression. However, the samples in GSE171011 were not staged, limiting the ability to assess circRNA expression across different PTC stages.

Data filtering and normalization

Data filtering was applied to remove low-quality samples, excluding those with total read counts below 50% of the median library size. The remaining data was processed using *edgeR* package¹⁷. Raw counts were organized into a DGEList object, and lowly expressed genes were filtered out using the *filterByExpr* function to improve statistical reliability while preserving biological relevance. To account for differences in library composition, normalization was performed using the weighted trimmed mean of M-values (TMM) method. The TMM-normalized data were then prepared for linear modeling with the *voom* function, which transformed the counts into $\log_2$ -counts per million (logCPM) and assigned precision weights to each observation based on the mean-variance relationship of the RNA-seq data. This transformation addressed non-constant variance and produced a matrix of normalized, weighted logCPM values that was subsequently used for differential expression analysis, ensuring statistical robustness and validity¹⁸.

Identifying DEGs across all stages of PTC

Normalized expression data were analyzed for differential expression using the *Limma* package¹⁹. Multiple testing correction was performed using the *Benjamini-Hochberg* procedure to control the false discovery rate (FDR). To identify stage-specific DEGs, contrast matrices were generated, and each PTC stage (I, II, III, and IV) was compared individually against normal thyroid tissue. Protein-coding DEGs (DEmRNAs) were defined as transcripts with a $\log_2$ fold change (logFC) greater than 1.0 and an FDR < 0.05. Noncoding RNAs—including miRNAs, lncRNAs, and pseudogenes—were designated as DEmiRs, DELncs, and DEpseudogenes, respectively, and filtered using an FDR < 0.05 irrespective of logFC. After identifying stage-specific RNAs, intersections across the four stages were examined to identify overlapping molecules. Using comparable approaches, consistently deregulated noncoding RNAs (DELncs, DEmiRs, and DEpseudogenes) shared among all PTC stages were determined. Differentially expressed circRNAs (DECircs) were defined as transcripts with logFC > 1.0 and FDR < 0.05.

Selection of hub genes via network analysis

DEmRNAs were used to construct a PPI network with a minimum interaction confidence of 0.4²⁰. Important nodes were prioritized using complementary centrality measures spanning local, semilocal, and global topology. Maximal clique centrality (MCC) was applied as a local measure given its strong performance in detecting biologically meaningful hubs²¹. The Integrated Value of Influence (IVI) served as an integrative index, computed with the *Influential* R package, to capture multiple topological dimensions by combining key local (degree centrality, ClusterRank), semilocal (neighborhood connectivity, local H-index), and global (betweenness centrality, collective influence) metrics in a bias-mitigating manner²². Candidate genes were defined as the overlap between the top 10% ranked by MCC and the top 10% ranked by IVI.

To detect densely connected modules, the MCODE plugin in Cytoscape²³ was used, and the highest-scoring cluster was selected for downstream analysis. Hub genes were defined at the intersection of the MCODE-derived cluster and the candidate gene set.

Verification of hub genes by RNAseq and microarray data analysis

Gene Expression Profiling Interactive Analysis (GEPIA) was used to validate the expression levels of PTC hub genes, leveraging RNA-seq data from the TCGA-THCA project²⁴. To increase the number of normal tissue samples to 337, additional normal tissue data were incorporated from the Genotype-Tissue Expression (GTEx) portal (<https://www.gtexportal.org>). Differential expression analysis was performed using the *Limma* package in a two-step process. First, a single-dataset analysis compared PTC tumors with normal tissues, both obtained from the TCGA-THCA database. Second, a combined analysis compared PTC tissues from TCGA-THCA with pooled normal tissues from both TCGA and GTEx databases. Genes with a $\text{LogFC} > 1.0$ and $P < 0.05$ were considered GEPIA-validated. Additional validation was conducted using microarray data from the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov>). Datasets involving treatments or therapies were excluded. Seven datasets were included for analysis: GSE3678, GSE9115, GSE27155, GSE33630, GSE53157, GSE60542, and GSE138198 (Table S1). Data were normalized using log2 transformation and quantile normalization. Hierarchical clustering and principal component analysis (PCA) were performed to identify outliers. Metadata from individual datasets were integrated to create a unified microarray dataset, with batch effects corrected using the SVA²⁵ and *Limma* packages. DEGs between PTC and normal tissues were identified using *Limma* ($\text{FDR} < 0.05$). Hub genes were defined as those overlapping between the DEGs from microarray analysis and the genes validated via GEPIA.

Assessment of confounding factors

The *recount* package was employed to assess potential confounding effects of sex and age on hub gene expression levels²⁶. Following integration of TCGA and GTEx data, batch effects were identified and corrected using the SVA package, thereby enhancing analytic precision across cohorts. To evaluate the influence of sex, four comparisons were performed: i) PTC tissues from female patients versus corresponding normal female tissues, ii) PTC tissues from male patients versus corresponding normal male tissues, iii) all PTC tissues versus all normal tissues, irrespective of sex, and iv) direct comparison of hub gene expression between male and female subjects. For age-related analysis, a comparative study was conducted to assess the impact of age on hub gene expression. The median age among PTC patients in both the TCGA/GTEx dataset and fresh tissue samples was 46 and 46.5 years, respectively. Accordingly, PTC tissues were divided into two groups: patients aged 46 years or older and those younger than 46 years.

Identification of ceRNA network components

In addition to hub genes, regulatory molecules targeting these genes—including miRNAs, lncRNAs, pseudogenes, and circRNAs—were identified. Four databases (miRDB²⁷, RNAInter²⁸, miRTarBase²⁹, and miRWalk³⁰) were used to predict miRNA targets. lncRNAs and pseudogenes were identified using the RNAInter, StarBase³¹, and miRcode³² databases. MiRNA selection followed two criteria: first, a hub gene had to be targeted by miRNA(s) present in at least two of the databases; second, the identified miRNAs were required to overlap with DE miRs. MiRNAs meeting both criteria were considered common miRs and used to predict the corresponding lncRNAs, pseudogenes, and circRNAs. Common lncRNAs were defined as those overlapping between predicted lncRNAs and DE lncs, while common pseudogenes were those overlapping between predicted pseudogenes and DE pseudogenes. To identify DE circs potentially interacting with the common miRs, the IntaRNA tool³³ was employed to assess circRNA–miRNA interactions. CircRNAs with predicted interaction energies below -7 kcal/mol were incorporated into the ceRNA network.

Coexpression analysis and construction of a ceRNA network

Normalized expression values of genes within the ceRNA network were used to construct a coexpression network. Data analysis was performed using the *Hmisc* package³⁴, while the *corrplot* package³⁵ was employed to generate graphical representations of the correlation matrix. Visualization of the ceRNA network, including hub genes, common miRs, common lncRNAs, common pseudogenes, and DE circRNAs, was carried out using Cytoscape software.

Pan-cancer analysis of the ceRNA network

To assess the expression levels of ceRNA network components across various cancer types, the *TIMER 2.0* resource was utilized³⁶. Statistical significance was determined using the Wilcoxon test, with significance levels indicated as follows: * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ (denoted by asterisks).

DE-TF network construction

The Open Access Repository of Transcriptional Interactions (ORTI) was used to identify experimentally validated transcription factors (TFs)³⁷. Two separate networks were constructed: one involving TFs associated with hub genes and another related to ceRNA network genes. A differentially expressed TF (DE-TF) network was created by identifying components of the TF networks that overlapped with DE mRNAs. Visualization and identification of interactions within the DE-TF network were facilitated using the *ggalluvial* package.

Interstage analysis of the ceRNA/DE-TF network components

RNA-seq data from TCGA and GTEx were analyzed using GEPIA. The expression levels of genes within the ceRNA and DE-TF networks were evaluated across the four clinical stages of PTC. Statistical significance was defined as a P value of the F statistic ($\Pr(>F)$) less than 0.05.

Development of a prognostic model for the components of the ceRNA and DE-TF networks

To develop a robust prognostic model and mitigate the risk of overfitting, a comprehensive internal validation was performed using a stratified tenfold cross-validation (CV) approach^{38,39}. To this end, the patient cohort was partitioned into ten folds, and the analytical pipeline was iterated ten times, each time training on nine folds and testing on the held-out fold. The pipeline utilized univariable Cox regression analysis to identify genes with P values below 0.05. Further evaluation of significant genes was conducted using the least absolute shrinkage and selection operator (LASSO) model, which identifies the most stable predictors. LASSO applies a penalty that improves model accuracy by shrinking some coefficients to zero, thereby eliminating non-significant predictors and resulting in a model that retains only a subset of predictors with nonzero coefficients. The optimal penalty parameter (λ) was determined through tenfold cross-validation, based on the minimum partial likelihood deviance. LASSO analysis was performed using the *glmnet* package. The key genes identified from the LASSO model were then used to construct the risk assessment model, with risk scores calculated by formula (1):

$$Rikscore = \sum_{k=1}^m \beta_k * E_k \quad (1)$$

In this context, m represents the total number of genes identified by the LASSO model, β_k denotes the risk coefficient, and E_k indicates the expression level of each gene included in the risk model. Patients were then stratified into high-risk and low-risk groups based on the median risk score as the cutoff. Kaplan–Meier survival analysis was performed via the *survminer* package⁴⁰ to compare the survival rates between these groups, and the log-rank test was used for statistical evaluation. A time-ROC analysis was conducted to assess the accuracy of the risk model predictions. The area under the curve (AUC) values were computed at 12, 36, and 60 months following PTC diagnosis using the *timeROC* package⁴¹. The aggregated out-of-sample predictions were used to provide an unbiased assessment of the model's performance and confirm the signature's stability. After addressing the risk of overfitting, the entire patient cohort was utilized to develop a gene prognostic signature.

Identification of independent prognostic factors and the construction of a nomogram

Cox proportional hazards regression analysis was performed in two phases to identify independent prognostic factors for PTC. Initially, univariable Cox regression was conducted on a range of clinicopathological variables, including patient age, sex, TNM staging, tumor size, nodal invasion, and the SERN model. Variables with a P value less than 0.05 in the univariable analysis were subsequently included in the multivariable Cox regression model. A nomogram was then developed using the results of the multivariable Cox regression to estimate overall survival (OS) rates at 12, 36, and 60 months post-PTC diagnosis. The nomogram's performance was assessed by calibration curves generated from 1,000 bootstrap resamples. Discriminatory ability was further evaluated using time-dependent ROC curve analysis. All nomogram construction and calibration analyses were performed using the *rms* package in R.

Evaluation of the diagnostic performance of genes

ROC curve analysis was conducted to evaluate the diagnostic performance of components within the ceRNA and DE-TF networks using RNA-seq data. A robust tenfold CV strategy was implemented via the *caret* package to reliably estimate the classification performance of each gene and reduce the risk of model overfitting. For each gene, the RNA-seq dataset was randomly partitioned into 10 equal subsets; nine were used for training and one for validation in each iteration. This procedure was repeated so that every subset served once as the validation set, and results from all ten test folds were aggregated using the *pROC* package⁴². Genes with an AUC greater than 0.6 were selected as candidates. A pooled ROC curve was created, and AUC values with 95% confidence intervals (CIs) were calculated using the DeLong method. The optimal classification threshold was established by maximizing the Youden index ($J = \text{Sensitivity} + \text{Specificity} - 1$). Sensitivity and specificity metrics were also reported for each gene. To experimentally validate these findings, ROC curve analysis using RT-qPCR data from fresh tissue samples was performed for three selected ceRNA network components: *GAS5*, *PKMYT1*, and *circMET*. This validation cohort comprised paired samples independent from those used in the RNA-seq dataset. ΔCt values were used as predictors to assess each gene's discriminatory power between PTC and normal tissues. AUC values and their 95% CIs were estimated via the non-parametric trapezoidal rule and DeLong method, respectively. Due to the limited sample size in this cohort, cross-validation was not applied. All statistical analyses and graphical representations were carried out using GraphPad Prism (version 10.2.3.402; GraphPad Software, San Diego, CA, USA). Results with P values less than 0.05 were considered statistically significant.

Creation of a multi-gene diagnostic panel

A multi-gene panel was developed to assess whether combining multiple genes could improve diagnostic accuracy. This was achieved by augmenting the previously described tenfold CV framework with a nested feature selection process. Within each iteration, a backward stepwise logistic regression algorithm was applied to the training set (nine folds) to identify an optimal subset of predictive genes. The resulting multi-gene model was then used to generate predictions on the corresponding held-out test fold. The overall performance of the final multi-gene panel was determined by aggregating the out-of-sample predictions across all folds. The optimal

diagnostic threshold was selected by maximizing the Youden index. Sensitivity and specificity at this cutoff were reported. AUC values with 95% CIs were calculated using the DeLong method. Statistical significance was set at $P < 0.05$.

