International Journal of General Medicine

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Identification of Novel Gene Signature Predicting Lymph Node Metastasis in Papillary Thyroid Cancer via Bioinformatics Analysis and in vitro Validation

Hai Li[1,2], Dongnan Sun[3], Kai Jin[4], Xudong Wang[1]

[1]Department of Maxillofacial and Otorhinolaryngological Oncology, Tianjin Medical University Cancer Institute and Hospital, Key Laboratory of Basic and Translational Medicine on Head & Neck Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin Cancer Institute, National Clinical Research Center of Cancer, Tianjin Medical University, Tianjin, People's Republic of China; [2]Department of General Surgery, Baotou Mongolian Medicine and Traditional Chinese Medicine Hospital, Baotou, Inner Mongolia Autonomous Region, People's Republic of China; [3]Center for Translational Medicine, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai, People's Republic of China; [4]Department of Thyroid Neoplasms Surgery, Inner Mongolia Autonomous Region People's Hospital, Hohhot, Inner Mongolia Autonomous Region, People's Republic of China

Correspondence: Xudong Wang, Email wxd.1133@163.com

International Journal of General Medicine downloaded from https://www.dovepress.com/
For personal use only.

Background: Although with a good prognosis of papillary thyroid cancer (PTC), patients with PTC and also experiencing lymph node metastasis (LNM) had higher recurrence and mortality rates. Therefore it was essential to explore novel biomarkers or methods to predict and evaluate the situation in the stages of PTC.

Methods: In this study, mRNA sequence datasets from The Cancer Genome Atlas (TCGA) database and Gene Expression Omnibus (GEO) were utilized to obtain differentially expressed genes (DEGs) between PTC tumors and normal specimens and DEGs related to lymph node metastasis were identified using weighted gene co-expression network analysis (WGCNA) according to the clinical information. Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were applied to query the biological functions and pathways. Furthermore, a protein-protein interaction (PPI) network was constructed using a STRING database and a prognosis model was established using the least absolute shrinkage and selection operator (LASSO) Cox regression analysis based on the LNM-related DEGs. Finally, six hub genes were identified and verified in vitro experiments.

Results: A novel six-gene signature model including COL8A2, MET, FN1, MPZL2, PDLIM4 and CLDN10 was established based on a total of 52 DEGs from the intersection of LNM-related modules identified by WGNCA from TCGA, THCA and GSE60542 to predict the situation of lymph node metastasis in PTC. Those six hub genes were all more highly expressed in PTC tumors and played potential biological functions on the development of PTC in in vitro experiments, which had potential values as diagnostic and therapeutic targets.

Keywords: papillary thyroid cancer, lymph node metastasis, bioinformatics analysis, WGCNA

Introduction

The thyroid gland, consisting of two connected lobes, is one of the largest endocrine glands in the human body, weighing 20–30 g in adults. Thyroid lesions are mostly asymptomatic and do not affect thyroid hormone secretion, they are often found on the gland, with a prevalence of 4–7%.[1] As the most common malignant tumor in the head and neck, thyroid cancer (TC) is becoming increasingly prevalent owing to the development of ultrasound technology, and will become the second and third most common malignant tumor in female and male, respectively, by 2030.[2] In China, the incidence rate of TC is also increasing and it has become one of the top ten malignancy cancers threatening the health of people.[3] Papillary thyroid carcinoma (PTC) is the most common subtype of thyroid cancer, which has a great incidence rate in the endocrine tumor, accounting for approximately 90% of all TCs and about 40% of adult PTC patients have lymph node metastasis.[4] As an important risk factor for high recurrence and mortality rate, lymph node metastasis (LNM), including

Received: 18 November 2024
Accepted: 21 February 2025
Published: 15 March 2025

© 2025 Li et al. This work is published and licensed by Dove Medical Press Limited. The full terms of this license are available at https://www.dovepress.com/terms.php and incorporate the Creative Commons Attribution – Non Commercial (unported, v3.0) License (http://creativecommons.org/licenses/by-nc/3.0/). By accessing the work you hereby accept the Terms. Non-commercial uses of the work are permitted without any further permission from Dove Medical Press Limited, provided the work is properly attributed. For permission for commercial use of this work, please see paragraphs 4.2 and 5 of our Terms (https://www.dovepress.com/terms.php).

central and lateral cervical LNM, is an effective indicator of PTC prognosis.[5] However, there is still a lack of research about the underlying mechanism of the lymph node metastasis of PTC and there is an urgent need to identify reliable biomarkers to predict the occurrence of LNM on PTC.

