

Integrated Multi-Omics Bioinformatics and Machine Learning Framework Identifies SPATA18 and IGF2BP2 as Robust Prognostic Biomarkers for Papillary Thyroid Carcinoma

Xingxing Zhao

The Second Affiliated Hospital of Guizhou Medical University

Zhitao Chen

The Second Affiliated Hospital of Guizhou Medical University

Chuan Zhang

The Second Affiliated Hospital of Guizhou Medical University

Jie Peng

The Second Affiliated Hospital of Guizhou Medical University

Chunyan Mao

Xingyi People's Hospital

Guangmin Meng

The Second Affiliated Hospital of Guizhou Medical University

Jie Tian

The Second Affiliated Hospital of Guizhou Medical University

Xinying Quan

Beijing Jishuitan Hospital Guizhou Hospital

Lei Shi

The Second Affiliated Hospital of Guizhou Medical University

Lu Pan

The Second Affiliated Hospital of Guizhou Medical University

Xianjie Peng

Qiandongnan Prefecture People's HospitalKaili

Xiaoyao Cai

xiaoyaocai_mc@sina.com

Qiandongnan Prefecture People's HospitalKaili

Keywords: papillary thyroid carcinoma, SPATA18, IGF2BP2, machine learning, diagnostic and prognostic biomarkers

Posted Date: June 30th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9899741/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Papillary thyroid carcinoma (PTC) is marked by diverse molecular changes, yet there is a paucity of reliable biomarkers for its early diagnosis and prognosis. This study seeks to identify and validate robust transcriptomic signatures with both diagnostic and prognostic significance in PTC.

Methods

We integrated transcriptomic data from public repositories and conducted differential expression analysis, weighted gene co-expression network analysis (WGCNA), and machine learning-based feature selection techniques, including LASSO, Random Forest, and XGBoost, to pinpoint key genes. Functional enrichment, gene set enrichment analysis (GSEA), and gene set variation analysis (GSVA) were utilized to elucidate the underlying biological pathways. The diagnostic and prognostic efficacy of candidate genes was assessed through ROC analysis, nomogram construction, and decision curve analysis. Additionally, a lncRNA–miRNA regulatory network was constructed to investigate post-transcriptional mechanisms.

Results

The study identified two consensus genes, SPATA18 and IGF2BP2, as robust biomarkers with significant upregulation in papillary thyroid carcinoma (PTC), exhibiting exceptional diagnostic efficacy (AUC > 0.95). A prognostic nomogram incorporating these genes demonstrated high predictive accuracy (AUC = 0.945) and clinical applicability. Gene Set Enrichment Analysis (GSEA) indicated their involvement in oxidative phosphorylation, p53 signaling, extracellular matrix-receptor interaction, and hypoxia response pathways. Additionally, Gene Set Variation Analysis (GSVA) underscored widespread dysregulation in metabolic and immune-related pathways.

Conclusions

SPATA18 and IGF2BP2 have been identified as potential biomarkers with substantial diagnostic and prognostic significance in PTC. The developed nomogram and regulatory network elucidate underlying molecular mechanisms and potential therapeutic targets, thereby enhancing personalized risk stratification and informing clinical decision-making.

1 Introduction

Thyroid carcinoma (TC) represents the most prevalent malignancy of the endocrine system, with its global incidence exhibiting a marked upward trend over recent decades[1]. Among all histological

subtypes, papillary thyroid carcinoma (PTC) accounts for approximately 80% of cases[2]. While the majority of PTC patients enjoy favorable prognoses with standard treatment, a subset experiences aggressive disease progression, including lymph node metastasis, recurrence, and even dedifferentiation[3]. This clinical heterogeneity underscores the pressing need for reliable biomarkers capable of early detection, accurate risk stratification, and guiding personalized therapeutic interventions[4, 5].

The rapid advancement of high-throughput sequencing technologies and public genomic databases has catalyzed the discovery of molecular drivers in PTC. Previous investigations have identified several pivotal mechanisms underlying PTC pathogenesis. These include SUMOylation-related modifications implicating genes such as BMP8A, RGS8, and SERPIND1[6, 7]; dynamic alterations in the tumor immune microenvironment, particularly involving CD8 + T cell infiltration[8, 9]; metabolic reprogramming mediated by molecules like TBC1D4[10–12]; and epigenetic regulation, exemplified by DNMT3B-mediated methylation of FAM111B[13, 14]. Despite these significant strides, the molecular landscape of PTC remains incompletely elucidated, and clinically applicable biomarkers with robust diagnostic and prognostic performance are still lacking[15, 16].

The integration of multi-omics data with sophisticated bioinformatics approaches offers a powerful strategy for uncovering novel biomarkers. Weighted gene co-expression network analysis (WGCNA) enables the identification of functionally relevant gene modules associated with clinical traits, while machine learning algorithms such as LASSO, Random Forest, and XGBoost facilitate the selection of stable and interpretable feature genes from high-dimensional data[17, 18]. Furthermore, comprehensive functional enrichment analyses are indispensable for deciphering the biological pathways engaged by candidate biomarkers[19, 20].

In this study, we integrated transcriptomic data from public repositories and implemented a systematic bioinformatics framework incorporating differential expression analysis, WGCNA, and multiple machine learning algorithms. Our objectives were threefold: (1) to identify and validate core genes with diagnostic and prognostic significance in PTC; (2) to construct a reliable prognostic model based on these genes and assess its clinical utility; and (3) to explore the potential biological functions and regulatory networks of the identified biomarkers. We successfully pinpointed SPATA18 and IGF2BP2 as promising biomarkers and developed a prognostic nomogram, providing new insights into the molecular mechanisms of PTC and potential tools for personalized patient management.

2 Materials and methods

2.1 Data acquisition and preprocessing

Public RNA-seq data encompassing papillary thyroid carcinoma (PTC) and normal thyroid tissue samples were retrieved from the Gene Expression Omnibus (GEO) (accession number: GSE33630 and GSE29265). After quality control and removal of batch effects using the ComBat algorithm, normalized

expression matrices were generated. A total of 69 PTC samples and 65 normal thyroid tissue samples were included for subsequent analyses.

2.2 Differential expression analysis and functional enrichment

Differentially expressed genes (DEGs) between PTC and normal tissues were identified using the limma R package, with thresholds set at $|\log_2(\text{fold change})| > 1$ and adjusted p-value < 0.05 [21, 22]. Functional enrichment analysis, including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses, was performed using the clusterProfiler package. Terms with an adjusted p-value < 0.05 were considered significantly enriched.

2.3 Weighted gene co-expression network analysis (WGCNA)

A weighted gene co-expression network was constructed using the WGCNA R package on the GSE33630 dataset. Gene modules were identified through hierarchical clustering and dynamic tree cutting, with a minimum module size of 50 genes. Module-trait relationships were assessed by correlating module eigengenes with clinical traits, focusing on modules significantly associated with PTC status.

2.4 Machine learning-based feature selection

To identify robust prognostic biomarkers, we employed three machine learning algorithms: LASSO (Least Absolute Shrinkage and Selection Operator) regression, Random Forest (RF), and eXtreme Gradient Boosting (XGBoost). LASSO regression was performed using the glmnet package with ten-fold cross-validation to select the optimal penalty parameter (λ)[23]. Random Forest modeling was implemented with optimized decision trees, and feature importance was evaluated using Mean Decrease Gini and Mean Decrease Accuracy. [24]. XGBoost analysis was conducted with default parameters, and SHAP (SHapley Additive exPlanations) analysis was applied to interpret feature importance. Genes selected by all three algorithms were considered consensus feature genes[25].

2.5 Construction and validation of a prognostic nomogram

A prognostic nomogram was developed by integrating the expression levels of the consensus feature genes using the rms R package. The model's performance was evaluated through internal validation. Calibration curves were plotted to compare predicted probabilities with observed outcomes. Decision curve analysis (DCA) and clinical impact curve (CIC) analysis were conducted to assess the clinical utility of the model. The discriminative ability was quantified using receiver operating characteristic (ROC) curve analysis in both training and validation cohorts[26].

2.6 Immune infiltration and tumor microenvironment analysis

The relative abundances of 22 immune cell types in each sample were estimated using the CIBERSORT algorithm. The ESTIMATE algorithm was applied to calculate stromal, immune, and ESTIMATE scores. Spearman correlation analysis was performed to evaluate the relationships between key biomarker expression and immune cell infiltration levels.

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

GSEA was performed using the HALLMARK gene set database to identify biological pathways associated with high and low expression of key biomarkers. A normalized enrichment score (NES) with $|\text{NES}| > 1.5$ and a false discovery rate (FDR) < 0.25 were considered significant. GSVA was conducted using the GSVA R package with the HALLMARK gene set collection to estimate pathway activity variations across all samples[27, 28].

2.8 Construction of a lncRNA-miRNA interaction network

Potential lncRNA-miRNA interactions were predicted using the ENCORI database with a threshold of 10. The regulatory network was visualized using Cytoscape software. Hub molecules within the network were identified based on degree centrality[29].