Ethics statement

Informed consent was obtained from all participants and their legal guardians. All subjects were fully informed of the study's objectives, methods, funding sources, the institutional affiliations of the researchers, anticipated benefits, potential risks, and arrangements for post-study care. Participants were also notified of their right to withdraw from the study at any time. The study protocol was approved by the Iran National Committee for Ethics in Biomedical Research and the Ethics Committee of the Research Institute for Endocrine Sciences, Isfahan University of Medical Sciences, Isfahan, Iran (IR.UI.REC.1403.006). All experiments involving human subjects were conducted in compliance with the principles of the Declaration of Helsinki (7th edition) and its subsequent amendments.

Tissue acquisition and pathological affirmation

Thirty paired samples were collected from patients undergoing thyroidectomy at Sina Hospital, Isfahan Province, Iran. Samples were strategically selected to encompass a broad range of clinicopathological characteristics, thereby ensuring that the cohort represented the spectrum of disease observed in the broader patient population. The study included 16 low-aggressive PTC tissues (stage I), 14 high-aggressive PTC tissues (stages II, III, and IV), and 30 corresponding normal tissues adjacent to PTC tumors, providing a balanced representation of low- and high-aggressive PTC cases and effectively capturing clinical heterogeneity. Inclusion criteria required a confirmed histopathological diagnosis of PTC, sufficient tissue for molecular analysis, and no history of chemotherapy or radiotherapy. Fresh tissue samples were immediately submerged in RNeasy Lysis Reagent (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Samples were incubated overnight at 4 °C, then rinsed with normal saline to remove residual RNeasy Lysis Reagent, and stored at -70 °C until total RNA extraction.

Tissue pathology

Histopathological examination of the paraffin-embedded tissues was conducted by a special pathologist. Tumor staging was performed via the 7th edition of the American Joint Committee on Cancer tumor-node-metastasis (TNM) staging system. Patient information and tumor pathological characteristics are detailed in Table S2.

Blood biochemistry

Venous blood samples were collected from PTC patients prior to surgery. Blood calcium levels were measured using a calcium diagnostic kit (Pars-Azmoun Company, Tehran, Iran) according to the manufacturer's protocol. Calcium concentrations were determined by colorimetric assay using a biochemical autoanalyzer (BT 1500, Biotecnica, Italy). Based on these measurements, PTC tissues were categorized into two groups: a low-calcium group (LCG) and a normal-calcium group (NCG), with classification criteria based on clinical reference ranges established by the American Association for Clinical Chemistry. Specifically, NCG was defined as blood calcium levels between 8.5 and 10.5 mg/dL, while LCG was defined as levels below 8.5 mg/dL. To ensure accuracy and reproducibility, each sample was analyzed in duplicate, and internal quality control samples were included in each assay run. The coefficient of variation for the calcium assay was maintained below 3%. Comparative analysis of mean expression levels for *GAS5*, *PKMYT1*, and *circMET* genes was performed between the LCG/NCG PTC tissues and their matched adjacent normal tissues using Student's t-test. Additional analyses were conducted to compare gene expression between LCG and NCG PTC tissues.

RNA extraction and quality control

Total RNA was extracted from RNeasy Lysis Reagent-treated tissues using a one-step RNA extraction reagent (BIOBASIC, Canada) in accordance with the manufacturer's instructions. The integrity and purity of the extracted RNAs were evaluated via agarose gel electrophoresis and spectrophotometry. RNA fragmentation was assessed via gel electrophoresis. Protein and salt/phenol contaminations were identified by calculating the absorbance ratios of 260/280 and 260/230, respectively.

Primer Designing

Primers targeting mature miR-3671 were designed using the stem-loop methodology⁴³. Divergent primers specific for *circMET* (hsa_circ_0082002) were generated with the CircInteractome web tool⁴⁴, and their specificity was further assessed using circPrimer2⁴⁵. Primers for *GAS5*, *PKMYT1*, *SYMPK*, and *RNAU6* were designed using Beacon Designer 8.1 (Premier Biosoft International, Palo Alto, CA, USA). All primers were subsequently evaluated with Oligo7 software (Molecular Biology Insights, Colorado Springs, USA) to minimize the potential for secondary structures, such as primer dimers, stem loops, and hairpins. Additionally, the NCBI Primer-BLAST tool was used to confirm the specificity of each primer at the genomic level. *RNU6* was utilized as a housekeeping gene for miRNA analysis, whereas *SYMPK* functioned as a housekeeping gene for circRNAs, lncRNAs, and mRNAs during normalization. Table S3 summarizes the sequences of all primers.

cDNA synthesis

Total RNA (1,000 ng) was reverse transcribed using the AddScript cDNA Synthesis Kit (AddBio, Canada) with random hexamers. For miR-3671 cDNA synthesis, specific RT primers were first heated to 85 °C for 5 min and then cooled prior to reverse transcription. For *circMET* cDNA synthesis, a branched rolling-circle amplification approach was used, employing a primer spanning the backsplice junction⁴⁶.

Reverse transcription quantitative polymerase chain reaction (RT-qPCR)

Expression levels of ceRNA network genes were analyzed using the Rotor-Gene platform and SYBR Green qPCR Master Mix (Amplicon, Denmark). All RT-qPCR experiments were conducted in duplicate. Relative gene expression was calculated using the $2^{-\Delta\Delta Ct}$ method⁴⁷. Analyses were performed to compare gene expression between low-aggressive tumors (stage I) and normal tissues, high-aggressive tumors (stages II, III, and IV) and normal tissues, as well as between low- and high-aggressive tumors. Statistical significance of expression differences was assessed using REST 2009 v2.0.13 software⁴⁸. Boxplots visualizing gene expression levels across groups were generated using GraphPad Prism.

Experimental validation of the most influential component of the ceRNA network

Experimental validation of the in silico-constructed ceRNA network was performed using RT-qPCR on paired clinical PTC tissues and their adjacent normal counterparts. The analysis examined elements of the five-layer ceRNA network axis, specifically the central hub gene *PKMYT*, its regulator *mir3671*, and the miR sponges *GAS5* lncRNA and *circMET*. *LINC02910* lncRNA and the *FOLH1B* pseudogene were excluded due to undetectable expression in fresh tissues. For RT-qPCR data, the Student's t-test was used for pairwise group comparisons, preceded by assessment of data normality using the Shapiro–Wilk test. Homogeneity of variance was evaluated using Levene's test. Three types of pairwise comparisons were performed: high-aggressive PTC tissues versus adjacent normal tissues, low-aggressive PTC tumors versus adjacent normal tissues, and high-aggressive versus low-aggressive PTC tumors.

Results

Differentially expressed RNA molecules were identified

We identified DEmRNAs, DELncs, DEmiRs, and DEpseudogenes in four stages of PTC by analyzing RNAseq data (Figure S1). When stage I PTC tissues were compared with normal tissues, 1425 DEmRNAs, 28 DEmiRs, 4041 DELncs, and 1295 DEpseudogenes were identified. Compared with normal tissue, stage II PTC contained 903 DEmRNAs, 22 DEmiRs, 3066 DELncs, and 968 DEpseudogenes. Differential expression analysis between stage III PTC and normal tissues identified 1590 DEmRNAs, 27 DEmiRs, 4065 DELncs, and 1364 DEpseudogenes. Analysis of stage IV PTC tissues revealed 1815 DEmRNAs, 26 DEmiRs, 3947 DELncs, and 1239 DEpseudogenes that were differentially expressed relative to normal tissues. Four stages were overlapped to identify commonly expressed molecules, resulting in the identification of 771 DEmRNAs (Figure S1a), 19 DEmiRs (Figure S1b), 2593 DELncs (Figure S1c), and 751 DEpseudogenes (Figure S1d).

Twenty-three hub genes were recognized

Utilizing IVI, MCC, and MCODE, twenty-three hub genes were identified (Figure S2). A PPI network was constructed using 771 DEmRNAs (Figure S2a), with 661 nodes and 2148 edges. A total of 55 candidate genes were overlapped by the MCC and IVI methods, representing the top ten percent of nodes (Figure S2b). MCODE identified 25 clusters, with cluster 1 attaining the highest score of 19.818 (Table S4). MCODE-cluster 1 consists of 23 nodes, or genes, all situated within the overlapping region with candidate genes, identified as hub genes (Figure S2c). Hub genes exhibited greater interconnectivity compared to other candidate genes (Figure S2d).

Six of the 23 hub genes were validated via GEPIA

Six of the 23 hub genes were significantly upregulated in both the single and combined analyses (Figure S3). These genes included protein kinase membrane associated tyrosine/threonine 1 (*PKMYT1*, Figures S3a–b), E2F transcription factor 1 (*E2F1*, Figures S3c–d), thymidine kinase 1 (*TK1*, Figures S3e–f), antisilencing function 1B histone chaperone (*ASF1B*, Figures S3g–h), chromatin licensing and DNA replication factor 1 (*CDT1*, Figures S3i–j), and RecQ-mediated genome instability 2 (*RMI2*, Figures S3k–l).

GEPIA-validated hub genes were reinforced by microarray metadata analysis

Quality control of microarray metadata was done (Figure S4). Sample distributions before and after batch effect removal are shown in boxplots (Figures S4a and S4b, respectively) and PCA plots (Figures S4c and S4d, respectively). After removing outliers and addressing batch effects, a total of 267 samples remained, comprising 86 normal samples, 103 PTC samples, and 78 samples representing other THCA subtypes. Overall, 4,669 genes were found to be upregulated in PTC compared to normal tissues adjacent to PTC tumors. Two datasets, GSE27155 and GSE9115, were excluded from the microarray meta-analysis due to concerns regarding data quality and redundancy.

RNA-seq data were used to investigate the expression of hub genes across different PTC stages, while the microarray meta-analysis focused on PTC tissues without detailed staging information (Table 1). RNA-seq analysis revealed significant overexpression of GEPIA-validated hub genes across all four PTC stages relative to normal tissues. Importantly, all six hub genes validated by GEPIA were also identified as DEGs in the microarray meta-analysis, independent of detailed pathological annotation.

The expression of hub genes was not influenced by the sex or age of the patients

The influence of sex on hub gene expression is illustrated in Figure S5. Hub genes were consistently identified as DEGs in PTC tissues compared to normal tissues, irrespective of sex (Figure S5a). Both male and female PTC tissues exhibited significant differential expression of hub genes relative to matched normal tissues (Figures S5b and S5c, respectively). However, no significant differences in hub gene expression were observed when directly comparing PTC tissues from female versus male patients (Figure S5d).

Similarly, no differences in hub gene expression were found across PTC tissues from different age groups (Figure S6), indicating that age does not significantly impact the expression levels of hub genes.

Genes	RNAseq data								Microarray metadata	
	I - N		II - N		III - N		IV - N		P - N	
	logFC	FDR	logFC	FDR	logFC	FDR	logFC	FDR	logFC	FDR
PKMYT1	1.8	9.3e-26	1.4	4.9e-10	1.9	9.3e-24	2.1	2e-23	0.21	1.8e-08
RMI2	1.8	5.6e-27	1.5	6.9e-13	1.8	1.5e-24	1.9	4.2e-22	1.02	3.7e-24
ASF1B	1.9	1.3e-25	1.5	1.9e-10	1.9	2.3e-23	2	4.4e-21	0.58	2.3e-12
CDT1	1.9	1.5e-26	1.7	6.5e-14	2	3.4e-25	2.2	2.3e-23	0.49	3.7e-14
E2F1	2.2	4.4e-32	2.1	2.5e-20	2.3	1.2e-30	2.4	5.9e-26	0.3	7.1e-13
TK1	1.8	1.6e-22	1.6	3.3e-12	1.9	1.2e-21	2	1e-19	0.62	9.6e-13

Table 1. Assessment of hub gene expression through the analysis of RNAseq data and microarray metadata. Hub genes were differentially expressed according to microarray metadata analysis. Hub genes also showed differential expression at stages I to IV after RNAseq data analysis. Microarray datasets were retrieved from the NCBI-GEO public repository and RNAseq data from the publicly available TCGA database. N: normal; I: stage I; II: stage II; III: stage III; IV: stage IV; P: papillary thyroid cancer; logFC: log2-fold change; FDR: false discovery rate; *PKMYT1*: protein kinase membrane-associated tyrosine/threonine 1; *RMI2*: RecQ-Mediated Genome Instability 2. *ASF1B*: anti-silencing function 1B histone chaperone, *CDT1*: chromatin licensing and DNA replication factor 1, *TK1*: thymidine kinase 1, *E2F1*: E2F transcription factor 1.

PKMYT1 and RMI2 were the hub genes used for the construction of the ceRNA network

A ceRNA network was constructed centered on two hub genes: *PKMYT1* and *RMI2* (Fig. 2). The *PKMYT1* axis consists of five layers: *PKMYT1*, miRNAs, lncRNAs, circRNAs, and pseudogenes. The *RMI2* axis comprises three layers: *RMI2*, miRNAs, and circRNAs.

Within the *PKMYT1* axis, *miR-3671* emerged as a key regulator (Fig. 2a). RNAInter predicted six lncRNAs interacting with *miR-3671*, three of which overlapped with DElncs (Fig. 2b). They include growth arrest-specific 5 (*GAS5*), ArfGAP with a GTPase domain, ankyrin repeat, and PH domain 11 (*AGAP11*), and long intergenic nonprotein-coding RNA 2910 (*LINC02910*). Neither StarBase nor miRcode identified additional lncRNAs. Two pseudogenes (folate hydrolase 1B: *FOLH1B* and NBPF member 8: *NBPF8*) were predicted via RNAInter, with *FOLH1B* being most prominent, overlapping with 751 DEpseudogenes (Fig. 2c).