With the development of diagnostic techniques, especially the widespread application of ultrasonography, the over-diagnosis of LNM of PTC had become a prominent issue leading to total thyroidectomy and lymph node dissection from PTC patients, resulting in severe overtreatment.[6] Therefore, it was essential to conduct accurate assessments of the prognosis of cancer for the high-risk PTC patients. Identification of specific biomarkers related to PTC metastasis had become feasible due to high-throughput sequencing and bioinformatics technology development and several studies have screened gene signatures predicting the LNM of PTC. Kun Wang et al systematically evaluated the expression and potential mechanisms of the transmembrane 4 superfamily (TM4SF) and revealed that a five-gene prognostic model involved in TM4SFs could assess PTC progression effectively.[7] Wenjie Chen et al identified the plasma exosomal miRNAs, miR-6774-3p and miR-6879-5p as potential diagnostic markers of LNM in PTC via miRNA microarray and quantitative real-time polymerase chain reaction (qRT-PCR) and proved that those two miRNAs and their combination had better diagnostic performance than that of total circulating miRNAs to distinguish PTC-N1 patients from PTC-N0 patients using receiver operating characteristic (ROC) curves analysis.[8] Yang Yu et al established a three hub-gene model including PTGS, MET and ICAM1 from immune-related genes (IRGs) by least absolute shrinkage and selection operator (LASSO) and Random Forest methods and validated that those signature immune genes had potential functions for lymphatic metastasis in papillary thyroid cancer.[9] However, there were more studies needed to provide strong enough evidence to prove the accuracy of the model and the impact of key signatures on lymphatic metastasis in PTC.

In this study, we explored RNA-seq data from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) datasets comprehensively with clinical information by weighted gene co-expression network analysis (WGCNA) and other bioinformatic methods and identified key modules associated with lymph node metastasis of PTC. A risk score model was constructed to predict the prognosis of PTC and six hub genes including COL8A2, MET, FN1, MPZL2, PDLIM4 and CLDN10 might have a functional impact on PTC lymphatic metastasis by in vitro experimental verification, which might be novel therapeutic targets for clinical diagnosis and treatment of PTC.

Materials and Methods

Tissue Samples and Cell Lines

Three groups namely normal, non-LNM and LNM including 12 tissue samples respectively were collected from Tianjin Medical University Cancer Institute & Hospital and used for immunohistochemical (IHC) analysis. K1, TPC-1 and B-CPAP cell lines were purchased from the American Type Culture Collection (ATCC). Tianjin Medical University Cancer Institute & Hospital granted ethical approval to carry out this study within its facilities (Ethical Application Ref: bc20240045) and all the patients from who tissue samples were separated have written informed consents in accordance with the Declaration of Helsinki.

Data Collection and Preprocessing

The workflow of this study is shown in Figure 1. The thyroid carcinoma (THCA) RNA-seq datasets including 495 tumor samples and corresponding clinical information were downloaded from The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/repository) for WGCNA analysis after deleting duplicate samples. Three GEO datasets, namely GSE3467, GSE3678 and GSE60542, which were on the same annotation platform GLP570, were downloaded from GEO (http://www.ncbi.nlm.nih.gov/geo) to identify the differentially expressed genes (DEGs) between PTC tumors and normal samples.

Differentially Expressed Genes Screening

R package "limma" was used to screen the differentially expressed genes in the three GEO datasets (GSE3467/GSE3678/GSE60542) setting the threshold value of $|\log FC|>1.5$ and $p < 0.05$. A volcano plot was plotted by using R package "ggplot2". The statistical power of GSE3467, GSE3678 and GSE60542 was respectively 0.97, 0.92 and 0.99 calculated by R package "RNASeqPower".