3 Results

3.1 Transcriptomic profiling identifies widespread dysregulation in PTC

Differential expression analysis identified a total of 841 significantly dysregulated genes (459 upregulated and 382 downregulated) in PTC compared to normal thyroid tissues ($|\log_2\text{FC}| > 1$, adj. $p < 0.05$) (Fig. 1A). KEGG pathway enrichment analysis of these DEGs revealed significant involvement in pathways related to extracellular matrix (ECM) organization, cancer pathogenesis (e.g., p53 signaling, transcriptional misregulation), and immune response (e.g., complement and coagulation cascades) (Fig. 1B). GO analysis further highlighted dominant roles in biological processes such as extracellular matrix organization and synaptic signaling, and molecular functions including serine-type peptidase activity (Fig. 1C, D).

3.2 WGCNA identifies key gene modules correlated with PTC

Sample clustering revealed no outliers (Fig. 2A). WGCNA constructed a scale-free co-expression network with a soft-thresholding power of 13 (Fig. 2B). eigengene clustering revealed their relationships (Fig. 2C), and Hierarchical clustering identified 6 distinct co-expression modules (Fig. 2D). Module-trait association analysis demonstrated that several modules, most notably the 'blue' and 'magenta' modules, showed highly significant correlations with the PTC phenotype (Fig. 2E). Within the blue module, a strong

positive correlation was observed between gene significance for PTC and module membership, validating its biological relevance (Fig. 2F).

3.3 LASSO-RF-XGBoost integrated framework identifies SPATA18 and IGF2BP2 as robust PTC biomarkers

To identify robust diagnostic and prognostic biomarkers for papillary thyroid carcinoma (PTC), we established an integrated framework combining three high-performance machine learning algorithms: least absolute shrinkage and selection operator (LASSO), random forests (RF), and eXtreme Gradient Boosting (XGBoost), to systematically filter candidate feature genes. The LASSO algorithm, a penalized regression method optimizing model interpretability and accuracy via a contraction penalty function, was implemented with the “glmnet” R package. Guided by coefficient profiles and binomial deviance curves, the optimal penalty parameter was determined, and a 10-gene signature was ultimately identified (Figs. 3A, 3B). For RF analysis, we used the “randomForest” R package, with the number of decision trees optimized via the error convergence curve (Fig. 3C). Feature importance was ranked by Mean Decrease Gini (Fig. 3D) and Mean Decrease Accuracy (Fig. 3E), with SPATA18 and IGF2BP2 consistently ranked as the top influential genes. The XGBoost model, built on regularized gradient-boosted decision trees, was interpreted via SHAP analysis (Fig. 3F) and feature importance metrics (Figs. 3G, 3H), showing consistent key gene rankings with the aforementioned algorithms. A Venn diagram visualized the consensus genes from the three high-performance classifiers, and their intersection pinpointed SPATA18 and IGF2BP2 as high-confidence PTC biomarkers (Fig. 3I). The overlapping candidates proceeded to subsequent validation stages, with diagnostic accuracy evaluated via ROC curves and corresponding AUC values using the “pROC” R package.

3.4 SPATA18 and IGF2BP2 exhibit superior diagnostic performance

Receiver operating characteristic (ROC) curve analysis was performed to evaluate the diagnostic potential of the consensus genes. Both SPATA18 and IGF2BP2 demonstrated exceptional capability in distinguishing PTC tissues from normal controls, with Area Under the Curve (AUC) values of 0.951 each (Figs. 4A, B), confirming their outstanding power as diagnostic biomarkers.

3.5 A prognostic nomogram integrating SPATA18 and IGF2BP2 shows high clinical utility

A prognostic nomogram was constructed by integrating the expression levels of SPATA18 and IGF2BP2 to predict individual PTC risk (Fig. 5A). The model showed excellent calibration, with the calibration curve closely aligning with the ideal line (Fig. 5B). Decision curve analysis (DCA) indicated that the nomogram provided a higher net benefit across a wide range of risk thresholds compared to “treat-all” or “treat-none” strategies (Fig. 5C). Clinical impact curves further visualized its potential utility (Fig. 5D). The

model achieved an AUC of 0.945 in the training set (Fig. 5E) and maintained robust performance (AUC = 0.940) in the validation set (Fig. 5F).

3.6 GSEA elucidates functional roles of SPATA18 and IGF2BP2 in key pathways

GSEA revealed distinct pathway associations for the two biomarkers. High expression of SPATA18 was positively associated with 'Oxidative phosphorylation' (NES = 2.21), 'p53 signaling pathway' (NES = 1.81), and 'Chemokine signaling pathway' (NES = 1.63) (Fig. 6A). Its low expression was linked to the inhibition of pathways like 'TGF-beta signaling' (NES=-1.76) and 'Focal adhesion' (NES=-1.72) (Fig. 6B). For IGF2BP2, high expression strongly enriched 'ECM-receptor interaction' (NES = 2.63), 'Focal adhesion' (NES = 2.29), and 'Response to hypoxia' (NES = 1.80) (Fig. 6C). Conversely, pathways including 'Autoimmune thyroid disease' (NES=-1.79) and 'TCA cycle' (NES=-1.83) were suppressed in the IGF2BP2 low-expression group (Fig. 6D).

3.7 GSVA reveals global pathway dysregulation in the PTC cohort

Gene Set Variation Analysis (GSVA) provided a panoramic view of pathway activity alterations. Unsupervised clustering of GSVA enrichment scores for HALLMARK gene sets clearly separated PTC samples from normal tissues (Fig. 7A). Notably, pathways such as 'MYC TARGETS', 'EPITHELIAL MESENCHYMAL TRANSITION', and 'HYPOXIA' were significantly activated in tumors, while metabolic pathways like 'OXIDATIVE PHOSPHORYLATION' and 'FATTY ACID METABOLISM' were suppressed (Fig. 7B), indicating profound metabolic reprogramming in PTC.

3.8 The core lncRNA-miRNA interaction network suggests potential regulatory mechanisms

A core lncRNA-miRNA interaction network was constructed to explore the post-transcriptional regulatory mechanisms of the key biomarkers in PTC. The network comprised 31 lncRNAs and 18 miRNAs based on predictions (Fig. 8). Topological analysis identified multiple hub nodes with high connectivity, including miRNAs that directly target IGF2BP2 and SPATA18, highlighting their critical roles in the ceRNA regulatory circuitry of PTC.

3.9 Validation of biomarker expression in an independent cohort

Differential expression analysis in an independent validation dataset (GSE33630) confirmed the significant upregulation of both SPATA18 and IGF2BP2 mRNA levels in PTC tissues (n = 49) compared to

normal thyroid tissues ($n = 45$ ($p < 0.001$)) (Figs. 9A, B), providing robust independent confirmation of our bioinformatics findings.

4 Discussion

In this study, we implemented a comprehensive bioinformatics framework integrating transcriptomic profiling, co-expression network analysis, and multiple machine learning algorithms to identify and validate SPATA18 and IGF2BP2 as robust diagnostic and prognostic biomarkers in PTC. Our prognostic nomogram, combining these two genes, demonstrated exceptional predictive accuracy ($AUC > 0.94$) and clinical utility, providing a potential tool for personalized risk stratification.

Mechanistically, The pivotal role of SPATA18, a gene implicated in mitochondrial function and oxidative metabolism, in cancer is increasingly recognized[30]. Our GSEA results, showing significant enrichment of 'Oxidative Phosphorylation' and 'p53 Signaling Pathway' with high SPATA18 expression, align with recent studies highlighting metabolic reprogramming as a hallmark of PTC[31]. Concurrently, the oncogenic role of IGF2BP2 as an RNA-binding protein regulating mRNA stability and translation is well-established in various cancers[32, 33]. Our findings that IGF2BP2 high expression enriches 'ECM-Receptor Interaction' and 'Focal Adhesion' pathways corroborate its involvement in modulating the tumor microenvironment (TME), a critical aspect of PTC progression.

Notably, Post-translational modifications are key regulators of oncogenic signaling in thyroid cancer, and may serve as upstream modulators of the biomarkers identified in our study. SUMOylation, one of the critical post-translational modifications, has been reported to participate in thyroid carcinogenesis, and is involved in regulating metabolic enzymes and extracellular matrix remodeling factors—core biological processes associated with SPATA18 and IGF2BP2 in our enrichment analysis[6]. This suggests a potential regulatory link between SUMOylation and the oncogenic pathways driven by our two biomarkers, which warrants further exploration.

Beyond their direct roles in regulating intrinsic tumor cell pathways, Our immune infiltration analysis revealed significant correlations between the biomarker expression signatures and altered immune cell abundances. This is consistent with the growing understanding that the immune landscape is a key determinant of PTC behavior and patient outcomes[34]. The observed associations with specific immune cells, particularly those with anti-tumor properties such as CD8 + T cells, resonate with studies highlighting the role of immune surveillance in containing PTC progression[9]. These findings further support that SPATA18 and IGF2BP2 may influence PTC progression not only through intrinsic tumor cell regulation, but also by remodeling the tumor immune microenvironment.