For the *RMI2* axis, *miR-616-5p* was identified as a common regulator overlapping between predicted miRNAs and DEmiRs (Fig. 2d). Among eight lncRNAs predicted by RNAInter, only *TSPEAR* antisense RNA 2 (*TSPEAR-AS2*) overlapped with DElncs (Fig. 2e). No common pseudogene sponges were identified for *miR-616-5p* (Fig. 2f).

Correlation analysis was performed to finalize network components (Fig. 2g), based on the ceRNA hypothesis: miRNAs should be negatively correlated with target hub genes, while miRNA sponges should exhibit positive correlations. The final *PKMYT1* axis comprised *miR-3671* and four sponges (*GAS5*, *LINC02910*, *AGAP11*, and *FOLH1B*), while the *RMI2* axis included *miR-616-5p* and one sponge (*TSPEAR-AS2*). However, *AGAP11* and *TSPEAR-AS2* were excluded due to negative correlations with their respective hub genes.

Regarding circRNAs, eleven DEcircs were shown to interact with *miR-3671*, with four novel circRNAs (Table S5), notably *novel_circ_011147*, which displayed the strongest binding affinity. Among known DEcircs, *hsa_circ_0003571* had the highest interaction energy (-13.51 kcal/mol), while *hsa_circ_001640* presented the most interaction sites with *miR-3671*. Seventeen DEcircs were found to interact with *miR-616-5p*, including four novel circRNAs (Table S6). *Novel_circ_010456* and *hsa_circ_000013* exhibited the strongest affinity (-8.81 kcal/mol and -12.54 kcal/mol), while *hsa_circ_000609* had the most interaction sites with *miR-616-5p*.

A ceRNA network containing two hub genes (*PKMYT1* and *RMI2*), two miRNAs (*miR-3671* and *miR-616-5p*), two lncRNAs (*GAS5* and *LINC02910*), twenty-three circRNAs, and a pseudogene (*FOLH1B*) was constructed (Fig. 2h). *Novel_circ_004588* and *novel_circ_013288*, *circ_0007444*, *circ_001209*, and *hsa_circ_000609* were found to have binding sites for both *miR-3671* and *miR-616-5p*.

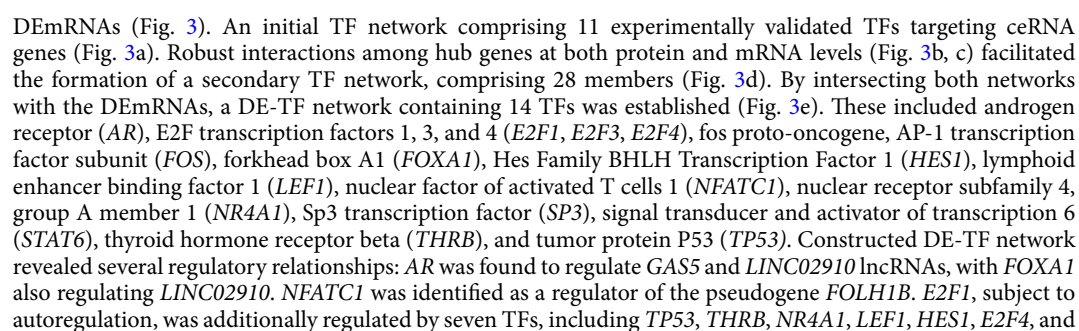
Validation analyses of ceRNA network genes (excluding circRNAs) were completed both with and without PTC tumor stage annotations (Table 2). All network components were identified as DEGs across various PTC stages and in comparisons of normal versus cancerous tissues, regardless of pathological classification. Other hub genes did not meet the miRNA selection criteria (Figures S7a-d).

Deregulation of ceRNA network genes was observed in the pan-cancer analysis

The expression levels of ceRNA network genes across various cancer types are shown in Figure S8. *PKMYT1* was significantly overexpressed in tumors of all types compared to adjacent normal tissues (Figure S8a). *GAS5* also demonstrated significant upregulation in multiple cancers, including THCA (Figure S8b). *RMI2* was consistently overexpressed across all cancer types evaluated, including THCA (Figure S8c). Both *FOHL1B* and *LINC02910* exhibited significantly increased expression in THCA and several other cancers (Figures S8d and S8e, respectively).

Fourteen TFs were incorporated into the DE-TF network construction

To identify potential regulators, two types of TF networks were constructed for ceRNA components and hub genes followed by the development of a DE-TF network incorporating those regulators identified within



◀**Fig. 2.** Construction of the ceRNA network. (a) *miR-3671* is a microRNA that targets the *PKMYT1* hub gene, as indicated in both the RNAInter and miRDB databases. It is also identified as a DE miRs. (b) *AGAP11*, *LINC02910*, and *GAS5* are three lncRNAs identified by RNAInter as potential sponges for *miR-3671*, thereby influencing the regulation of *PKMYT1*. These lncRNAs are also present among DE lncs. (c) *FOLH1B* is identified by RNAInter as a pseudogene that can act as a sponge for *miR-3671*. It also exists among DE pseudogenes. (d) *miR-616-5p* is a microRNA that targets the *RMI2* hub gene, as identified in the RNAInter, miRWalk, and miRDB databases. It is also present among DE miRs. (e) *TSPEAR-AS2* is an lncRNA proposed by RNAInter to act as a sponge for *miR-616-5p* and regulate *RMI2*. This lncRNA is also present among DE lncs. (f) RNAInter-predicted pseudogenes do not overlap with DE pseudogenes, indicating that no pseudogene is currently identified as a sponge for *miR-616-5p*. (g) *GAS5*, *LINC02910*, and *FOLH1B* demonstrate positive correlations with *PKMYT1*, whereas *PKMYT1* exhibits a negative correlation with *miR-3671*. *RMI2* exhibits negative correlations with *miR-616-5p* and *TSPEAR-AS2*. (h) A ceRNA network is observed to consist of two hub genes (*PKMYT1* and *RMI2*), two miRNAs (*miR-3671* and *miR-616-5p*), two lncRNAs (*GAS5* and *LINC02910*), twenty-three circRNAs, and one pseudogene (*FOLH1B*). Red circles denote circRNAs, orange arrows indicate pseudogenes, blue ovals represent miRNAs, brown hexagons signify lncRNAs, and light brown squares correspond to mRNA. ceRNA: competitive endogenous RNA; DE miRs: differentially expressed microRNAs; DE lncs: differentially expressed long noncoding RNAs; DE pseudogenes: differentially expressed pseudogenes; *PKMYT1*: protein kinase membrane associated tyrosine/threonine 1; *RMI2*: recQ mediated genome instability 2; *GAS5*: growth arrest specific 5; *AGAP11*: ArfGAP with a GTPase domain, ankyrin repeat, and PH domain 11; *LINC02910*: long intergenic nonprotein-coding RNA 2910; *FOLH1B*: folate hydrolase 1B; *miR-3671*: microRNA 3671; *miR-616-5p*: microRNA 616-5p; *TSPEAR-AS2*: TSPEAR antisense RNA 2.

Genes	Considering the stage of the PTC tumor				Excluding staging
	I – N	II – N	III – N	IV – N	P – N
	logFC (FDR)	logFC (FDR)	logFC (FDR)	logFC (FDR)	logFC (FDR)
<i>PKMYT1</i>	1.8 (9.3e-26)	1.4 (4.9e-10)	1.9 (9.3e-24)	2.1 (2e-23)	1.8 (7.3e-25)
<i>RMI2</i>	1.8 (5.6e-27)	1.5 (6.9e-13)	1.8 (1.5e-24)	1.9 (4.2e-22)	1.8 (1.7e-27)
<i>GAS5</i>	0.71 (3.9e-14)	0.49 (0.0001)	0.93 (2.3e-18)	1.06 (1.0e-17)	0.72 (1.4e-14)
<i>LINC02910</i>	0.9 (3.7e-09)	0.7 (0.0007)	0.74 (1.7e-05)	0.55 (0.006)	0.9 (2.2e-09)
<i>FOLH1B</i>	0.97 (0.0004)	1.14 (0.0017)	1.07 (0.0011)	0.98 (0.006)	0.96 (0.0005)
<i>miR-3671</i>	– 0.91 (5.8e-09)	– 0.94 (0.0006)	– 1.09 (2.7e-09)	– 1.3 (1.2e-09)	– 0.87 (1.2e-08)
<i>miR-616-5p</i>	– 1.08 (1.7e-08)	– 1.21 (1.1e-07)	– 1.7 (2.09e-13)	– 1.12 (2.8e-05)	– 1.07 (4.6e-08)

Table 2. RNAseq data analysis in the ceRNA network with and without tumor staging. A comparative gene expression analysis using RNA-seq data is presented for essential ceRNA network components in PTC, excluding circRNAs. In the analysis, tumor tissues are compared to normal tissues irrespective of staging characteristics, as well as comparing tumors at various stages to normal tissues. P: papillary thyroid cancer; N: normal; I: stage I; II: stage II; III: stage III; IV: stage IV; logFC: log2 fold change; FDR: false discovery rate; *PKMYT1*: protein kinase membrane associated tyrosine/threonine 1; *RMI2*: recQ mediated genome instability 2; *GAS5*: growth arrest specific 5; *LINC02910*: long intergenic nonprotein-coding RNA 2910; *FOLH1B*: folate hydrolase 1B; *miR-3671*: microRNA 3671; *miR-616-5p*: microRNA 616-5p.

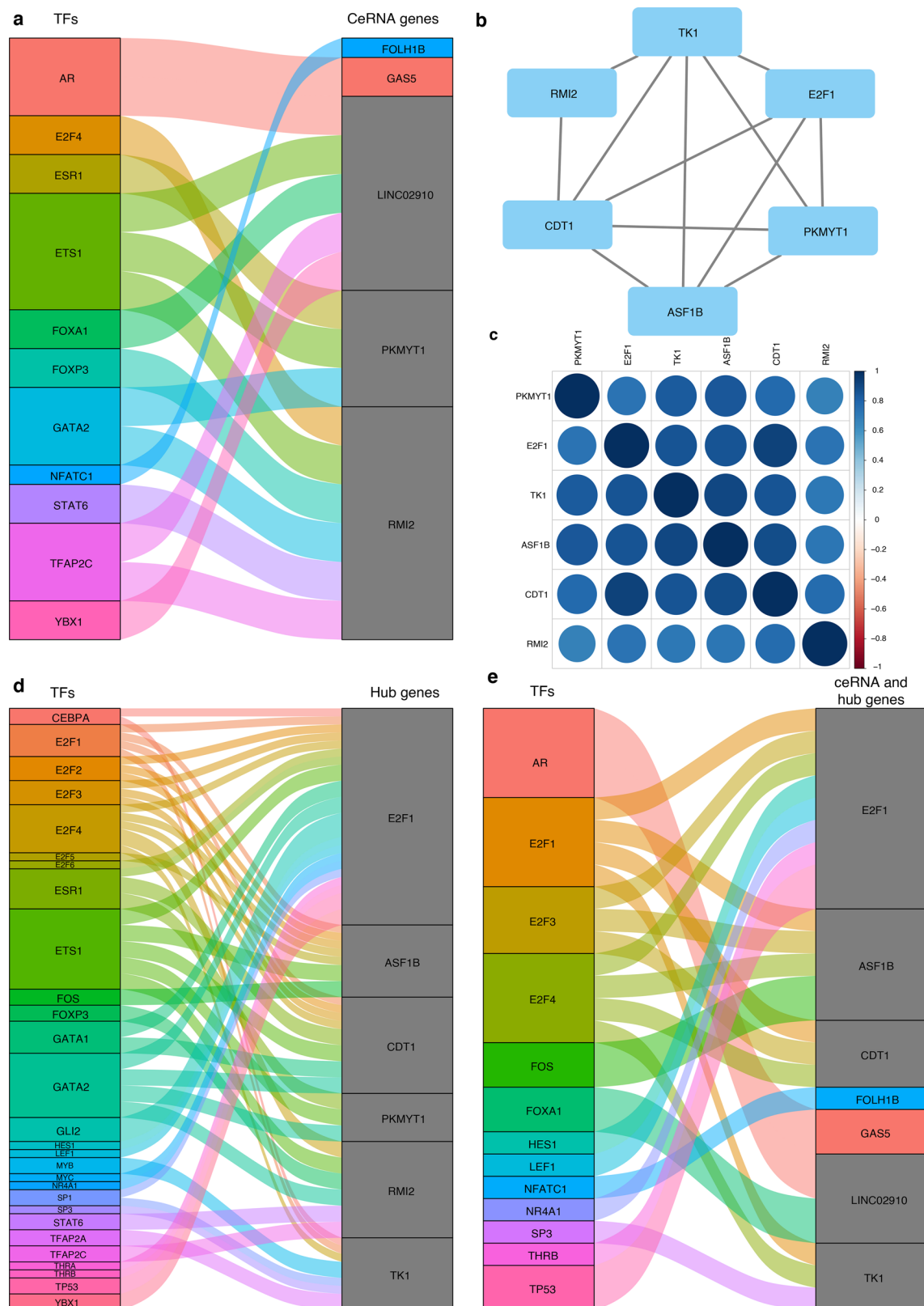
E2F3, *ASF1B* was regulated by four TFs: *E2F1*, *E2F3*, *E2F4*, and *FOS*. Both *CDT1* and *TK1* were under the control of *E2F1* and *E2F4*, respectively. *STAT6* was identified as a regulator of *RMI2*.