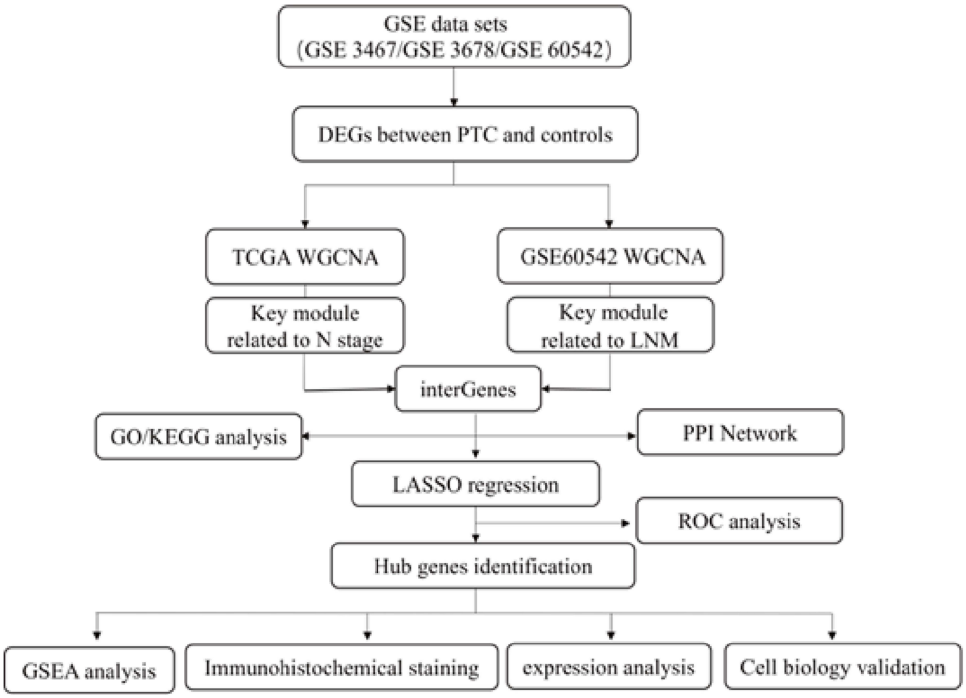

Figure 1 Workflow of study analysis.

Weighted Gene Co-Expression Network Construction

All DEGs from three GEO datasets were prepared for further WGCNA based on 495 tumor samples from TCGA, THCA and 38 tumor samples including 24 N1-LNM and 14 N0 tumor samples from GSE60542 separately with detailed pathological T and N stage clinical information according to a process described previously.[10] Outlier samples were removed by sample clustering and soft threshold power was set as 6 for TCGA and 12 for GSE60542 using mean connectivity and scale independence analysis. Then gene modules with similar expression profiles setting module size >30 were built followed by transforming the adjacency matrix of genes into a topological overlap matrix (TOM). Therefore, the gene modules associated with LNM of TCGA and GSE 60542 were identified by Pearson's correlation analysis with $p < 0.01$.

GO and KEGG Functional Analysis

The inter-genes from WGCNA analysis of TCGA and GSE60542 underwent gene ontology (GO) analysis including cellular component (CC), molecular function (MF), and biological process (BP) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis[11] by R package "clusterProfiler"[12] with the statistical significance set to $p < 0.05$ and minimum count >3.

PPI Network Construction

The protein-protein interaction (PPI) networks of the inter-genes from WGCNA analysis were constructed by STRING (http://string-db.org)[13] after hiding the disconnected nodes by setting 0.4 as the interaction score of the networks. The network was further analyzed by the program Cytoscape and visualized depending on the degree of nodes in the network.[14]

Prognostic Signature Construction

R package "glmnet" was used to analyze the inter-genes related with lymphatic metastasis by LASSO Cox regression analysis. The expression data of TCGA THCA was randomly divided into two partswith 70% as training data and the

remaining 30% data as a validation set. A 10-fold internal cross-validation was used, and the process was repeated 1000 times to evaluate the model.

ROC Validation of Prognostic Signature

R package "pROC" was used to plot the ROC curve of six gene signatures according to mRNA data of GSE 60542 to evaluate whether prognostic risk model had good diagnostic value or not.

Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) was implemented to explore the underlying biological function of the risk model and hub genes using R package "clusterProfiler". The mRNA data of TCGA THCA were grouped into high and low expression of the risk model and each hub gene and "c2.cp.kegg. v7.1. entrez.gmt" from the Molecular Signature Database was chosen as the reference gene set.

RT-qPCR

Total RNA of three THCA cell lines (K1, TPC-1 and B-CPAP) which were cultured using RPMI1640 (Gibco, USA) with 10% of fetal bovine serum (FBS) (ScienCell, USA) below 37 °C with 5% carbon dioxide was extracted using TRIzol reagent (Invitrogen, America, 15596018) and quantified by Nanodrop 2000 spectrophotometer (Thermo Fisher, America). Reverse transcription was conducted using ReverTra Ace qPCR RT Master Mix with gDNA Remove (TOYOBO, Japan, FSQ-301) and the cDNA generated was used for quantitative real-time PCR using 2* SYBR Green qPCR Master Mix (Low ROX) (Servicebio, China, G3321-15) and the conditions of reaction were as follows: initial denaturation at 95 °C for 3 min, followed by 40 cycles of 95 °C for 15 s, 60 °C for 30 s and 72 °C for 30 s using the ABI 7500 RT-PCR system (Applied Biosystems, America, 7500). The 2-ΔΔCt method was used to measure the mRNA expression with β-actin as internal reference. The primers of six hub genes are listed in Table 1.

Transwell Assay and Wound Healing

The function of migration and invasion that hub genes might impact was verified in the B-CPAP cell which was established from metastasizing papillary thyroid carcinoma and cultured in RPMI1640 (Gibco, USA) with 10% of fetal bovine serum (FBS) (ScienCell, USA) below 37 °C with 5% carbon dioxide. Wound healing and transwell assays were conducted to estimate the migration and invasive abilities of B-CPAP cells after transfection with siRNAs and siControls of hub genes as a standard protocol. Briefly, siRNAs of hub genes and corresponding siControls were transfected into B-CPAP cells and 1.5×10^5 cells were seeded into the chamber and cultured for 24 hours. Then, the cells were fixed with methanol and stained with crystal violet solution. Meanwhile, 5×10^5 B-CPAP cells were seeded into 6-well plates and transfected as the same as transwell assay, then pictures were taken at 0-hour and 24-hour points after scratching using a pipette tip. The sequences of siRNAs of hub genes are listed in SI Table 1.

Table 1 The Primer Sequences of the Hub Genes

Gene Name	Forward Primer	Reverse Primer
β-actin	CATGTACGTTGCTATCCAGGC	CTCCTTAATGTCACGCACGAT
COL8A2	GAGCCAGGAATACGAGGGGA	TCCAGGGATAGTAATGCCTGAG
MET	TCCAGGGATAGTAATGCCTGAG	AGCAATGGGGAGTGTAAAGAGG
MPZL2	AGCAATGGGGAGTGTAAAGAGG	CCCAGTCTTGTACTCAGCAAC
PDLIM4	CCCAGTCTTGTACTCAGCAAC	GGAATCCTGAGCGGTACGATG
FN1	GGAATCCTGAGCGGTACGATG	CTGGCAGGTGTATGTCCCATT
CLDN10	CTGGCAGGTGTATGTCCCATT	GCTGCCACGATCACCTCAC

Immunohistochemical Staining (IHC)

Tissue wax samples of three groups namely normal, non-LNM and LNM including 12 tissue samples respectively were heated in citrate buffer for 15 min and rinsed using PBS after soaking in xylene for 40 min and in anhydrous ethanol for 20 min. Then the sections were incubated with primary antibodies including anti-MET (Proteintech, 25869-1-AP, PBS 1:500 dilution), anti-FN1 (Proteintech, 15613-1-AP, PBS1:300 dilution), anti-MPZL2 (Proteintech, 11787-1-AP, PBS 1:200 dilution), anti-PDLIM4 (ThermoFisher, bs-6093R, PBS 1:200 dilution), anti-CLDN10 (ThermoFisher, bs-13739R, PBS 1:200 dilution), and anti-COL8A2 (Affinity Biosciences, DF8903, PBS 1:100 dilution) at 4 °C overnight and goat anti-mouse/rabbit IgG polymer antibody (Immunoway, RS0011, PBS 1:100 dilution) was then added and incubated for 1 h at room temperature after three times washing. After PBS rinsing, they were incubated with 3,3-diaminobenzidine tetrahydrochloride (DAB, Beytime, P0202) for 5 min. The scores of IHC results were evaluated based on the intensity and area of dyeing. The intensity of dyeing ranged from 0 point as positive cells with no tinting; 1 point was light; 2 points was medium; and 3 points was strong. The area of dyeing followed the rules that the area <5% cell staining was 0 point; 6 ~25% cells staining was 1 point; 26~50% cells staining was 2 points; 51~75% cells staining was 3 points; and 76~100% cell staining was 4 points. Final results were the product of the intensity and area of dyeing with a score of 0–4 being negative and 5–12 was positive.