From a multi-omics perspective, Our study, primarily focused on the transcriptomic layer, is supported by recent multi-omics investigations in thyroid cancer. A 2025 study characterizing the epigenetic landscapes of benign thyroid nodules versus PTCs revealed that chromatin architecture and somatic mutation hotspots are spatially coordinated in malignant transformation[35]. While our work did not include epigenomic profiling, the dysregulated pathways we identified, such as p53 signaling and ECM

organization, are frequently under epigenetic control. The potential regulation of SPATA18 and IGF2BP2 by DNA methylation or other epigenetic mechanisms, as suggested by studies like the one on DNMT3B-mediated methylation of FAM111B in PTC, presents an exciting avenue for future research[13, 14].

Clinically, The high diagnostic and prognostic accuracy of our model underscores its translational potential. The nomogram could aid clinicians in identifying high-risk patients who might benefit from more intensive surveillance or adjuvant therapy. From a therapeutic perspective, the close association of our biomarkers with core oncogenic pathways (metabolic reprogramming and TME remodeling) hints at their potential as promising therapeutic targets for aggressive PTC. Our regulatory network analysis further suggests potential for post-transcriptional modulation of these biomarkers using miRNA- or lncRNA-based strategies, providing additional directions for targeted intervention development.

Nevertheless, Several limitations of our study should be acknowledged. Firstly, as a retrospective analysis based on public databases, prospective studies with larger, multi-center cohorts are required to validate our findings. Secondly, the precise molecular mechanisms through which SPATA18 and IGF2BP2 contribute to PTC pathogenesis need to be fully elucidated through systematic functional experiments in vitro and in vivo. Future research should integrate multi-omics data (e.g., epigenomics, proteomics) to paint a more comprehensive picture of the upstream and downstream regulatory networks of these two biomarkers. Additionally, exploring the potential of these biomarkers in liquid biopsies could enhance their clinical applicability for non-invasive diagnosis and long-term monitoring of PTC patients.

5 Conclusions

In conclusion, our integrative multi-omics bioinformatics analysis has successfully identified and validated SPATA18 and IGF2BP2 as two robust, high-confidence biomarkers for papillary thyroid carcinoma. The study demonstrates that a prognostic model integrating these two genes exhibits exceptional diagnostic and prognostic performance, highlighting its potential for enhancing personalized risk assessment and clinical decision-making in PTC management.

The functional characterization of these biomarkers reveals their involvement in distinct yet complementary oncogenic pathways. SPATA18 is predominantly linked to metabolic reprogramming, specifically oxidative phosphorylation, and the p53 signaling pathway. In contrast, IGF2BP2 plays a critical role in remodeling the tumor microenvironment through its strong association with extracellular matrix-receptor interactions and focal adhesion pathways. Furthermore, our discussion situates these findings within the broader molecular context of PTC, suggesting their potential regulation by key post-translational modification mechanisms implicated in thyroid carcinogenesis[7].

While this study provides valuable insights derived from a comprehensive bioinformatics framework, it also underscores the necessity for future research. The precise molecular mechanisms through which SPATA18 and IGF2BP2 exert their functions warrant rigorous experimental validation using in vitro and in vivo models. Prospective studies with larger, multi-center cohorts are essential to further solidify the

clinical translatability of our prognostic model. Ultimately, targeting the SPATA18/IGF2BP2 axis or their associated regulatory networks may emerge as a novel therapeutic strategy for aggressive PTC.

Declarations

Ethical approval and Accordance Statement

Not applicable. This study utilized publicly available datasets and did not involve direct human subjects research. Therefore, ethical approval was not required.

Consent to participate

Not applicable, as this study is based entirely on publicly available transcriptomic datasets.

Consent to publish

Not applicable.

Competing interests

The authors declare that they have no financial interests or personal affiliations that could have influenced the outcomes of this study.

Funding

This work was supported by the Qianxinan Prefecture Science and Technology Planning Project (No. 2023-3-52) and the Guizhou Medical University National Natural Science Foundation Cultivation Project(No.25NSFCP35)

Author Contribution

XXZ: conceived the study, collected and processed the data, performed the key bioinformatics analyses, interpreted the main findings, and drafted the manuscript. ZTC, CZ, JP and CYM: Methodology and Data process. JP, CYM, GMM, JT, XYQ, LS, LP, and XJP: Analysis and Visualization. XYZ: designed, Investigation and Supervision. all authors: Writing-review & editing.

Acknowledgement

We thank the researchers who provided sequencing data as well as all the participants for their contributions to this study.

Data Availability

The datasets analyzed during the current study are available in the NCBI Gene Expression Omnibus (GEO) repository: GSE33630: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE33630> GSE29265: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29265> The ENCORI database used for lncRNA-miRNA interaction prediction is available at <https://starbase.sysu.edu.cn/>. The HALLMARK gene set collections used for GSEA and GSVA analysis are available at <https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp#H>. The original R and Python code used for data analysis, machine learning modeling and figure generation is available upon reasonable request from the corresponding author.

References

1. Li M, Dal Maso L, Pizzato M, Vaccarella S. Evolving epidemiological patterns of thyroid cancer and estimates of overdiagnosis in 2013-17 in 63 countries worldwide: a population-based study. *Lancet Diabetes Endocrinol*. 2024;12(11):824–36. [https://doi.org/10.1016/s2213-8587\(24\)00223-7](https://doi.org/10.1016/s2213-8587(24)00223-7).
2. Kim J, Gosnell JE, Roman SA. Geographic influences in the global rise of thyroid cancer. *Nat Rev Endocrinol*. 2020;16(1):17–29. <https://doi.org/10.1038/s41574-019-0263-x>.
3. Xu S, Huang H, Huang Y, Qian J, Wang X, Xu Z, Liu S, Liu J. Comparison of Lobectomy vs Total Thyroidectomy for Intermediate-Risk Papillary Thyroid Carcinoma With Lymph Node Metastasis. *JAMA Surg*. 2023;158(1):73–9. <https://doi.org/10.1001/jamasurg.2022.5781>.
4. Liu W, Zhu J, Wu Z, Yin Y, Wu Q, Wu Y, Zheng J, Wang C, Chen H, Qazi TJ, Wu J, Zhang Y, Liu H, Yang J, Lu D, Zhang X, Ai Z. Insight of novel biomarkers for papillary thyroid carcinoma through multiomics. *Front Oncol*. 2023;13:1269751. <https://doi.org/10.3389/fonc.2023.1269751>.
5. Ji G, Sun H, Chen S, Sun X, Chang L, Xie R, Huang R, Zheng L, Chang Z. Single-cell RNA sequencing and multi-omics analysis of prognosis-related staging in papillary thyroid cancer. *Cancer Immunol Immunother*. 2025;74(8):267. <https://doi.org/10.1007/s00262-025-04101-4>.
6. Li X, Ding Z, Tong Y. Identification of SUMOylation-related biomarkers in papillary thyroid carcinoma. *Cancer Cell Int*. 2024;24(1):149. <https://doi.org/10.1186/s12935-024-03323-3>.
7. Rodrigues JS, Chenlo M, Bravo SB, Perez-Romero S, Suarez-Fariña M, Sobrino T, Sanz-Pamplona R, González-Prieto R, Blanco Freire MN, Nogueiras R, López M, Fugazzola L, Cameselle-Teijeiro JM, Alvarez CV. dsRNAi-mediated silencing of PIAS2beta specifically kills anaplastic carcinomas by mitotic catastrophe. *Nat Commun*. 2024;15(1):3736. <https://doi.org/10.1038/s41467-024-47751-1>.
8. Jing R, Wu N, Zhang Q, Liu J, Zhao Y, Zeng S, Wu S, Wu Y, Yi S. DPP4 promotes an immunoenhancing tumor microenvironment through exhausted CD8 + T cells with activating IL13-IL13RA2 axis in papillary thyroid cancer. *Int Immunopharmacol*. 2025;145:113760. <https://doi.org/10.1016/j.intimp.2024.113760>.
9. Chen Z, Guo ML, Li YY, Yan K, Li L, Shen F, Guan H, Liu QZ, Xu B, Lian ZX. Immune profiling identifies CD8(+) T-cell subset signatures as prognostic markers for recurrence in papillary thyroid cancer.