RNA-seq data revealed that *GAS5*, *AR*, and *ASF1B* were expressed differently during PTC stages

Stage-wise expression plots for one ceRNA network component and two DE-TF network components are shown in Figure S9. *GAS5* expression (Figure S9a) was found to increase significantly with tumor aggressiveness ($\text{Pr}(> F) = 3.11 \times 10^{-5}$). *ASF1B* and *AR* also exhibited stage-dependent differential expression in PTC (Figures S9b and S9c, respectively). In contrast, other DE-TF network components demonstrated uniform expression levels across different PTC stages (data not shown).

A PTC prognostic model was developed containing four genes

A prognostic gene signature for PTC was developed and validated through the integration of ceRNA and DE-TF network components (Fig. 4). A tenfold cross-validation approach employing univariate Cox regression and LASSO regression identified 9 of 24 genes within the ceRNA and DE-TF networks (Table S7). Four genes—



STAT6, E2F1, RMI2, and NR4A1—emerged as the most stable and consistently significant predictors of PTC survival. The Kaplan–Meier (KM) plot demonstrated a reduced OS rate for high-risk PTC relative to low-risk PTC (Fig. 4a, log-rank $p = 0.039$). Time-dependent ROC analysis revealed AUC values of 0.60 at 12 months, 0.75 at 36 months, and 0.58 at 60 months (Fig. 4b).

After internal validation of the risk model, the complete patient cohort was evaluated to establish a gene prognostic signature. Out of 24 genes within ceRNA and DE-TF networks, five candidates (*RMI2*, *E2F1*, *ASF1B*, *NR4A1*, and *STAT6*) were identified as prognostically significant through univariable Cox regression analysis

◀ **Fig. 3.** Development of the DE-TF network. (a) An initial TF network identifies 11 experimentally validated transcription factors that target ceRNA genes. (b) The protein–protein interaction network illustrates the connections between two ceRNA components, *PKMYT1* and *RMI2*, and other hub genes at the protein level. (c) Two ceRNA components, *PKMYT1* and *RMI2*, exhibit strong correlations with other hub genes at the mRNA level. (d) A second TF network identifies 28 experimentally validated transcription factors that regulate hub gene expression. (e) A DE-TF network consists of 14 transcription factors that are included among the DEmRNAs. DE-TF: differentially expressed transcription factor; TFs: transcription factors; DEmRNAs: differentially expressed messenger RNAs; ceRNA: competitive endogenous RNA; *PKMYT1*: protein kinase membrane associated tyrosine/threonine 1; *RMI2*: recQ mediated genome instability 2.

(Fig. 4c). To determine the optimal gene combination, LASSO regression was applied to these five genes, resulting in the exclusion of *ASF1B* from further analysis (Fig. 4d, e). A risk score model—termed the SERN model—was subsequently developed, comprising *STAT6*, *E2F1*, *RMI2*, and *NR4A1*. Patients in the low-risk group demonstrated significantly better OS than those in the high-risk group (Fig. 4f). A strong association was observed between higher risk scores and increased mortality. In the risk formula (Eq. 1), negative coefficients for *STAT6* (−1.32), *E2F1* (−0.45), and *RMI2* (−0.05) indicated their protective roles, with higher expression in low-risk patients. Conversely, the positive coefficient for *NR4A1* (0.16) identified it as a risk gene, showing lower expression in the low-risk group (Fig. 4g). The SERN model demonstrated robust predictive accuracy, yielding AUC values of 80.43%, 84.82%, and 76.97% at 12, 36, and 60 months post-PTC diagnosis, respectively (Fig. 4h). The high sensitivity and specificity of these four-gene signatures confirmed the SERN model as a reliable predictor of OS in PTC patients.

Correlation analysis between the SERN model and clinicopathological features

Both univariable and multivariable Cox regression analyses were conducted to assess the independent prognostic value of the SERN model in PTC (Table 3). In univariable analysis, the SERN model was significantly associated with OS (HR=1.43). Other clinicopathological features, including age (HR=1.01), TNM stage (HR=2.67), and tumor size (HR=3.12), were also significantly associated with OS. Multivariable Cox analysis revealed that the SERN model retained a significantly increased hazard ratio (HR=1.78, $p<0.001$), with age remaining a significant factor (HR=1.01). TNM stage and tumor size were no longer significant predictors in the multivariable context ($p>0.05$). These findings indicate that the SERN model and patient age are independent prognostic factors for OS in PTC.

A nomogram was developed to estimate the probability of OS at 12, 36, and 60 months following PTC diagnosis (Fig. 5). This tool was constructed based on the independent prognostic factors identified in multivariable Cox regression analysis, specifically incorporating the SERN model risk score and patient age as key predictors (Fig. 5a). The discriminative performance of the nomogram was assessed using time-dependent ROC curves, demonstrating excellent predictive accuracy: the AUC was 95.93% for 1-year OS, 96.85% for 3-year OS, and 94.46% for 5-year OS (Fig. 5b). Calibration plots further confirmed strong agreement between the nomogram's predicted survival probabilities and actual observed outcomes at 1, 3, and 5 years post-diagnosis (Fig. 5c–e).

Components of the ceRNA and DE-TF networks have potential for PTC diagnosis

A comprehensive tenfold CV analysis demonstrated that all genes within the ceRNA and DE-TF networks possessed significant diagnostic capability for distinguishing PTC from normal tissue (Table S8). In the ceRNA network, *RMI2* emerged as the leading biomarker, achieving a cross-validated AUC of 0.88. Other network components, *GAS5* (AUC=0.86), *PKMYT1* (AUC=0.81), *miR-3671* (AUC=0.75), *LINC02910* (AUC=0.71), *miR-616-5p* (AUC=0.73), and *FOLH1B* (AUC=0.66), also displayed substantial diagnostic accuracy. DE-TF genes yielded comparably robust results, with *E2F1* showing particularly strong performance (AUC=0.92). Other prominent DE-TF biomarkers included *CDT1* (AUC=0.90), *STAT6* (AUC=0.89), and *TP53* (AUC=0.86), while the remaining DE-TF candidates were ranked by descending AUC values.

A gene panel including components of the ceRNA and DE-TF networks has enhanced diagnostic accuracy for PTC

A multi-gene panel was established to enhance diagnostic accuracy using backward stepwise logistic regression within a tenfold CV framework (Fig. 6). This process identified an optimal panel of eight genes: *PKMYT1*, *LINC02910*, *GAS5*, *E2F1*, *NFATC1*, *STAT6*, *E2F3*, and *TK1*. As the components of the ceRNA network, *GAS5* (AUC=0.861), *PKMYT1* (AUC=0.817), and *LINC02910* (AUC=0.715) exhibited substantial diagnostic accuracy (Figs. 6a–c). As the components of the DE-TF network, *E2F1* (AUC=0.929), *STAT6* (AUC=0.898), *TK1* (AUC=0.846), *NFATC1* (AUC=0.837), and *E2F3* (AUC=0.761) were identified as additional high-performing DEGs (Figs. 6d–h). The combined eight-gene panel exhibited a notable synergistic effect with enhanced diagnostic performance of 0.969 (Fig. 6i). At the optimal Youden index threshold, the panel yielded a sensitivity of 97.0% and a specificity of 91.5%.

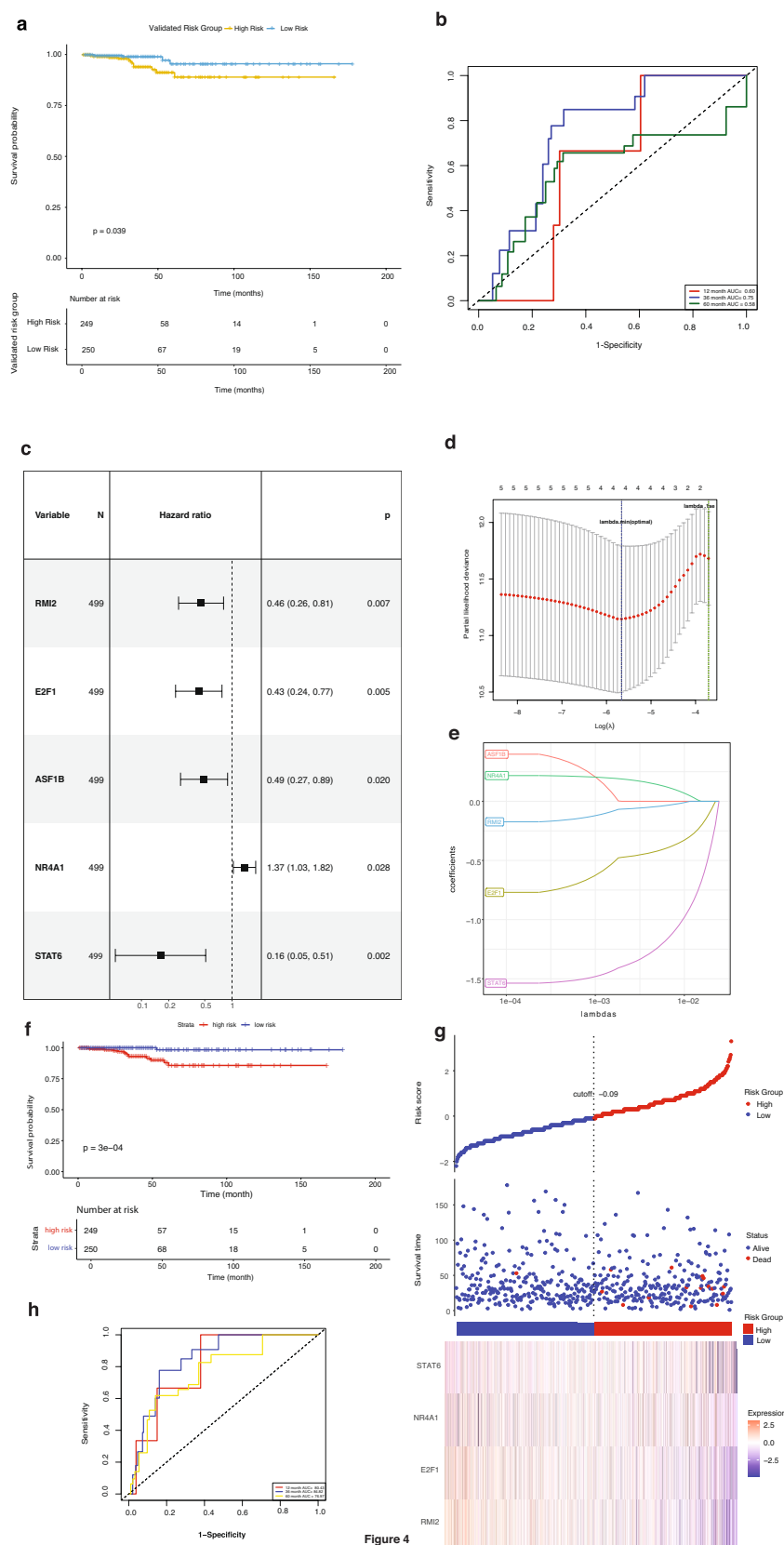


Figure 4

RT-qPCR revealed a stage-specific expression pattern for *PKMYT1* and *GAS5*

RNA-seq-derived hub genes were not identified as DEGs in the comparison of fresh PTC tissues with adjacent normal tissues, excluding considerations of tumor aggressiveness (data not shown). The initial discrepancy was resolved by categorizing PTC tissues according to their aggressiveness (Fig. 7). Both *PKMYT1* and *GAS5* were significantly downregulated in low-aggressive tumors compared to normal tissues (Fig. 7a, d, respectively). In

◀**Fig. 4.** Development and validation of a prognostic gene signature in PTC. Internal validation confirms the model’s predictive accuracy through (a) the KM plot and (b) the time-dependent ROC curve. (c) A forest plot presents the findings of univariable Cox regression analysis for the entire cohort, emphasizing five candidate genes. (d) The tenfold CV method is employed to identify the optimal tuning parameter (lambda) in the LASSO regression model. (e) The LASSO coefficient profiles of the five candidate genes identified a prognostic gene signature comprising four genes. (f) The KM plot indicates significantly different survival outcomes between the high-risk and low-risk groups within the entire cohort, as determined by the prognostic gene signature. (g) The upper panel presents the risk score and survival status for each sample. The lower panel displays the mRNA expression levels of the prognostic gene signature. (h) Time-dependent ROC curves confirm the prognostic model’s performance across the entire cohort at 12 months (red), 36 months (blue), and 60 months (yellow), with all AUC values surpassing 0.70. PTC: papillary thyroid cancer; KM: Kaplan-Meier; ROC: receiver operating characteristic; tenfold CV: tenfold cross validation; LASSO: least absolute shrinkage and selection operator; AUC: area under the curve.