Statistical Analysis

Statistical analyses were carried out by R software version 3.5.3 and GraphPad Prism 9.0. The values of mean ± standard deviation (SD) of at least three individuals were expressed in the results. Student's *t*-test and the chi-square test were used in comparison between the two groups, and one-way analysis of variance (ANOVA) was used for multiple groups with two-sided $p < 0.05$ as a statistically significant difference.

Results

Identification of DEGs from GEO Datasets

As the threshold of adjusted P-value < 0.05 and |log FC| > 1.5, there were 265 significant DEGs including 129 up-regulated and 136 down-regulated DEGs for GSE 3467, 235 significant DEGs including 179 up-regulated and 56 down-regulated DEGs for GSE 3678 and 269 significant DEGs including 120 up-regulated and 149 down-regulated DEGs for GSE 60542 (Figure 2A–C). In total, 460 DEGs from all three GES datasets were used for the following analysis (Figure 2D). The heatmap of all 460 DEGs expressed in TCGA samples is shown in Figure 2E.

Identification Key Modules Related to LNM from Weighted Co-Expression Network

460 DEGs were used for WGCNA analysis for lymph node metastasis of HNSCC following differential analysis. For TCGA THCA, two modules, namely blue including 84 genes and turquoise including 201 genes were significantly related to lymph node metastasis with the appropriate soft threshold power β of 6 (Figure 3A–C) and the scatter diagrams of GS for N grade vs MM in the two modules were plotted, respectively (Figure 3D and E). For GSE 60542, the expression data of 24 N1-LNM samples and 14 N0 tumor samples were used to construct in WGCNA on lymph node metastasis with threshold power β of 12 and the brown module including 74 genes was identified as an LNM-related module (Figure 3F–H) and the genes in brown module were intensively correlated to LNM from the scatter diagram of GS for LNM vs MM (Figure 3I). The detailed genes were listed in SI Table 2, Finally, there were 52 interGenes in the intersection of these three modules (Figure 3J).

Functional Enrichment Analysis and PPI Network Construction of interGenes

The function of 52 interGenes was further investigate by performing GO and KEGG analysis with a p-value of 0.05 used as the significant difference. From the results of GO analysis, these genes were mainly enriched in cell junction assembly, epidermis development and skin development in biological process (BP) (Figure 4A), and apical plasma membrane, endoplasmic reticulum lumen and endoplasmic reticulum-Golgi intermediate compartment were the top 3 GO terms in cellular component (CC) (Figure 4B). In the aspect of molecular function (MF), these genes were enriched in peptidase