- Front Immunol. 2022;13:894919. <https://doi.org/10.3389/fimmu.2022.894919>.
10. Agarwal S, Gupta S, Raj R. Identification of potential targetable genes in papillary, follicular, and anaplastic thyroid carcinoma using bioinformatics analysis. *Endocrine*. 2024;86(1):255–67. <https://doi.org/10.1007/s12020-024-03836-x>.
 11. Song H, Qiu Z, Wang Y, Xi C, Zhang G, Sun Z, Luo Q, Shen C. HIF-1 α /YAP Signaling Rewrites Glucose/Iodine Metabolism Program to Promote Papillary Thyroid Cancer Progression. *Int J Biol Sci*. 2023;19(1):225–41. <https://doi.org/10.7150/ijbs.75459>.
 12. Wen SS, Wu YJ, Wang JY, Ni ZX, Dong S, Xie XJ, Wang YT, Wang Y, Huang NS, Ji QH, Ma B, Qu N. BRAF(V600E)/p-ERK/p-DRP1(Ser616) Promotes Tumor Progression and Reprogramming of Glucose Metabolism in Papillary Thyroid Cancer. *Thyroid*. 2024;34(10):1246–59. <https://doi.org/10.1089/thy.2023.0700>.
 13. Zhu X, Xue C, Kang X, Jia X, Wang L, Younis MH, Liu D, Huo N, Han Y, Chen Z, Fu J, Zhou C, Yao X, Du Y, Cai W, Kang L, Lyu Z. DNMT3B-mediated FAM111B methylation promotes papillary thyroid tumor glycolysis, growth and metastasis. *Int J Biol Sci*. 2022;18(11):4372–87. <https://doi.org/10.7150/ijbs.72397>.
 14. Wei F, Li W, Zhou T, Yuan X, Zhang L, Unveiling. FAM111B: A Pan-Cancer Biomarker for DNA Repair and Immune Infiltration. *Int J Mol Sci*. 2025;26(7). <https://doi.org/10.3390/ijms26073151>.
 15. Liao M, Wang Z, Yao J, Xing H, Hao Y, Qiu B. Identification of potential biomarkers for papillary thyroid carcinoma by comprehensive bioinformatics analysis. *Mol Cell Biochem*. 2023;478(9):2111–23. <https://doi.org/10.1007/s11010-022-04606-x>.
 16. Forma A, Kłodnicka K, Pająk W, Flieger J, Teresińska B, Januszewski J, Baj J. Thyroid Cancer: Epidemiology, Classification, Risk Factors, Diagnostic and Prognostic Markers, and Current Treatment Strategies. *Int J Mol Sci*. 2025;26(11). <https://doi.org/10.3390/ijms26115173>.
 17. Zhang Y, Xie X, Zhou H, Li B, Ding L, Cai Z, Song H, Zhao S, Xu H. Identification of SERPINA1 promoting better prognosis in papillary thyroid carcinoma along with Hashimoto's thyroiditis through WGCNA analysis. *Front Endocrinol (Lausanne)*. 2023;14:1131078. <https://doi.org/10.3389/fendo.2023.1131078>.
 18. Lin J, Cai B, Lin Q, Lin X, Wang B, Chen X. TLE4 downregulation identified by WGCNA and machine learning algorithm promotes papillary thyroid carcinoma progression via activating JAK/STAT pathway. *J Cancer*. 2024;15(14):4759–76. <https://doi.org/10.7150/jca.95501>.
 19. Landa I, Ibrahimpasic T, Boucai L, Sinha R, Knauf JA, Shah RH, Dogan S, Ricarte-Filho JC, Krishnamoorthy GP, Xu B, Schultz N, Berger MF, Sander C, Taylor BS, Ghossein R, Ganly I, Fagin JA. Genomic and transcriptomic hallmarks of poorly differentiated and anaplastic thyroid cancers. *J Clin Invest*. 2016;126(3):1052–66. <https://doi.org/10.1172/jci85271>.
 20. Wang Z, Wang H, Zhou Y, Li L, Lyu M, Wu C, He T, Tan L, Zhu Y, Guo T, Wu H, Zhang H, Sun Y. An individualized protein-based prognostic model to stratify pediatric patients with papillary thyroid carcinoma. *Nat Commun*. 2024;15(1):3560. <https://doi.org/10.1038/s41467-024-47926-w>.

21. Zhang Q, Li J, Shen H, Bai X, Zhang T, Liu P. Screening and validation of lymph node metastasis risk-factor genes in papillary thyroid carcinoma. *Front Endocrinol (Lausanne)*. 2022;13:991906. <https://doi.org/10.3389/fendo.2022.991906>.
22. Yan C, Sun J, He X, Jia L. An age-and sex-matched postoperative therapy should be required in thyroid papillary carcinoma. *Front Endocrinol (Lausanne)*. 2024;15:1339191. <https://doi.org/10.3389/fendo.2024.1339191>.
23. Liu C, Zhou Y, Zhou Y, Tang X, Tang L, Wang J. Identification of crucial genes for predicting the risk of atherosclerosis with system lupus erythematosus based on comprehensive bioinformatics analysis and machine learning. *Comput Biol Med*. 2023;152:106388. <https://doi.org/10.1016/j.compbiomed.2022.106388>.
24. Li S, Long Q, Nong L, Zheng Y, Meng X, Zhu Q. Identification of immune infiltration and cuproptosis-related molecular clusters in tuberculosis. *Front Immunol*. 2023;14:1205741. <https://doi.org/10.3389/fimmu.2023.1205741>.
25. Tan X, Zhang N, Zhang G, Xu S, Zeng Y, Bian F, Tang B, Wang H, Fan J, Bo X, Fu Y, Fan H, Zhou Y, Kang P. Exploring the diagnostic and immune infiltration roles of disulfidptosis related genes in pulmonary hypertension. *Respir Res*. 2024;25(1):365. <https://doi.org/10.1186/s12931-024-02978-w>.
26. An X, Zeng X, Yi Z, Cao M, Wang Y, Yu W, Ren Z. Integrative bioinformatics and machine learning identify iron metabolism genes MAP4, GPT, and HIRIP3 as diagnostic biomarkers and therapeutic targets in Alzheimer's disease. *Front Cell Neurosci*. 2025;19:1610682. <https://doi.org/10.3389/fncel.2025.1610682>.
27. Lin R, Fogarty CE, Ma B, Li H, Ni G, Liu X, Yuan J, Wang T. Identification of ferroptosis genes in immune infiltration and prognosis in thyroid papillary carcinoma using network analysis. *BMC Genomics*. 2021;22(1):576. <https://doi.org/10.1186/s12864-021-07895-6>.
28. He W, Sun Y, Ge J, Wang X, Lin B, Yu S, Li Y, Hong S, Xiao H. STRA6 regulates tumor immune microenvironment and is a prognostic marker in BRAF-mutant papillary thyroid carcinoma. *Front Endocrinol (Lausanne)*. 2023;14:1076640. <https://doi.org/10.3389/fendo.2023.1076640>.
29. An X, Wang Y, Cao M, Yi Z, Zeng X, Yu W, Ren Z. Exploring BSCL2 and associated genes in Alzheimer's disease by integrative analysis of bioinformatics, sn-RNAseq and machine learning approach. *J Alzheimers Dis Rep*. 2026;10:25424823261423079. <https://doi.org/10.1177/25424823261423079>.
30. Dan X, Babbar M, Moore A, Wechter N, Tian J, Mohanty JG, Croteau DL, Bohr VA. DNA damage invokes mitophagy through a pathway involving Spata18. *Nucleic Acids Res*. 2020;48(12):6611–23. <https://doi.org/10.1093/nar/gkaa393>.
31. Zhan ZJ, Li N, Sun Y, Chen L, Yin JD, Shan YH, Zeng JX, Li ZC, Tan HL, Tang N, Chang S, Huang P. Single-Cell Transcriptomics Reveals ITGA2-Mediated Metabolic Reprogramming and Immune Crosstalk in Pediatric Thyroid Carcinogenesis. *Adv Sci (Weinh)*. 2025;12(45):e04088. <https://doi.org/10.1002/advs.202504088>.

32. Yao B, Zhang Q, Yang Z, An F, Nie H, Wang H, Yang C, Sun J, Chen K, Zhou J, Bai B, Gu S, Zhao W, Zhan Q. CircEZH2/miR-133b/IGF2BP2 aggravates colorectal cancer progression via enhancing the stability of m(6)A-modified CREB1 mRNA. *Mol Cancer*. 2022;21(1):140. <https://doi.org/10.1186/s12943-022-01608-7>.
33. Xia T, Dai XY, Sang MY, Zhang X, Xu F, Wu J, Shi L, Wei JF, Ding Q. IGF2BP2 Drives Cell Cycle Progression in Triple-Negative Breast Cancer by Recruiting EIF4A1 to Promote the m6A-Modified CDK6 Translation Initiation Process. *Adv Sci (Weinh)*. 2024;11(1):e2305142. <https://doi.org/10.1002/advs.202305142>.
34. Sun J, Shi R, Zhang X, Fang D, Rauch J, Lu S, Wang X, Käsmann L, Ma J, Belka C, Su C, Li M. Characterization of immune landscape in papillary thyroid cancer reveals distinct tumor immunogenicity and implications for immunotherapy. *Oncoimmunology*. 2021;10(1):e1964189. <https://doi.org/10.1080/2162402x.2021.1964189>.
35. Yang P, Gao F, Peng S, Wei G, Bao G, Yuan L. Discriminating epigenetic landscapes: multi-omics characterization of benign thyroid nodules versus papillary thyroid carcinomas. *Thyroid Res*. 2025;18(1):49. <https://doi.org/10.1186/s13044-025-00267-2>.