Characteristic (N)	Univariable			Multivariable		
	HR	95% CI	p	HR	95% CI	p
SERN model (449)	1.43	1.17, 1.75	<0.001	1.78	1.31, 2.42	<0.001
Age (449)	1.01	1.01, 1.02	<0.001	1.01	1.01, 1.02	<0.001
TNM staging (449)	2.67	1.57, 4.53	<0.001	1.17	0.33, 4.14	0.8
Size of tumors (449)	3.12	1.55, 6.30	0.001	1.89	0.59, 6.07	0.3
Nodal invasion (449)	1.46	0.48, 4.47	0.5			
Sex (449)	1.98	0.64, 6.08	0.2			

Table 3. Univariable and multivariable Cox regression analysis of prognostic factors. The HR values for the SERN model (includes *STAT6*, *E2F1*, *RMI2*, and *NR4A1* genes) are presented. Univariable regression models for various clinicopathological features, including age, TNM staging, size of tumor, nodal invasion, and sex, are analyzed separately. Key features identified in univariable regression models are incorporated into multivariable regression analysis. The SERN model shows an enhanced hazard ratio of 1.78 ($P < 0.001$) in multivariable regression analysis, with age held constant ($HR = 1.01$). P values less than 0.05 were deemed statistically significant. N: number of individuals; HR: hazard ratio; 95% CI: 95% confidence interval; SERN: *STAT6*, *E2F1*, *RMI2*, and *NR4A1* genes.

contrast, these genes were significantly upregulated in high-aggressive PTC tissues relative to normal tissues (Figs. 7b, e, respectively). A direct comparison between high- and low-aggressive tumors revealed notable upregulation of *PKMYT1* and *GAS5* in the high-aggressive subgroup (Fig. 7c, f, respectively). Among circRNAs, *circMET* was selected for further experimental validation due to its highest logFC in the RNA-seq data (Table S9). *circMET* expression was significantly elevated in both low- and high-aggressive PTC tissues compared to adjacent normal tissues (Fig. 7g, h, respectively). However, no significant differential expression of *circMET* was observed between high- and low-aggressive tumors (Fig. 7i). For *miR-3671*, expression levels remained consistent across all PTC subgroups and normal tissues (data not shown).

***PKMYT1*, *GAS5*, and *circMET* demonstrated diagnostic capabilities in clinical specimens**

Diagnostic performance analysis demonstrated that the expression of *PKMYT1*, *GAS5*, and *circMET* could effectively predict PTC incidence and progression (Fig. 8). *PKMYT1* exhibited strong diagnostic accuracy for distinguishing high-aggressive from low-aggressive PTC, with an AUC of 0.81 (Fig. 8a). *GAS5* also showed high diagnostic accuracy, with an AUC of 0.87 for differentiating high-aggressive from low-aggressive PTC tumors (Fig. 8b). *circMET* demonstrated robust capability to distinguish low-aggressive PTC from normal thyroid tissues, achieving an AUC of 0.81 (Fig. 8c).

Blood calcium levels influence the expression patterns of *GAS5* in PTC tissues

Differential expression of *GAS5* was observed in PTC patients based on their blood calcium levels (Fig. 9). In PTC tissues from the LCG, *GAS5* was significantly downregulated compared to adjacent normal tissues (relative expression = 0.38, $p = 0.02$; Fig. 9a). In contrast, PTC tissues from the NCG exhibited significant upregulation of *GAS5* relative to normal tissues (relative expression = 3.01, $p = 0.01$; Fig. 9b). Moreover, direct comparison between LCG and NCG PTC tissues revealed a significant downregulation of *GAS5* in the LCG group (relative expression = 0.34, $p = 0.02$; Fig. 9c). These findings suggest a regulatory role for Ca^{2+} in modulating *GAS5* expression in PTC patients.

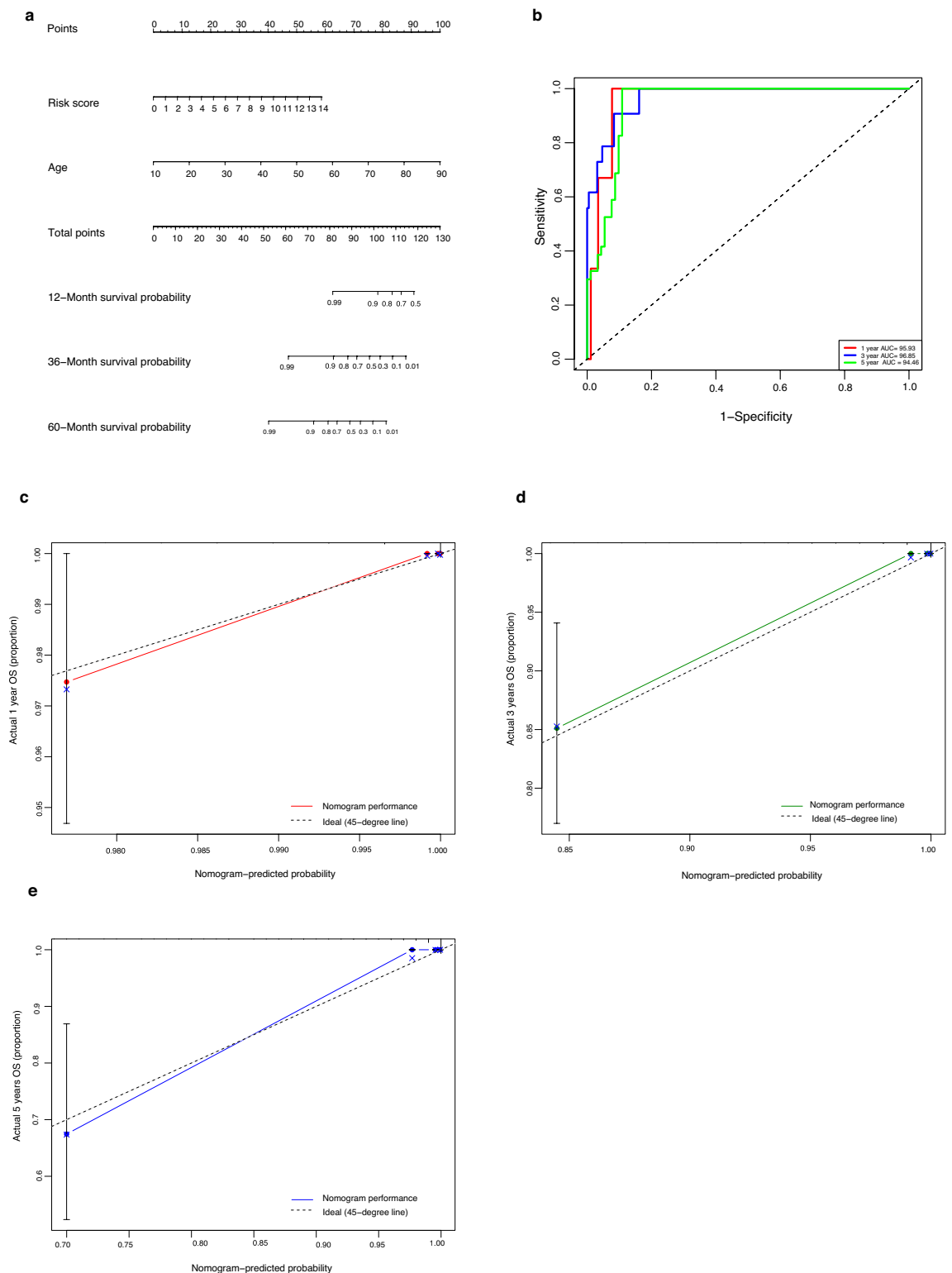


Fig. 5. Assessment of the Predictive Nomogram. **(a)** A nomogram predicts OS at 12, 36, and 60 months in patients with PTC. **(b)** Time-dependent ROC curves evaluate the nomogram's discrimination at 1, 3, and 5 years. Calibration curves show the nomogram's performance at **(c)** 1 year, **(d)** 3 years, and **(e)** 5 years. The x-axis denotes the survival probability predicted by the nomogram, while the y-axis indicates the actual observed survival rate. The solid line illustrates the nomogram's performance, while the vertical bars denote the 95%CI. The dashed 45-degree line indicates a perfectly calibrated model, wherein predicted outcomes align precisely with actual outcomes. PTC: papillary thyroid cancer; OS: overall survival; ROC: receiver operating characteristic; 95%CI: 95% confidence intervals.

Discussion

Numerous studies have highlighted the importance of ceRNAs in tumor development, emphasizing the interplay among lncRNAs, miRNAs, mRNAs, and circRNAs^{49–52}. This study represents the first report of a ceRNA network in PTC comprising five distinct RNA types: mRNAs, miRNAs, lncRNAs, pseudogenes, and circRNAs. The network benefitted from a stringent selection process for integrated hub genes, ensuring that their expression was not influenced by confounding variables such as patient sex or age. All components of the ceRNA network were dysregulated in PTC, exhibiting differential expression between cancerous and normal tissues—a pattern consistently observed across all PTC stages for every component except circRNAs. The identified ceRNA network features two regulatory axes (a five-layer axis and a three-layer axis), both of which are presented here for the first time in PTC.

Within the five-layer axis, *PKMYT1* emerged as a central oncogene, displaying overexpression in PTC as well as in various other cancers. This observation aligns with previous reports documenting *PKMYT1* upregulation in triple-negative breast, prostate, ovarian, oral, liver, and gastric cancers^{53–58}. Functional studies indicate that *PKMYT1* promotes tumorigenesis and cancer progression, such as enhancing growth in liver and ovarian cancer by inhibiting *GSK3β* (thereby activating the β -catenin/TCF pathway) and *SIRT3*^{55,56}. In summary, they established the oncogenic role of *PKMYT1* and its association with tumorigenesis and progression.

In our study, regulation of *PKMYT1* expression was attributed to *miR-3671* and its sponges, which included *FOLH1B*, *GAS5*, *LINC02910*, and eleven DEcircs. While the regulatory roles of *miR-3671* have been reported in gastric and breast cancers, its involvement in PTC has not been previously described^{59,60}. We found *miR-3671* to be downregulated, with concurrent upregulation of its sponges—*FOLH1B*, *GAS5*, *LINC02910*, and DEcircs—in PTC tissues. Notably, *FOLH1B* overexpression has, to date, only been associated with invasion and metastasis in prostate cancer^{61,62}. According to the ceRNA network hypothesis, miRNA sponges are typically coexpressed with their target mRNAs, which matches our observations^{63,64}.

There remain controversies regarding the roles of some lncRNAs identified. For instance, *GAS5* has been proposed as a tumor suppressor in PTC and other cancers^{65–68}, while *LINC02910* has primarily been reported as downregulated only in NSCLC⁶⁹. Gene polymorphisms can impact *GAS5*'s expression and cell-specific functionality⁷⁰; for example, rs55829688 has been shown to negatively regulate *GAS5* expression in colorectal cancer⁷¹. Our pan-cancer RNA-seq analysis reinforces our findings, revealing overexpression of both *GAS5* and *LINC02910* in PTC. These findings are consistent with those of previous studies of *GAS5* within the Iranian population⁷².

Among the eleven upregulated circRNAs identified as potent *miR-3671* sponges, four were novel. Several have previously been implicated as oncogenes—*circMET*⁷³, *circ_0003571*, and *circ_00060974*—influencing PTC metastasis to lymph nodes. We propose that the other circRNAs, not previously associated with PTC, may also serve as significant regulatory molecules due to their interaction with *miR-3671*.

Within the three-layer axis, *RM12* emerged as a central oncogene, exhibiting upregulation in PTC tumors as well as across a range of other cancer types. This finding is consistent with previous studies linking elevated *RM12* expression to tumor progression in cervical, lung, and prostate cancers⁷⁵. *RM12* has been implicated in tumorigenesis through various mechanisms, including disruption of the TP53 pathway⁷⁶ and activation of the PI3K/AKT pathway⁷⁷.

Our analysis identified *miR-616-5p*—which was found to be downregulated in PTC tumors—as a regulator of *RM12*. Prior research has shown that *miR-616-5p* plays a role in cancer initiation and progression, notably promoting ovarian and lung cancers by targeting the *GSK3β* pathway⁷⁸. Additionally, decreased *miR-616-5p* expression has been associated with increased cell viability and reduced apoptosis in gastric cancer, mediated by downregulation of *DUSP2*⁷⁹.

This study is the first to demonstrate the simultaneous dysregulation of *RM12*, *miR-616-5p*, *LINC02910*, *miR-3671*, and *FOLH1B* within a PTC-specific ceRNA network. Five out of twenty-three circRNAs identified were found to target both the *PKMYT1* and *RM12* axes, highlighting potential crosstalk between these regulatory pathways.