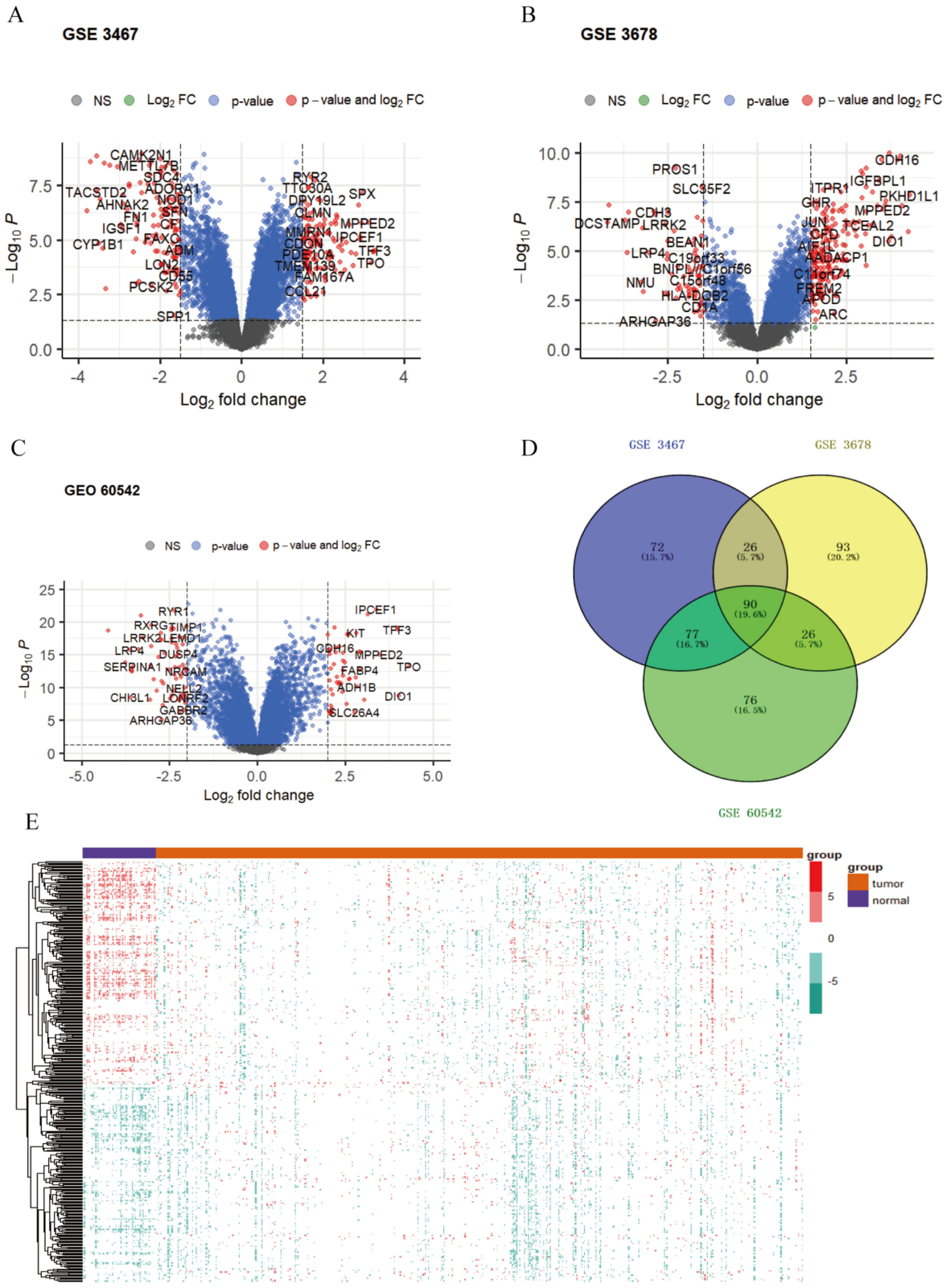

Figure 2 Identification of differentially expressed genes (DEGs) in GEO datasets. (**A**) Volcano plot showing DEGs in GEO 3467. (**B**) Volcano plot showing DEGs in GEO 3678. (**C**) Volcano plot showing DEGs in GEO 60542. (**D**) The union of DEGs from three GEO datasets. (**E**) Heatmap of 460 DEGs expression level in TCGA THCA.