Figures

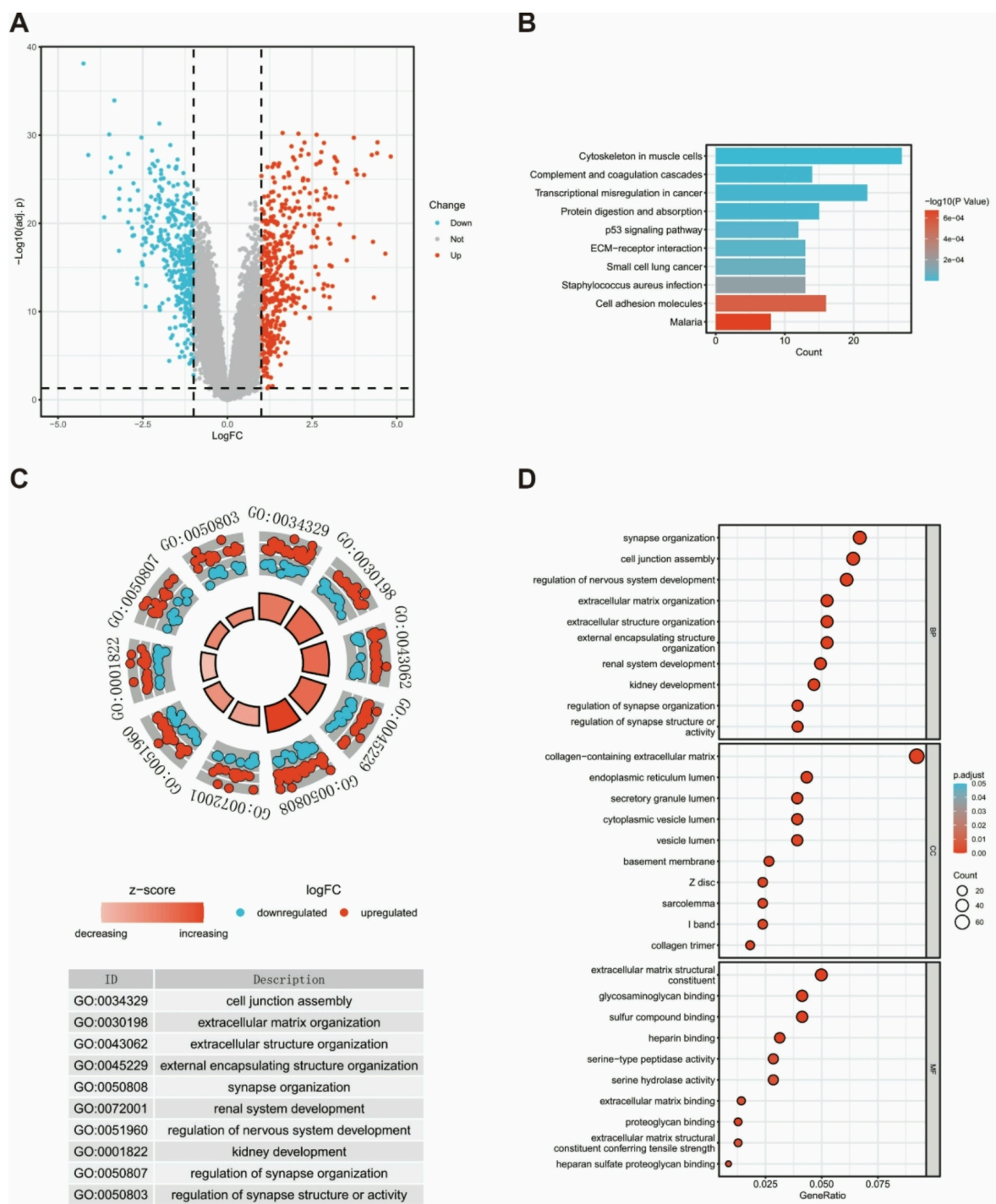


Figure 1

Transcriptomic Profiling Unveils Key Pathways in Thyroid Cancer Pathogenesis.

(A) Differentially expressed genes in thyroid cancer vs normal tissues $|\text{LogFC}| > 1$ adjusted $p < 0.05$ red upregulated blue downregulated.

- (B) KEGG enrichment DEGs enriched in ECM organization cancer pathogenesis and immune response pathways.
- (C) Biological process enrichment shows key pathways significance trend and expression change highlighting ECM and synaptic organization.
- (D) GO enrichment genes involved in ECM/synaptic organization serine peptidase activity and glycosaminoglycan binding.

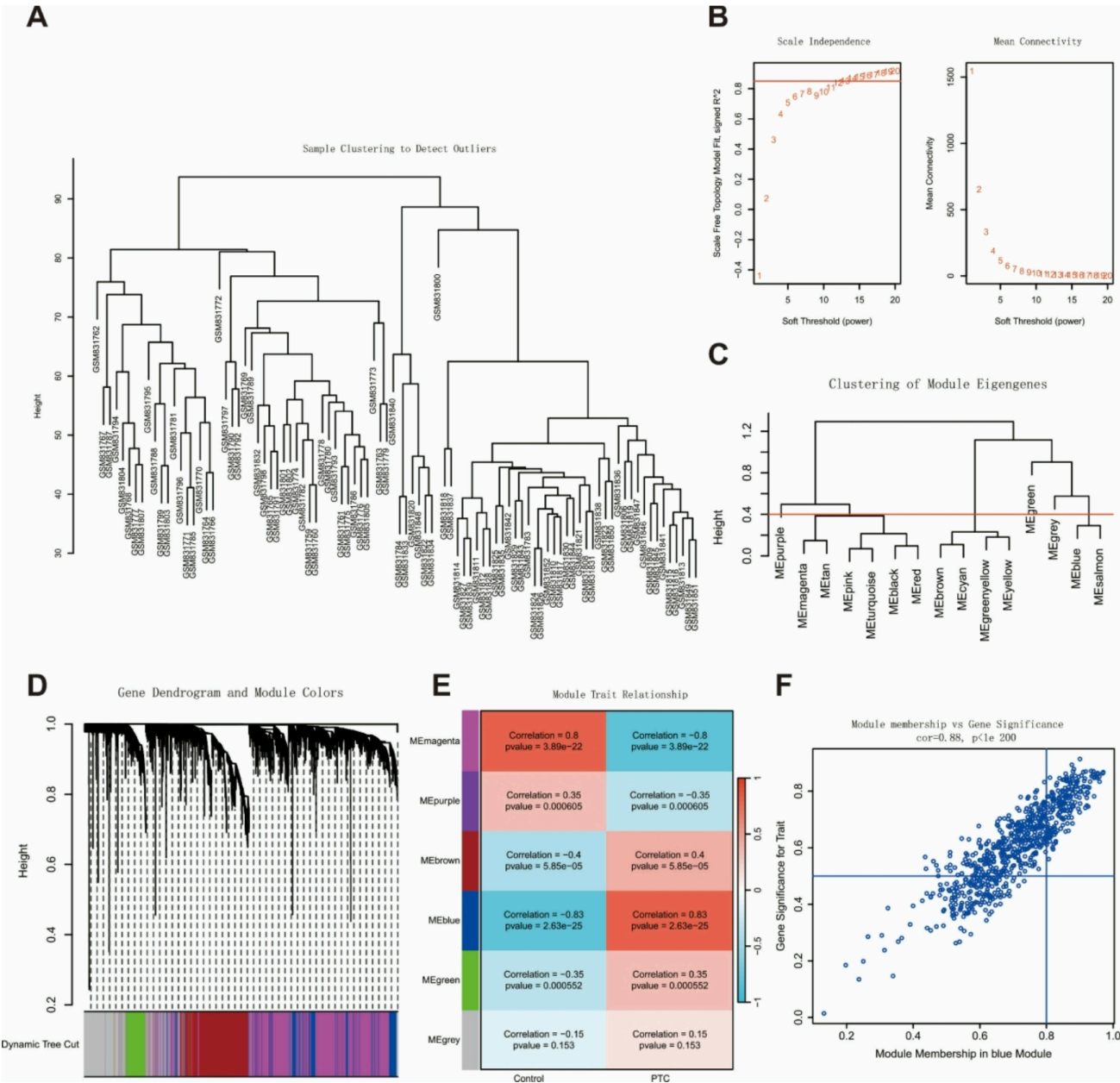


Figure 2

Weighted Gene Co-expression Network Analysis Identifies Clinically Relevant Gene Modules in Papillary Thyroid Cancer.

- (A) Sample clustering shows no outliers detected, all samples retained.
- (B) WGCNA soft threshold power of 13 selected for scale-free network construction.
- (C) Module eigengene clustering reveals biological relationships between modules.
- (D) Module detection by dynamic tree cutting identified distinct co-expression modules.
- (E) Module-trait correlation identifies MEblue and MEmagenta modules significantly correlated with PTC.
- (F) Blue module validation confirms core genes strongly associated with PTC.