A DE-TF network was established to clarify the regulation of hub genes and ceRNA network components. Of particular interest is *E2F1*, which functions not only as a component of the DE-TF network but also as one of the hub genes. Remarkably, *E2F1* regulates both itself and other hub genes, with the exception of *PKMYT1*. As a self-regulating gene, *E2F1* is further controlled by seven different transcription factors. Consistent with our findings, the overexpression of *E2F1* has been documented in gastric, lung, and PTC^{80–83}. For example, binding of *E2F1* to the promoter of the lncRNA *RGMB-AS1* has been shown to facilitate PTC metastasis and invasion⁸⁴.

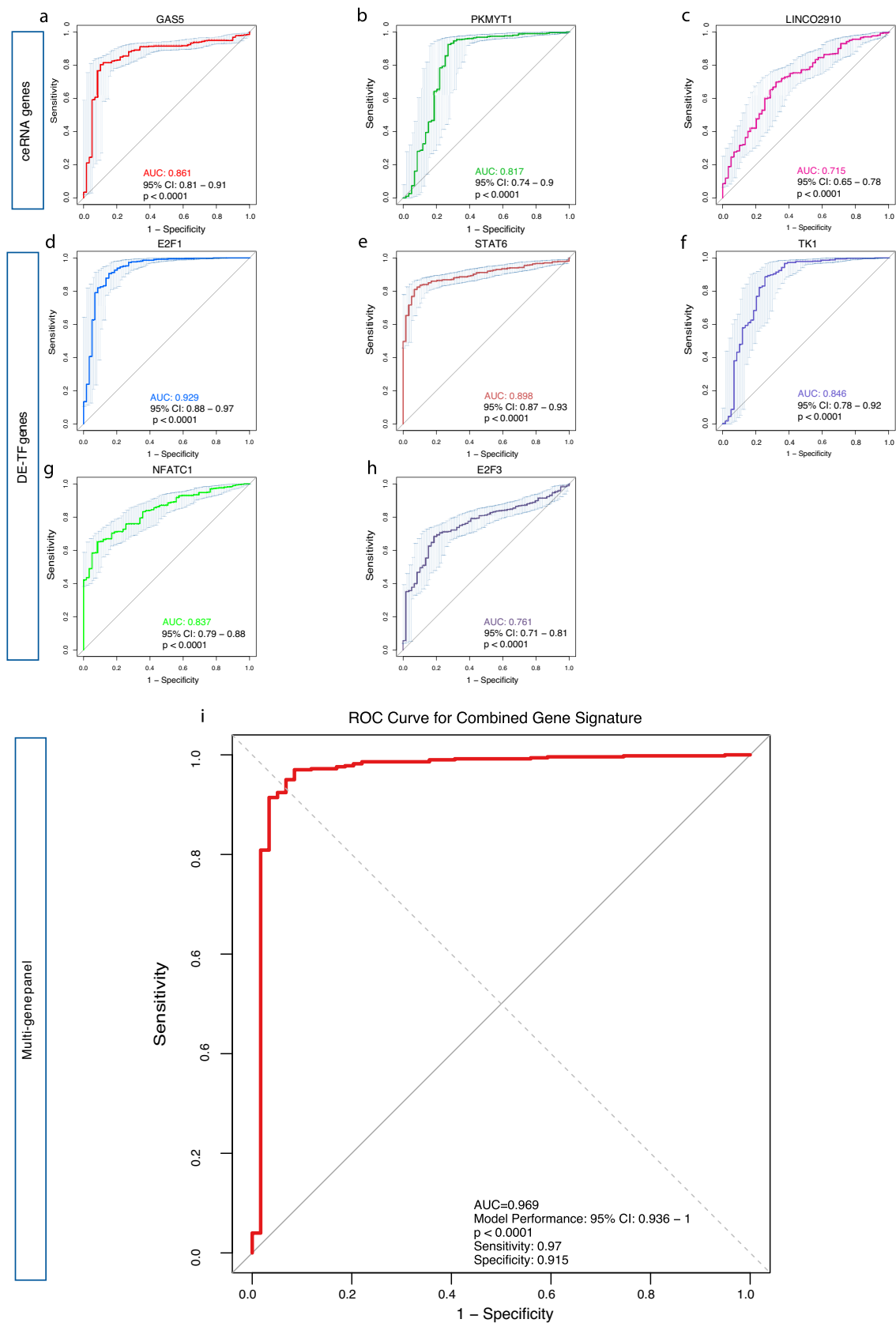


Fig. 6. Assessment of the effectiveness of PTC diagnosis utilizing a multi-gene panel. Backward stepwise logistic regression identifies an optimal eight-gene panel for diagnosing PTC, comprising *PKMYT1*, *E2F1*, *NFATC1*, *STAT6*, *E2F3*, *LINC02910*, *GAS5*, and *TK1*. (a–h) The ROC curves assess the diagnostic accuracy of each gene individually. (i) The multi-gene panel comprising eight genes demonstrates an enhanced AUC of 0.969. PTC: papillary thyroid cancer; ROC: receiver operating characteristic; AUC: Area under the curve; *PKMYT1*: protein kinase membrane associated tyrosine/threonine 1; *E2F1*: E2F transcription factors 1; *NFATC1*: nuclear factor of activated T cells 1; *STAT6*: signal transducer and activator of transcription 6; *E2F3*: E2F transcription factors 3; *LINC02910*: long intergenic nonprotein-coding RNA 2910; *GAS5*: growth arrest specific 5; *TK1*: thymidine kinase 1.

Beyond *E2F1*, *E2F4* was identified as the transcription factor regulating the largest number of ceRNA network components. This observation is supported by prior studies reporting *E2F4* upregulation in both liver and head and neck tumors, further implicating *E2F4* as a potential oncogene that promotes tumorigenesis and progression^{85,86}. *E2F4* participates in the regulation of critical cellular processes, including cell proliferation, cell cycle progression, and inhibition of apoptosis⁸⁵. Nevertheless, some studies have observed reduced *E2F4* expression in breast cancer tissues, suggesting that it may exert tumor-suppressive effects in specific contexts⁸⁷. Thus, the role of *E2F4* in cancer biology appears to be highly dependent on tumor type and cellular environment.

All genes included in the ceRNA and DE-TF networks demonstrated diagnostic potential for PTC, with AUC values ranging from 0.64 to 0.92. The integration of eight genes—*PKMYT1*, *LINC02910*, *GAS5*, *NFATC1*, *E2F3*, *STAT6*, *E2F1*, and *TK1*—further improved diagnostic accuracy, yielding a specificity of 96.9% for PTC detection.

Identifying robust gene signatures for predicting patient outcomes is fundamental to contemporary oncology. Recent studies have successfully identified *CIAPIN1* and *DBNDD1* as markers of survival in invasive breast carcinoma^{88,89}. Consistent with previous survival studies, our SERN prognostic model—comprising *STAT6*, *E2F1*, *RMI2*, and *NR4A1*—highlights the significance of these four genes as reliable prognostic markers for OS in patients with PTC. *STAT6*, *E2F1*, and *RMI2* had negative regression coefficients, identifying them as protective factors; higher expression levels were associated with lower risk scores and improved survival outcomes. In contrast, *NR4A1*, with its positive coefficient, served as a risk factor, where greater expression correlated with diminished survival prospects.

In our *in silico* analysis, we focused on the five-layer regulatory axis of the ceRNA network and validated its components using RT-qPCR in clinical samples. This work illuminated the complex regulatory roles of *PKMYT1* and *GAS5* in PTC, revealing their stage-specific expression patterns and highlighting potential diagnostic applications. For the first time, experimental validation via RT-qPCR demonstrated a dual expression pattern for *PKMYT1* and *GAS5*: both genes were downregulated in low-aggressive PTC tissues and significantly upregulated in high-aggressive PTC tissues. These findings suggest dynamic and complex regulatory roles for these genes during PTC progression. Additionally, *circMET* was found to be overexpressed in both low- and high-aggressive PTC tumors. A notable challenge in the RNA-seq data was the lack of pathological stage annotation, which hindered comprehensive analysis of circRNA expression across PTC stages. However, RT-qPCR analysis of fresh PTC tissues with known staging overcame this limitation, revealing that *circMET* expression was uniform across different PTC stages. Thus, in contrast to *PKMYT1* and *GAS5*, *circMET* does not appear to serve as a stage-specific marker for PTC. Overall, these results confirm that all ceRNA network components—including circRNAs—exhibited differential expression between PTC and normal tissues, with *PKMYT1* and *GAS5* specifically associated with tumor aggressiveness.

The integration of RT-qPCR and RNA-seq data yielded unexpected results. While *in silico* analyses identified *GAS5*, *circMET*, and *PKMYT1* as key modulators in PTC, initial RT-qPCR experiments did not detect significant deregulation of these genes when simply comparing PTC tissues to adjacent normal tissues. To resolve this discrepancy, we hypothesized that gene expression is correlated with PTC staging characteristics. Subsequent analyses confirmed distinct stage-specific expression patterns for these genes. Pairwise comparisons indicated that *GAS5* and *PKMYT1* exhibited differential expression between high-aggressive and low-aggressive tumors, as well as in comparisons of each group to normal tissues. In contrast, *circMET* did not show any significant difference between high-aggressive and low-aggressive tumors, although it was differentially expressed between tumor tissues (both high- and low-aggressive) and normal tissues. Rather than simply confirming RNA-seq predictions, our RT-qPCR results provided a significant enhancement, revealing a complex, stage-dependent expression profile not evident in the initial non-stratified analyses. Overall, these findings indicate that the dysregulation of these genes is closely linked to tumor aggressiveness in PTC.

Our results, corroborated by RNA-seq data analysis, highlight the significant potential of *PKMYT1*, *GAS5*, and *circMET* as valuable biomarkers for diagnosing and monitoring PTC. ROC curve analysis using RT-qPCR data demonstrated that both *PKMYT1* and *GAS5* possess strong diagnostic capabilities, particularly in distinguishing high-aggressive from low-aggressive PTC tissues. These findings emphasize the clinical relevance of these markers for stratifying tumor aggressiveness. In addition, *circMET* demonstrated robust diagnostic performance, with an accuracy exceeding 80% for PTC detection. Collectively, these results suggest that *PKMYT1* and its regulatory network components represent promising diagnostic biomarkers for PTC.

One of the most intriguing findings of our study is the calcium-dependent regulation of *GAS5* expression in PTC tissues. Specifically, *GAS5* expression was significantly reduced under low blood calcium conditions and upregulated when blood calcium levels were normal. This suggests a novel calcium-mediated regulatory mechanism involving *GAS5* in PTC development. Such calcium dependency may partly explain the variability