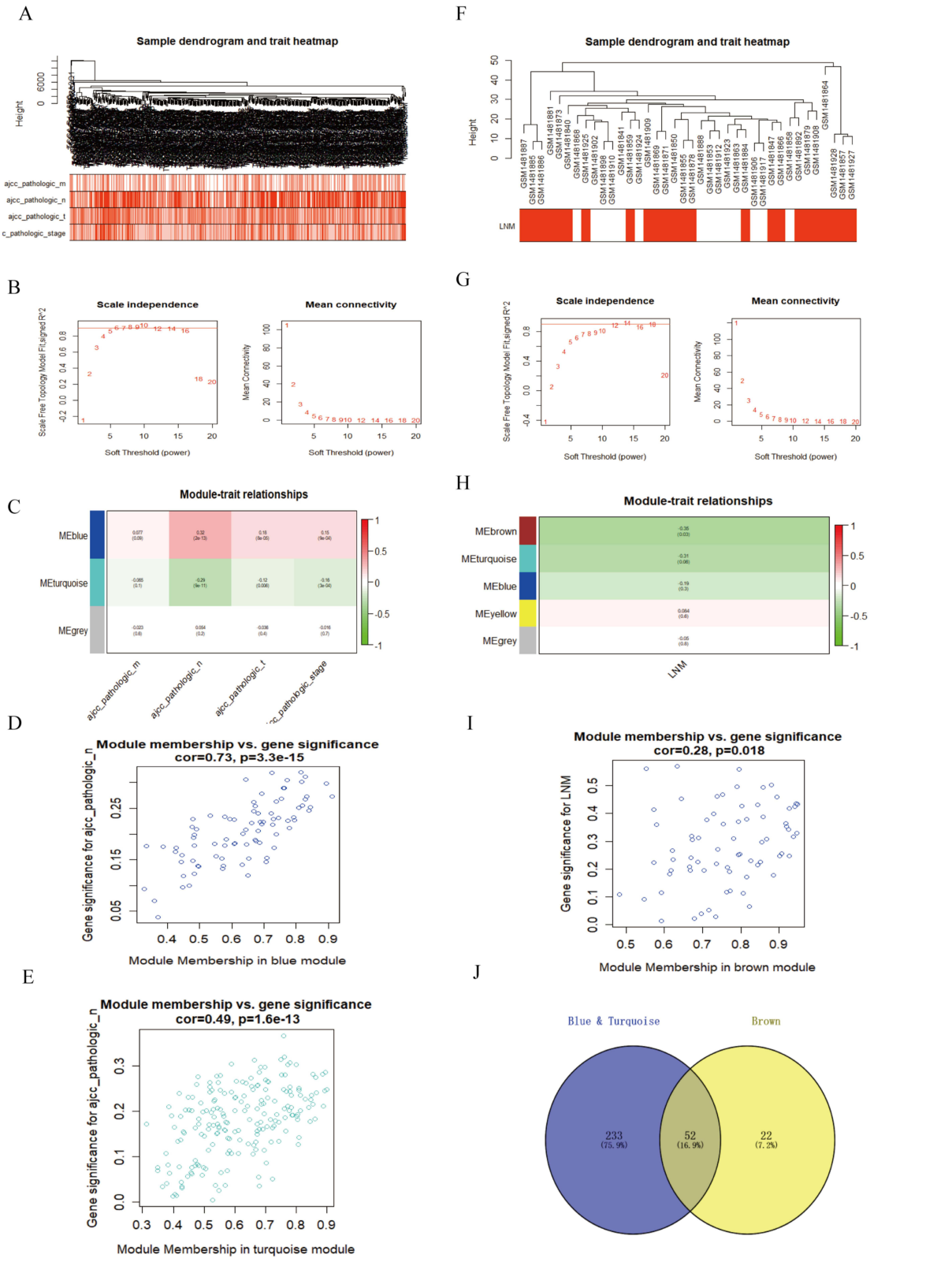

Figure 3 Weighted co-expression network analysis of TCGA and GEO databases. (**A**) Sample clustering of TCGA THCA. (**B**) Selection of optimal various soft-threshold powers in TCGA THCA. (**C**) Heatmap of the correlation between module eigengenes and N stage in TCGA THCA. (**D**) The correlation between blue module membership and N stage in TCGA THCA. (**E**) The correlation between turquoise module membership and N stage in TCGA THCA. (**F**) Sample clustering of GEO 60542. (**G**) Selection of optimal various soft-threshold powers in GEO 60542. (**H**) Heatmap of the correlation between module eigengenes and LMN in TCGA THCA. (**I**) The correlation between brown module membership and LMN in GEO 60542. (**J**) The intersection of genes in the blue, turquoise and brown modules.