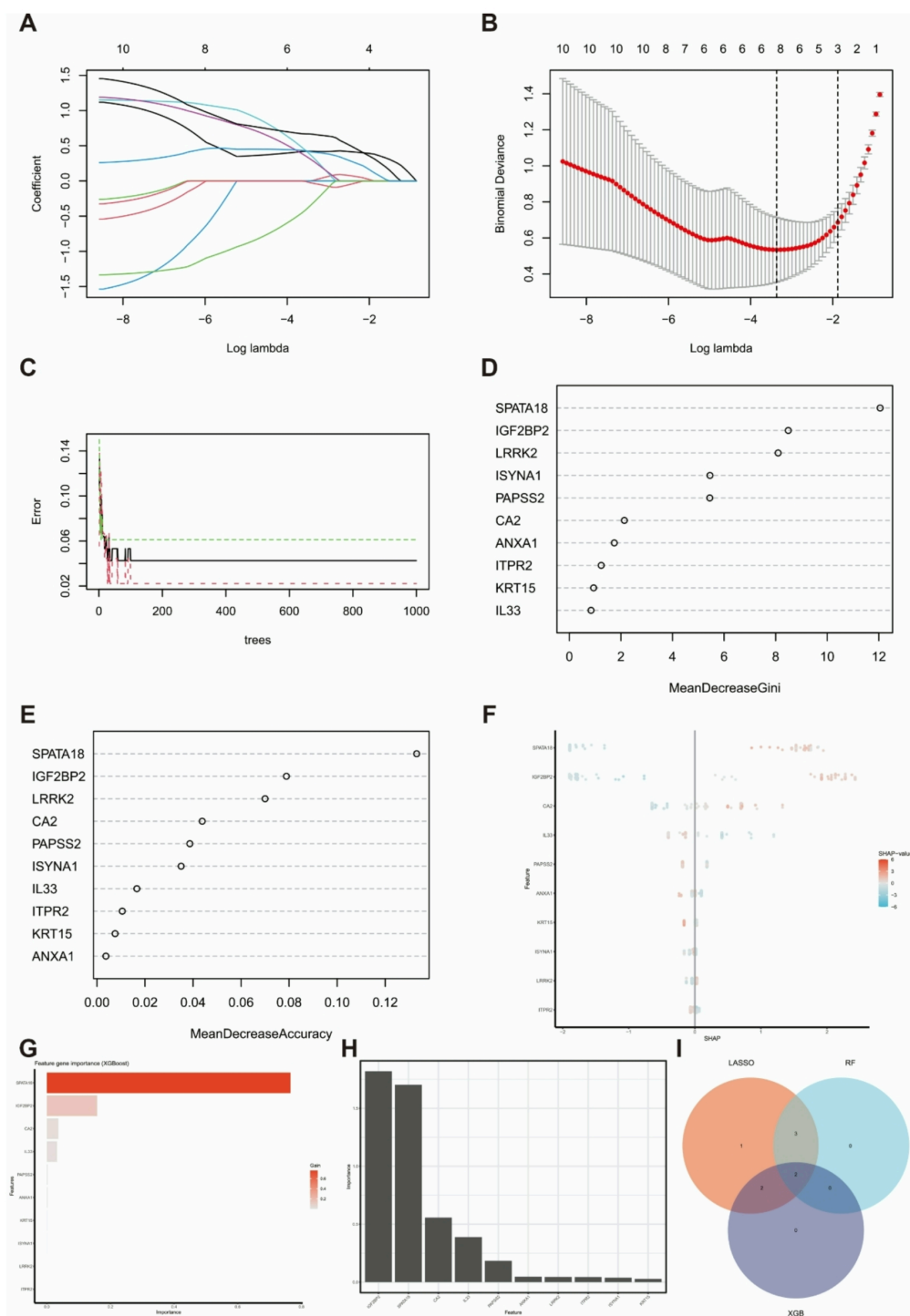


Figure 3

Machine Learning-Based Feature Selection Identifies Robust Prognostic Gene Signatures in Thyroid Cancer.

(A) LASSO regression feature selection Gene coefficients shrink with increasing regularization penalty Optimal λ selected by cross-validation for parsimonious prognostic features.

- (B) LASSO cross-validation curve Optimal λ selected at minimum deviance Top line shows non-zero coefficient count at each λ .
- (C) Random forest error rate vs number of trees Error converges with more trees demonstrating model robustness.
- (D) Top 10 random forest genes by Mean Decrease Gini ranking classification contribution.
- (E) Top 10 random forest genes by Mean Decrease Accuracy ranking prediction accuracy impact.
- (F) XGBoost SHAP summary plot Genes ordered by mean |SHAP| importance Colored by feature value red high expression blue low expression.
- (G) Top 10 XGBoost genes by Gain importance quantifying predictive accuracy contribution.
- (H) SHAP global feature importance Confirms SPATA18 and IGF2BP2 as most critical predictors.
- (I) Consensus genes from LASSO RF and XGBoost Central intersection identifies most robust candidates Prioritized for final prognostic signature.

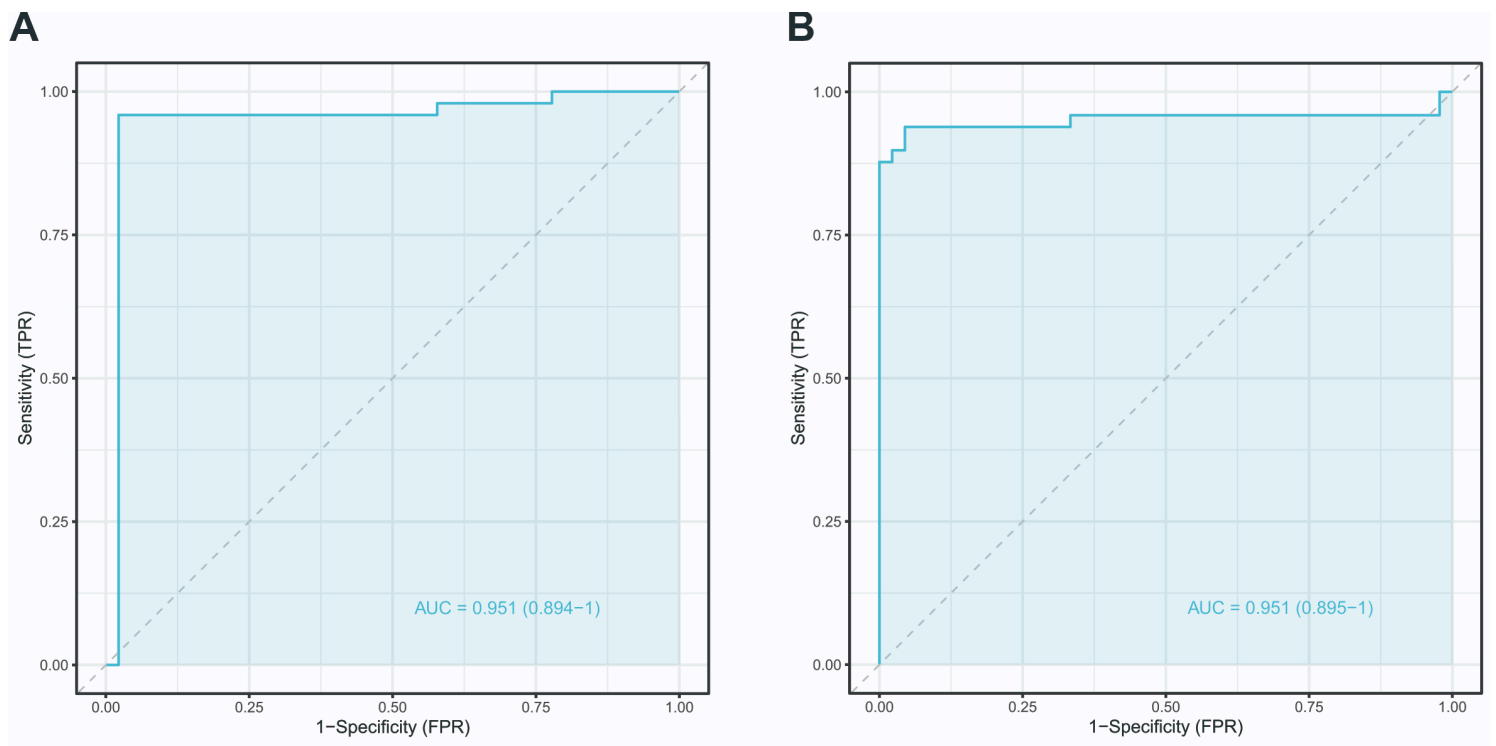


Figure 4

Diagnostic Performance Validation of Candidate Biomarkers Using ROC Curve Analysis

(A) SPATA18 ROC curve analysis AUC 0.951 excellent diagnostic power for thyroid cancer in GSE33630

(B) IGF2BP2 ROC curve analysis AUC 0.951 excellent diagnostic biomarker for thyroid cancer in GSE33630.