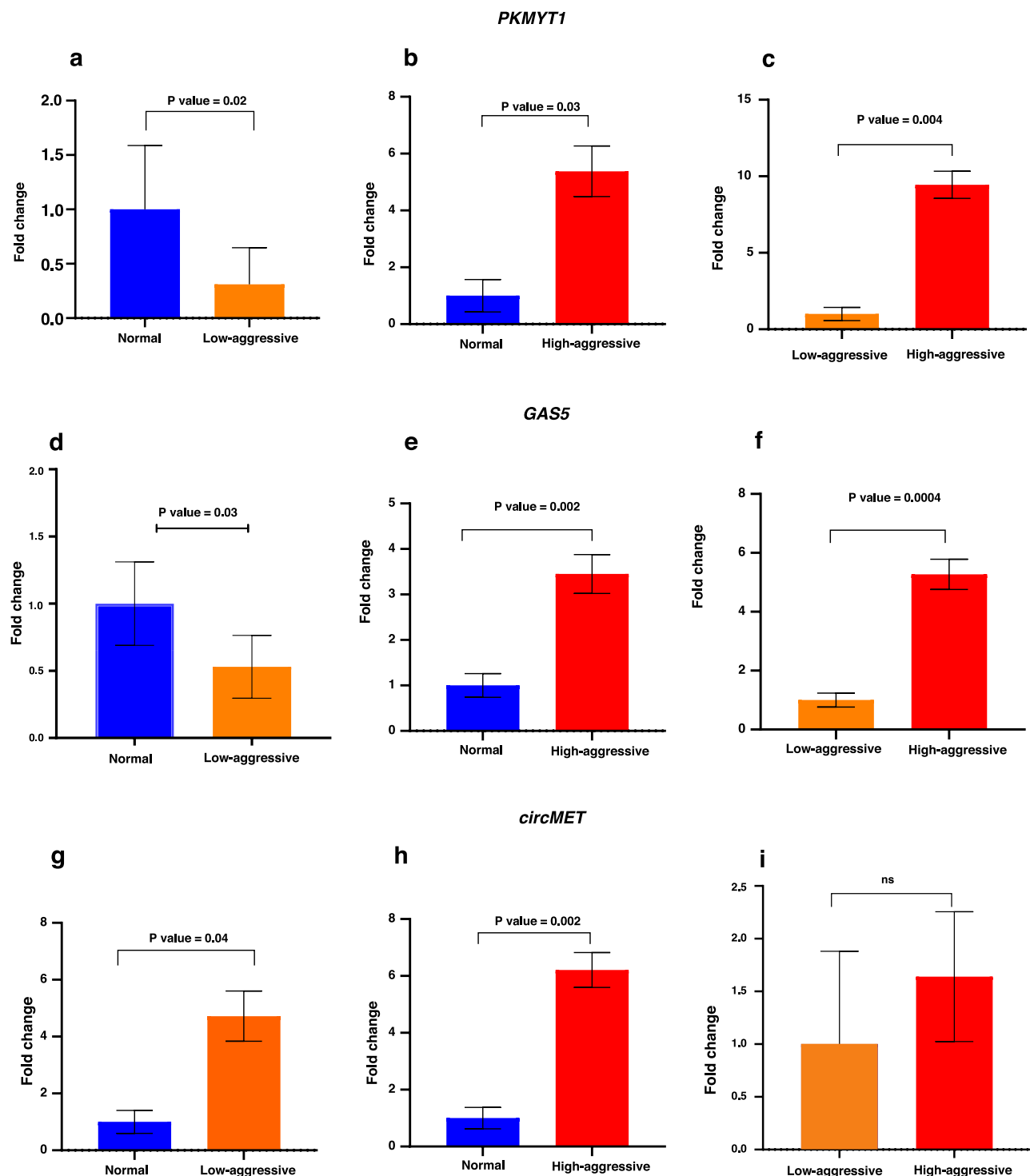


Fig. 7. Differential expression of *PKMYT1*, *GAS5*, and *circMET* in relation to PTC aggressiveness. (a) *PKMYT1* shows decreased expression in low-aggressive PTC tissues compared to normal thyroid tissue. (b) *PKMYT1* shows increased expression levels in high-aggressive PTC tissues relative to normal thyroid tissue. (c) *PKMYT1* shows increased expression levels in high-aggressive PTC tissues compared to those with low aggression. (d) *GAS5* expression is diminished in low-aggressive PTC tissues compared to normal thyroid tissue. (e) *GAS5* exhibits increased expression in high-aggressive PTC tissues relative to normal thyroid tissue. (f) *GAS5* shows markedly elevated expression levels in high-aggressive PTC tissues compared to low-aggressive PTC tissues. (g) *circMET* shows elevated expression in low-aggressive PTC tissues compared to normal thyroid tissue. (h) *circMET* exhibits increased expression in high-aggressive PTC tissues relative to normal thyroid tissue. (i) *circMET* expression remains unchanged across high-aggressive and low-aggressive PTC tissues. PTC: papillary thyroid cancer; *PKMYT1*: protein kinase membrane associated tyrosine/threonine 1; *GAS5*: growth arrest specific 5.

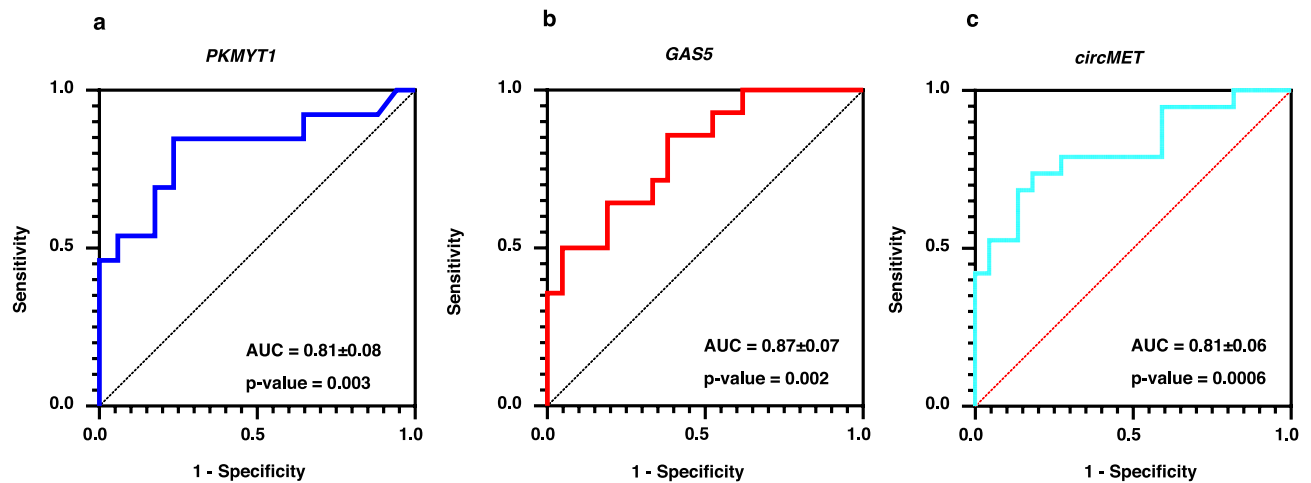


Fig. 8. Analysis of the diagnostic performance of *PKMYT1*, *GAS5*, and *circMET*. (a) *PKMYT1* exhibits notable diagnostic accuracy in distinguishing high-aggressive from low-aggressive PTC tissues. (b) *GAS5* demonstrates significant diagnostic accuracy in differentiating high-aggressive from low-aggressive PTC tissues. (c) *circMET* exhibits significant diagnostic effectiveness in distinguishing low-aggressive PTC tissues from normal thyroid tissues. PTC: papillary thyroid cancer; AUC: area under the curve; *PKMYT1*: protein kinase membrane associated tyrosine/threonine 1; *GAS5*: growth arrest specific 5.

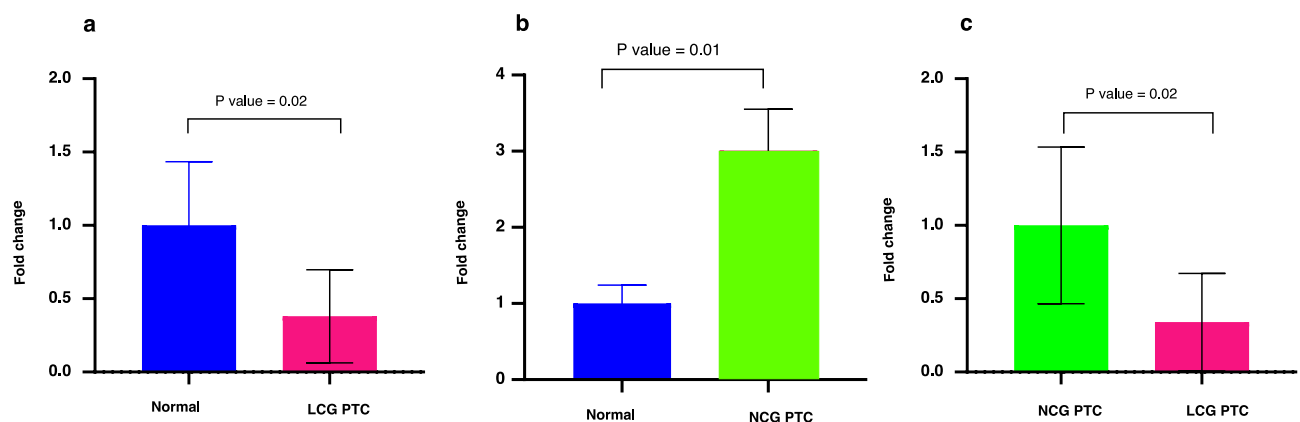


Fig. 9. Calcium-dependent differential expression of *GAS5* in PTC tissues. (a) *GAS5* expression is diminished in LCG PTC tissues compared to adjacent normal tissues. (b) *GAS5* shows elevated expression in NCG PTC tissues compared to adjacent normal tissues. (c) *GAS5* shows decreased expression in the LCG PTC tissues compared to the NCG PTC ones. LCG PTC: low-calcium level group papillary thyroid cancer; NCG PTC: normal-calcium level papillary thyroid cancer; *GAS5*: growth arrest specific 5.

observed in *GAS5* expression and offers important insights into its context-specific role in PTC progression. The discovery of calcium-dependent regulation of *GAS5* introduces a new layer of complexity to our understanding of PTC biology and suggests a potential therapeutic target. Future research should focus on elucidating the precise mechanisms underlying stage-specific and calcium-mediated expression patterns of *GAS5*. Additionally, validation in larger, independent patient cohorts will be important to confirm the clinical utility of these findings as diagnostic biomarkers.

Conclusion

A novel ceRNA and DE-TF network comprising components with diagnostic potential for PTC was developed. The integration of these components resulted in the creation of an eight-gene model that increased the diagnostic accuracy for PTC to 96.9%. *PKMYT1* and *GAS5* exhibited differential expression patterns correlated with the stages of PTC. In addition to *circMET*, these genes have demonstrated diagnostic potential, providing invaluable information about PTC progression and biomarker development. The calcium-dependent regulation of *GAS5* further emphasizes the importance of exploring metabolic factors in PTC biology.

Data availability

The original data presented in the study are openly available at [\[https://www.ncbi.nlm.nih.gov/geo/\]](https://www.ncbi.nlm.nih.gov/geo/), [\[https://www.gtexportal.org/\]](https://www.gtexportal.org/), [\[https://portal.gdc.cancer.gov/\]](https://portal.gdc.cancer.gov/)

Received: 5 April 2025; Accepted: 17 October 2025

Published online: 21 November 2025

References

- Hong, S. et al. Distinct molecular subtypes of papillary thyroid carcinoma and gene signature with diagnostic capability. *Oncogene* **41**, 5121–5132 (2022).
- Zaballos, M. A. & Santisteban, P. Key signaling pathways in thyroid cancer. *J. Endocrinol.* **235**, R43–R61 (2017).
- Toraih, E. A. et al. A miRNA-based prognostic model to trace thyroid cancer recurrence. *Cancers* **14**, 4128 (2022).
- Xu, L. et al. Comprehensive characterization of pathological stage-related genes of papillary thyroid cancer along with survival prediction. *J. Cell. Mol. Med.* **25**, 8390–8404 (2021).
- Liu, R. et al. A novel prognostic model for papillary thyroid cancer based on epithelial–mesenchymal transition-related genes. *Cancer Med.* **11**, 4703–4720 (2022).
- Prete, A. et al. Update on Fundamental Mechanisms of Thyroid Cancer. *Front. Endocrinol. (Lausanne)*. **11**, (2020).
- Salmena, L., Poliseno, L., Tay, Y., Kats, L. & Pandolfi, P. P. A ceRNA hypothesis: The rosetta stone of a hidden RNA language?. *Cell* **146**, 353–358 (2011).
- Shariatmadar Taleghani, A., Zohrab Beigi, Y., Zare-Mirakabad, F. & Masoudi-Nejad, A. Exploring ceRNA networks for key biomarkers in breast cancer subtypes and immune regulation. *Sci. Rep.* **13**, 20795 (2023).
- Liu, Y., Khan, S., Li, L., ten Hagen, T. L. M. & Falahati, M. Molecular mechanisms of thyroid cancer: A competing endogenous RNA (ceRNA) point of view. *Biomed. Pharmacother.* **146**, 112251 (2022).
- Qi, X., Lin, Y., Chen, J. & Shen, B. Decoding competing endogenous RNA networks for cancer biomarker discovery. *Brief. Bioinform.* **21**, 441–457 (2020).
- Li, Z., Zhou, J., Gu, L. & Zhang, B. Pseudogenes and the associated ceRNA network as potential prognostic biomarkers for colorectal cancer. *Sci. Rep.* **12**, 17787 (2022).
- An, Y., Furber, K. L. & Ji, S. Pseudogenes regulate parental gene expression via ceRNA network. *J. Cell. Mol. Med.* **21**, 185–192 (2017).
- Yang, Y. et al. A ceRNA network mediated by LINC00475 in papillary thyroid carcinoma. *Open Med.* **17**, 22–33 (2022).
- Xu, D. et al. The functional characterization of epigenetically related lncRNAs involved in dysregulated CeRNA–CeRNA networks across eight cancer types. *Front. Cell Dev. Biol.* **9**, 1–9 (2021).
- Zhong, Y. et al. Circular RNAs function as ceRNAs to regulate and control human cancer progression. *Mol. Cancer* **17**, 79 (2018).
- Colaprico, A. et al. TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71 (2016).
- Chen, Y., Chen, L., Lun, A. T. L., Baldoni, P. L. & Smyth, G. K. edgeR v4: Powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *Nucleic Acids Res.* **53**, gkaf018 (2025).
- Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
- Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
- Szklarczyk, D. et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).
- Chin, C. H. et al. cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* **8**, S11 (2014).
- Salavaty, A., Ramialison, M. & Currie, P. D. Integrated value of influence: An integrative method for the identification of the most influential nodes within networks. *Patterns* **1**, 100052 (2020).
- Shannon, P. et al. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
- Tang, Z. et al. GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* **45**, W98–W102 (2017).
- Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
- Collado-Torres, L. et al. Reproducible RNA-seq analysis using recount2. *Nat. Biotechnol.* **35**, 319–321 (2017).
- Chen, Y. & Wang, X. MiRDB: An online database for prediction of functional microRNA targets. *Nucleic Acids Res.* **48**, D127–D131 (2020).
- Kang, J. et al. RNAInter v40: RNA interactome repository with redefined confidence scoring system and improved accessibility. *Nucleic Acids Res.* **50**, D326–D332 (2022).
- Huang, H. Y. et al. MiRTarBase update 2022: An informative resource for experimentally validated miRNA–target interactions. *Nucleic Acids Res.* **50**, D222–D230 (2022).
- Sticht, C., De La Torre, C., Parveen, A. & Gretz, N. Mirwalk: An online resource for prediction of microRNA binding sites. *PLoS ONE* **13**, e0206239 (2018).
- Li, J. H., Liu, S., Zhou, H., Qu, L. H. & Yang, J. H. StarBase v20: Decoding miRNA–ceRNA, miRNA–ncRNA and protein–RNA interaction networks from large-scale CLIP–Seq data. *Nucleic Acids Res.* **42**, D92–D97 (2014).
- Jeggari, A., Marks, D. S. & Larsson, E. miRcode: A map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics* **28**, 2062–2063 (2012).
- Mann, M., Wright, P. R. & Backofen, R. IntaRNA 20: Enhanced and customizable prediction of RNA–RNA interactions. *Nucleic Acids Res.* **45**, W435–W439 (2017).
- Jr, H. Hmisc: Harrell Miscellaneous. <https://github.com/harrelfe/hmisc>. (2025).
- Wei, A. T. Visualization of a correlation matrix: Package ‘corrplot’. *CRAN: Contributed Packages* <https://cran.r-project.org/web/packages/corrplot/corrplot.pdf> (2015).
- Li, T. et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* **48**, W509–W514 (2020).
- Vafaee, F. et al. ORTI: An open-access repository of transcriptional interactions for interrogating mammalian gene expression data. *PLoS ONE* **11**, e0164535 (2016).
- Efthimiou, O. et al. Measuring the performance of survival models to personalize treatment choices. *Stat. Med.* **44**, e70050 (2025).
- Simon, R. M., Subramanian, J., Li, M. C. & Menezes, S. Using cross-validation to evaluate predictive accuracy of survival risk classifiers based on high-dimensional data. *Brief. Bioinform.* **12**, 203–214 (2011).
- Kassambara, A., Marcin Kosinski, M., Biecek, P. & Fabian, S. Package ‘survminer’ - Drawing Survival Curves using ‘ggplot2’. <https://rpkgs.datanovia.com/survminer/index.html>. (2024).
- Blanche, P., Dartigues, J. F. & Jacqmin-Gadda, H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat. Med.* **32**, 5381–5397 (2013).

42. Robin, X. et al. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011).
43. Chen, C. et al. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res.* **33**, e179–e179 (2005).
44. Dudekula, D. B. et al. Circinteractome: A web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biol.* **13**, 34–42 (2016).
45. Zhong, S. & Feng, J. CircPrimer 2.0: A software for annotating circRNAs and predicting translation potential of circRNAs. *BMC Bioinformatics* **23**, 215 (2022).
46. Boss, M. & Arenz, C. A fast and easy method for specific detection of circular RNA by rolling-circle amplification. *ChemBioChem* **21**, 793–796 (2020).
47. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402–408 (2001).
48. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, 45e–445 (2001).
49. Zhang, D. D. et al. Construction of a Myc-associated ceRNA network reveals a prognostic signature in hepatocellular carcinoma. *Mol. Ther. Nucleic Acids* **24**, 1033–1050 (2021).
50. Liu, Y. et al. Competitive endogenous RNA is an intrinsic component of EMT regulatory circuits and modulates EMT. *Nat. Commun.* **10**, 1637 (2019).
51. Li, L. et al. Construction of a four-mRNA prognostic signature with its ceRNA network in CESC. *Sci. Rep.* **12**, 10691 (2022).
52. Cheng, Y. et al. Identification of circrna-lncrna-mirna-mrna competitive endogenous rna network as novel prognostic markers for acute myeloid leukemia. *Genes* **11**, 1–18 (2020).
53. Li, H. et al. Overexpression of PKMYT1 associated with poor prognosis and immune infiltration may serve as a target in triple-negative breast cancer. *Front. Oncol.* **12**, 1–14 (2023).
54. Wang, J. et al. PKMYT1 is associated with prostate cancer malignancy and may serve as a therapeutic target. *Gene* **744**, 144608 (2020).
55. Liu, L. et al. PKMYT1 promoted the growth and motility of hepatocellular carcinoma cells by activating beta-catenin/TCF signaling. *Exp. Cell Res.* **358**, 209–216 (2017).
56. Xuan, Z. H. et al. PKMYT1 aggravates the progression of ovarian cancer by targeting SIRT3. *Eur. Rev. Med. Pharmacol. Sci.* **24**, 5259–5266 (2020).
57. Cai, Y. & Yang, W. PKMYT1 regulates the proliferation and epithelial-mesenchymal transition of oral squamous cell carcinoma cells by targeting CCNA2. *Oncol. Lett.* **23**, 63 (2022).
58. Zhang, Q. Y., Chen, X. Q., Liu, X. C. & Wu, D. M. PKMYT1 promotes gastric cancer cell proliferation and apoptosis resistance. *Oncotargets. Ther.* **13**, 7747–7757 (2020).
59. Zhu, T., Lou, Q., Shi, Z. & Chen, G. Identification of key miRNA-gene pairs in gastric cancer through integrated analysis of mRNA and miRNA microarray. *Am. J. Transl. Res.* **13**, 253–269 (2021).
60. Cilek, E. E., Ozturk, H. & Dedeoglu, B. G. Construction of miRNA-miRNA networks revealing the complexity of miRNA-mediated mechanisms in trastuzumab treated breast cancer cell lines. *PLoS ONE* **12**, 1–21 (2017).
61. Yu, Z., Zou, H., Wang, H., Li, Q. & Yu, D. Identification of key gene signatures associated with bone metastasis in castration-resistant prostate cancer using co-expression analysis. *Front. Oncol.* **10**, 571524 (2021).
62. Zhang, Z. et al. Identification of potential key genes and high-frequency mutant genes in prostate cancer by using RNA-seq data. *Oncol. Lett.* **15**, 4550–4556 (2018).
63. Guo, L. L. et al. Competing endogenous RNA networks and gastric cancer. *World J. Gastroenterol.* **21**, 11680–11687 (2015).
64. Gao, L., Zhao, Y., Ma, X. & Zhang, L. Integrated analysis of lncRNA-miRNA-mRNA ceRNA network and the potential prognosis indicators in sarcomas. *BMC Med. Genomics* **14**, 67 (2021).
65. Yang, X., Xie, Z., Lei, X. & Gan, R. Long non-coding RNA GAS5 in human cancer (Review). *Oncol. Lett.* **20**, 2587–2594 (2020).
66. Gupta, S., Panda, P. K., Luo, W., Hashimoto, R. F. & Ahuja, R. Network analysis reveals that the tumor suppressor lncRNA GAS5 acts as a double-edged sword in response to DNA damage in gastric cancer. *Sci. Rep.* **12**, 18312 (2022).
67. Wu, S. et al. lncRNA GAS5 represses stemness and malignancy of gliomas via elevating the SPACA6-miR-125a/let-7e Axis. *Front. Oncol.* **12**, 1–15 (2022).
68. Ji, J., Dai, X., Yeung, S. C. J. & He, X. The role of long non-coding RNA GAS5 in cancers. *Cancer Manag. Res.* **11**, 2729–2737 (2019).
69. Song, J. Y. et al. A fourteen-lncRNA risk score system for prognostic prediction of patients with non-small cell lung cancer. *Cancer Biomark.* **29**, 493–508 (2020).
70. Kaur, J. et al. Tumor suppressive effects of GAS5 in cancer cells. *Non-coding RNA* **8**, 39 (2022).
71. Wang, Y., Wu, S., Yang, X., Li, X. & Chen, R. Association between polymorphism in the promoter region of lncRNA GAS5 and the risk of colorectal cancer. *Biosci. Rep.* **39**, (2019).
72. Shams, F. & Ghorbani, S. Evaluation of prognostic usefulness of long noncoding RNA GAS5 and FAL1 in papillary thyroid carcinoma. *J. Cell. Biochem.* **120**, 11471–11477 (2019).
73. Dai, L. et al. The diagnostic potential of two exosome-derived circRNAs for papillary thyroid cancer. *Int. J. Clin. Oncol.* **28**, 1461–1474 (2023).
74. Lv, C. et al. Expression profiles of circular RNAs in human papillary thyroid carcinoma based on RNA deep sequencing. *Oncotargets. Ther.* **14**, 3821–3832 (2021).
75. Li, Y. et al. RMI2 is a prognostic biomarker and promotes tumor growth in hepatocellular carcinoma. *Clin. Exp. Med.* **22**, 229–243 (2022).
76. Tsai, H. W., Cheng, S. W., Chen, C. C., Chen, I. W. & Ho, C. L. A combined bioinformatics and experimental approach identifies RMI2 as a Wnt/ β -catenin signaling target gene related to hepatocellular carcinoma. *BMC Cancer* **23**, 1025 (2023).
77. Zhang, L. et al. RMI2 is a novel prognostic and predictive biomarker for breast cancer. *Cancer Med.* **12**, 8331–8350 (2023).
78. Ren, W. et al. miR-616-5p Promotes Invasion and Migration of Bladder Cancer via Downregulating NR2C2 Expression. *Front. Oncol.* **11**, 762946 (2021).
79. Zhang, F. et al. Apatinib combined with keytruda treatment induces apoptosis of gastric carcinoma cells through CES4/miR-616-5p/DUSP2 axis. *Basic Clin. Pharmacol. Toxicol.* **129**, 345–356 (2021).
80. Zhang, X. T. et al. Overexpression of E2F mRNAs associated with gastric cancer progression identified by the transcription factor and miRNA co-regulatory network analysis. *PLoS ONE* **10**, e0116979 (2015).
81. Hung, J.-J., Hsueh, C.-T., Chen, K.-H., Hsu, W.-H. & Wu, Y.-C. Clinical significance of E2F1 protein expression in non-small cell lung cancer. *Exp. Hematol. Oncol.* **1**, 18 (2012).
82. Yang, J. et al. Transcription factor E2F1 exacerbates papillary thyroid carcinoma cell growth and invasion via upregulation of LINC00152. *Anal. Cell. Pathol.* **2022**, 1–10 (2022).
83. Onda, M. et al. Up-regulation of transcriptional factor E2F1 in papillary and anaplastic thyroid cancers. *J. Hum. Genet.* **49**, 312–318 (2004).
84. Zhang, Z., Li, S. Y. & Zhang, L. B. lncRNA RGMB-AS1 is activated by E2F1 and promotes cell proliferation and invasion in papillary thyroid carcinoma. *Eur. Rev. Med. Pharmacol. Sci.* **22**, 1979–1986 (2018).
85. Qi, L., Ren, Z. & Li, W. E2F4 transcription factor is a prognostic biomarker related to immune infiltration of head and neck squamous cell carcinoma. *Sci. Rep.* **12**, 12132 (2022).
86. Liu, J. et al. E2F4 promotes the proliferation of hepatocellular carcinoma cells through upregulation of CDCA3. *J. Cancer* **12**, 5173–5180 (2021).

87. Sun, C. C. et al. Comprehensive analysis of the expression and prognosis for E2Fs in human breast cancer. *Mol. Ther.* **27**, 1153–1165 (2019).
88. Luo, Z. et al. Cytokine-induced apoptosis inhibitor 1: A comprehensive analysis of potential diagnostic, prognosis, and immune biomarkers in invasive breast cancer. *Transl. Cancer Res.* **12**, 1765–1786 (2023).
89. Huang, X. et al. Association of DBNDD1 with prognostic and immune biomarkers in invasive breast cancer. *Discov. Oncol.* **16**, 218 (2025).

Acknowledgements

The author expresses gratitude to the volunteers for their time and effort. The author acknowledges the Departments of Research, Technology, and Graduate Offices of the University of Isfahan, Iran, for their financial support of this study. We express our gratitude to Dr. Mohsen Kolahdouzan, the thoracic surgeon, for his kindness, patience, and responsibility in procuring thyroid tissues. Gratitude is also due to Dr. Keivan Shirneshan for the approval of the pathological results.

Author contributions

SEYR: Concept, design, experiments, data curation and analysis, results interpretation, original draft writing, SMJR: Concept, design, project administration, data curation and analysis, visualization, results interpretation, original draft writing & editing, and funding acquisition, final approval of the manuscript, MK: Concept, design, results interpretation, original draft writing & editing, final approval of the manuscript.

Funding

This study was conducted at Isfahan University of Iran and was supported financially by the Departments of Research, Technology, and Graduate Offices (Grant number: 59792306).

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

Informed consent was obtained from all subjects and their legal guardian(s). The institutional review board of Isfahan University (IR.UI.REC.1403.006) has authorized the procurement of human tissues. All experiments and procedures conducted in this study, particularly those involving human participants, adhered to the 1964 Helsinki Declaration and its subsequent amendments, or comparable ethical standards. The authors express gratitude to the volunteers for their participation.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-25114-0>.

Correspondence and requests for materials should be addressed to S.-M.J.

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

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

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

© The Author(s) 2025