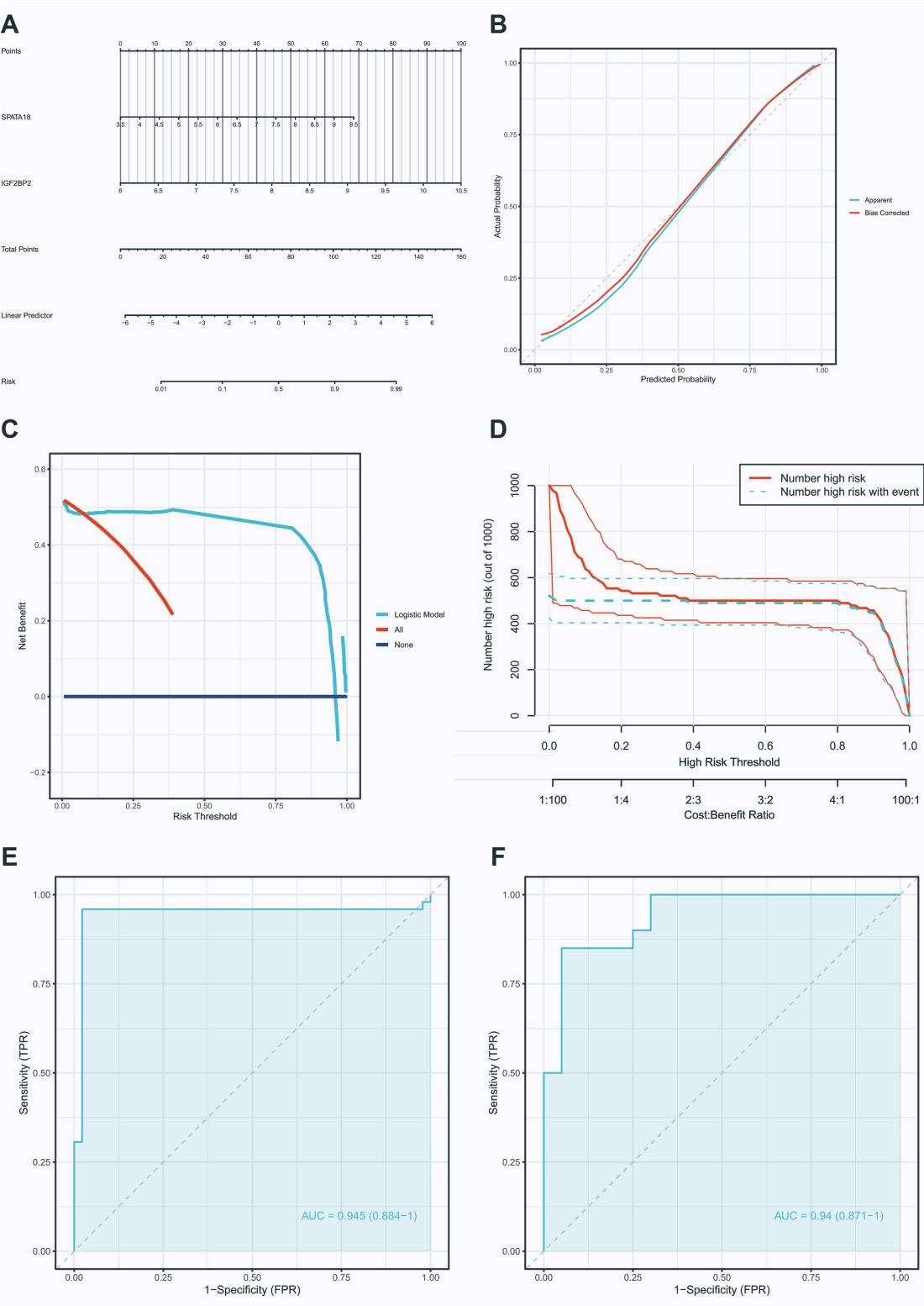


Figure 5

Construction and Comprehensive Validation of a Prognostic Nomogram for Thyroid Cancer Risk Prediction.

- (A) PTC risk prediction nomogram based on SPATA18 and IGF2BP2 Total points correspond to risk probability 0.01-0.99 for rapid assessment.
- (B) PTC risk model calibration curve Curves align closely with ideal line Strong accuracy minimal overfitting.
- (C) PTC risk model decision curve analysis Higher net benefit than all or none strategies across clinical thresholds.
- (D) PTC risk model clinical impact curve Simulated 1000 patients Shows high-risk individuals and true events across thresholds.
- (E) Training set ROC curve GSE33630 AUC 0.945 95% CI 0.884–1.000 Excellent discrimination.
- (F) Validation set ROC curve GSE29265 AUC 0.94 95% CI 0.871–1.000 Outstanding performance.

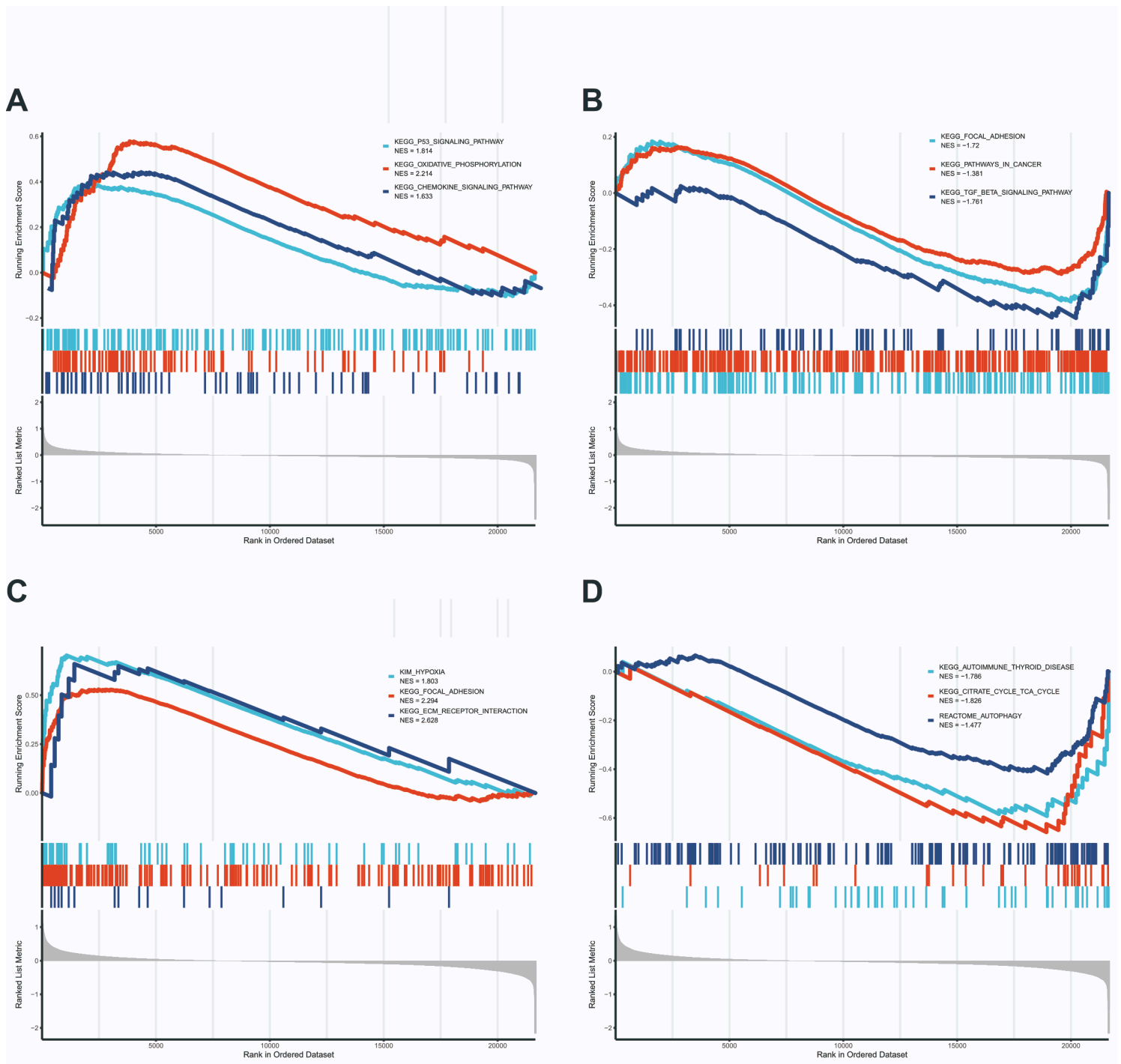


Figure 6

Gene Set Enrichment Analysis Reveals Pathway Regulation by SPATA18 and IGF2BP2 in Papillary Thyroid Cancer.

(A) GSEA of high SPATA18 expression Enriched pathways oxidative phosphorylation NES=2.21 p53 signaling NES=1.81 chemokine signaling NES=1.63 Positive association in PTC.

(B) GSEA of low SPATA18 expression Downregulated pathways TGF-beta signaling NES=-1.76 focal adhesion NES=-1.72 pathways in cancer NES=-1.38 Negative association in PTC.

(C) GSEA of high IGF2BP2 expression Enriched pathways ECM-receptor interaction NES=2.63 focal adhesion NES=2.29 hypoxia response NES=1.80 Implicated in ECM and microenvironment regulation.

(D) GSEA of low IGF2BP2 expression Downregulated pathways autoimmune thyroid disease NES=-1.786 TCA cycle NES=-1.826 autophagy NES=-1.477 Negative association in PTC.

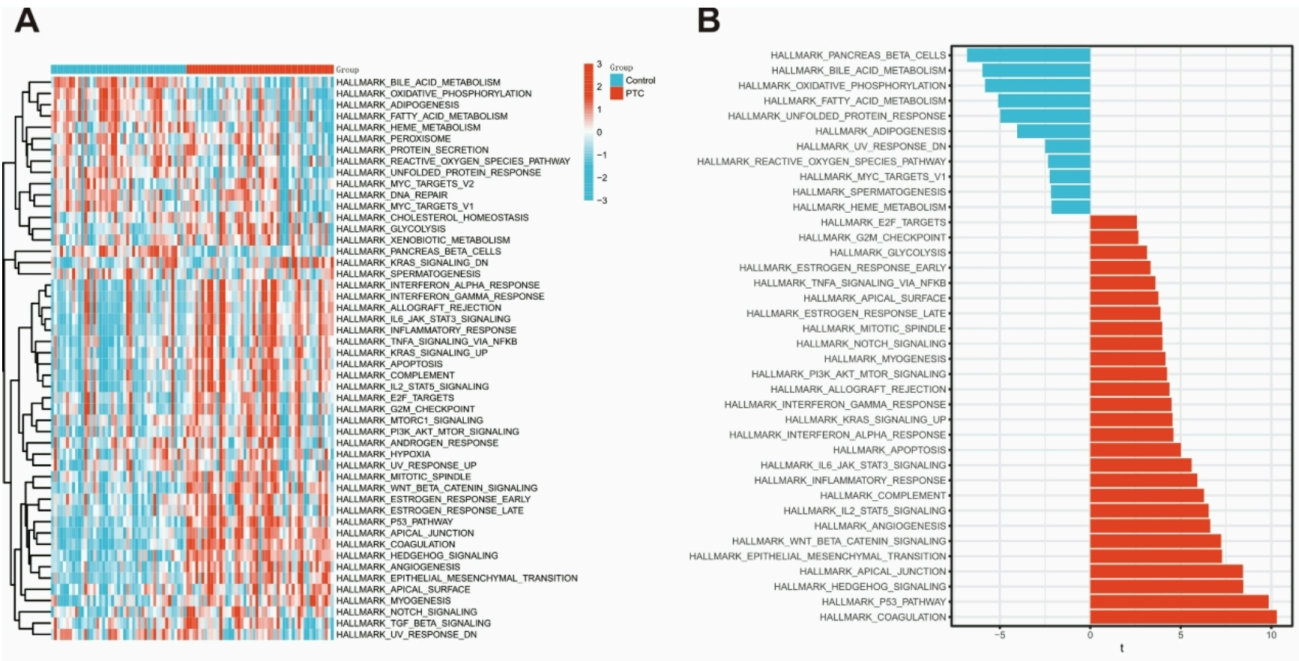


Figure 7

Gene Set Variation Analysis Reveals Global Pathway Dysregulation in Thyroid Cancer.

(A) GSVA heatmap of HALLMARK pathway activities Shows 50 gene set enrichment scores Highlights pathway activity differences between sample groups.

(B) Key dysregulated pathways by GSVA Bar plot ranks altered pathways by enrichment score Reveals inhibition and activation patterns Implicates metabolic reprogramming and pro-tumorigenic signaling.

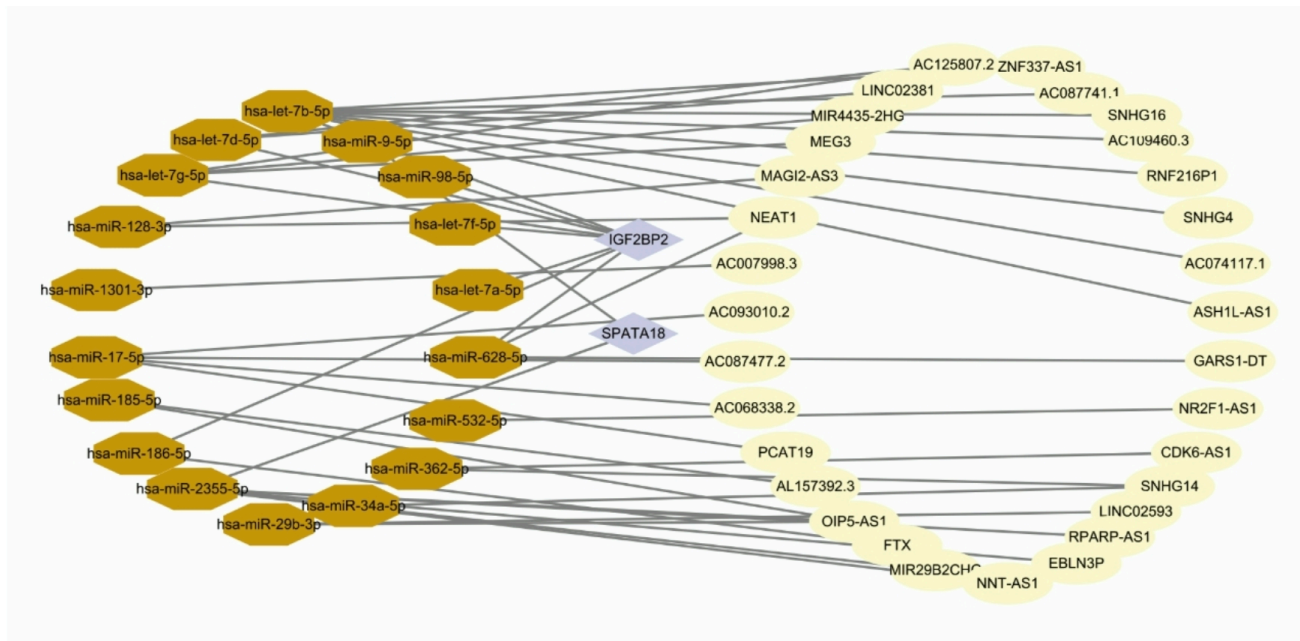


Figure 8

Core lncRNA-miRNA interaction network.

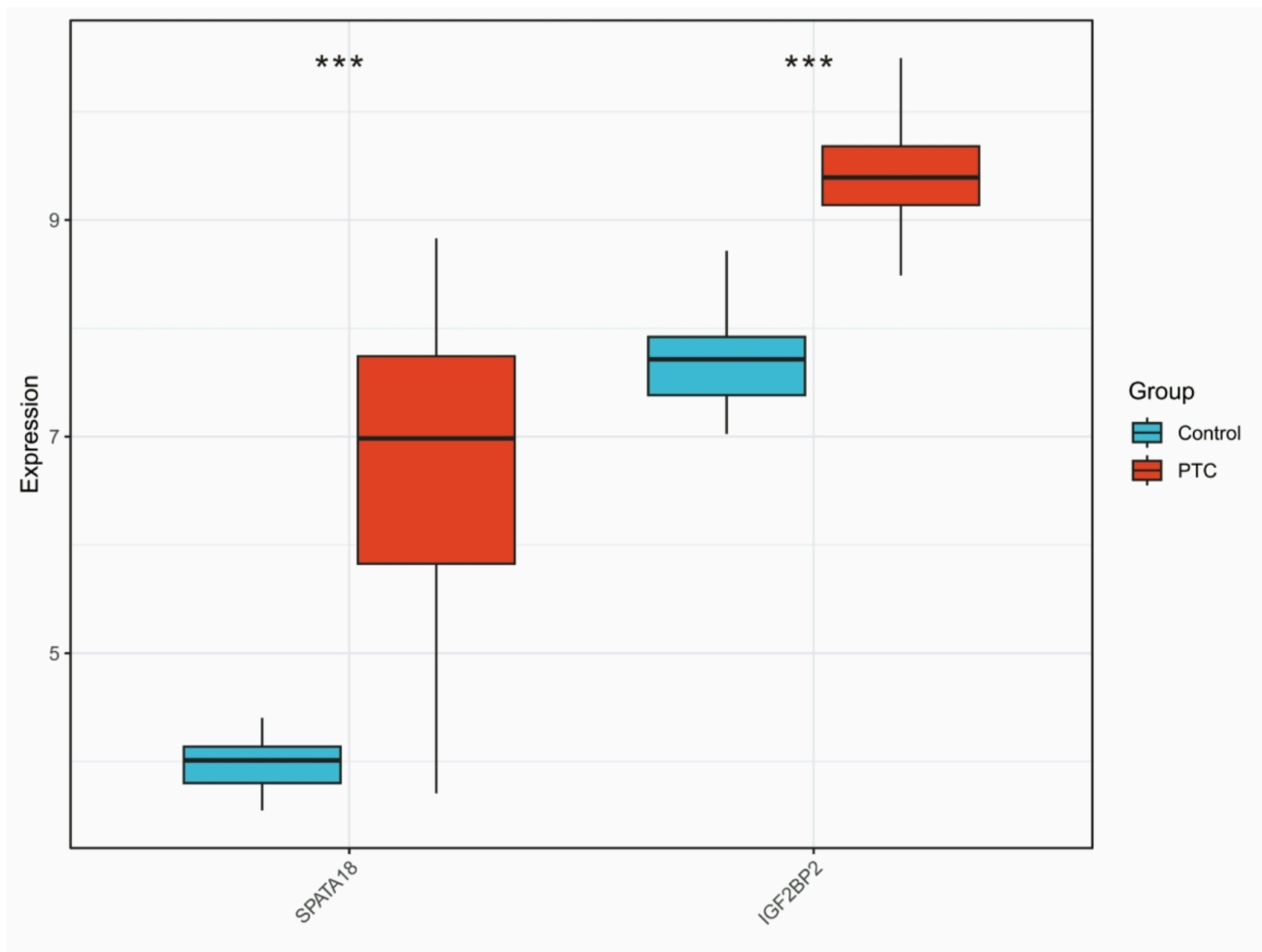


Figure 9

Differential Expression Validation of Key Biomarkers in Thyroid Cancer (p < 0.001).**