https://doi.org/10.2147/IJGM.S502480

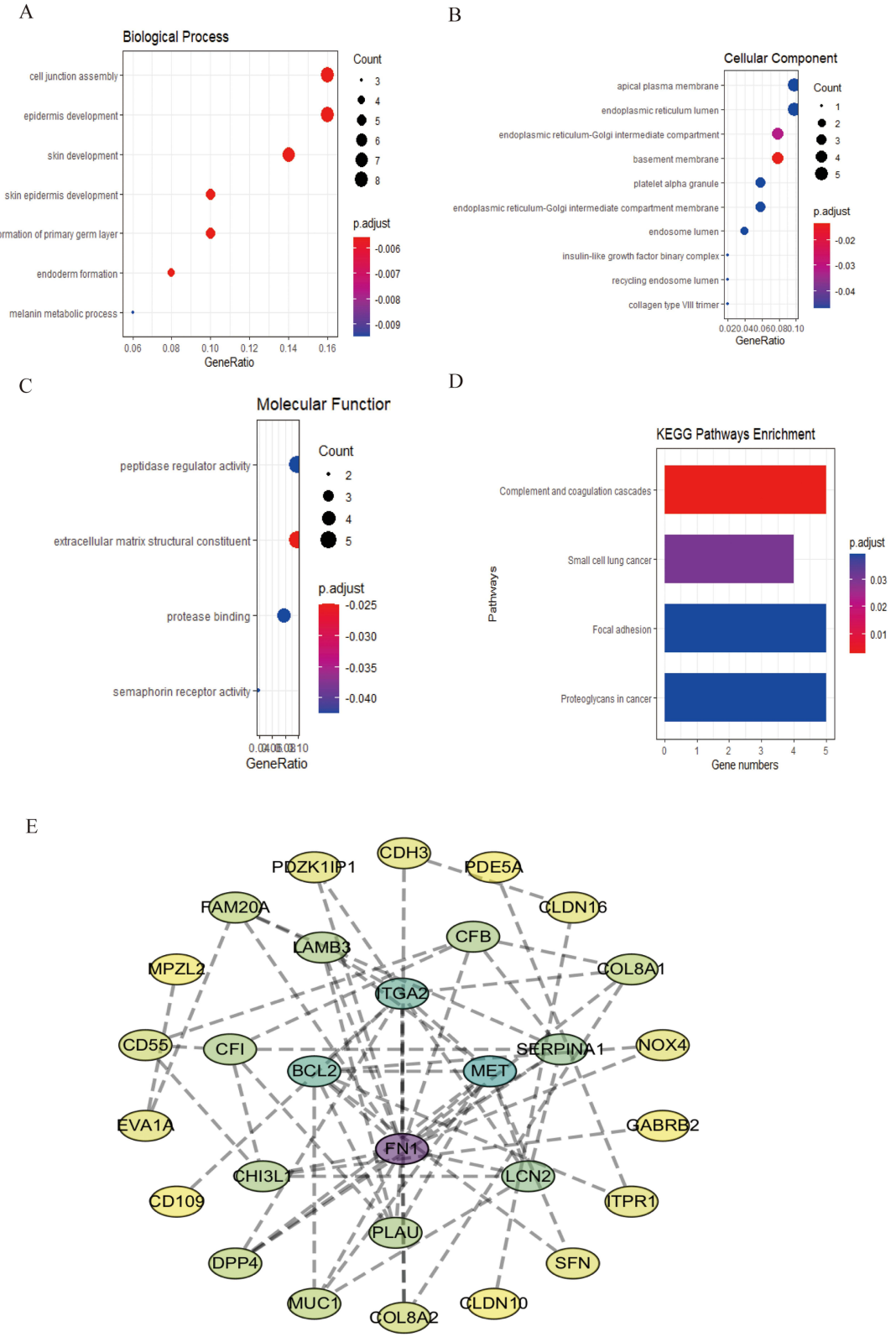

Figure 4 Enrichment analysis and PPI network construction of interGenes. (**A**) Analysis of biological process. (**B**) Analysis of cellular component. (**C**) Analysis of molecular function. (**D**) Analysis of KEGG pathway. (**E**) PPI network analysis of interGenes.

regulator activity, extracellular matrix structural constituent and protease binding (Figure 4C). For the KEGG pathway, complement and coagulation cascades, small cell lung cancer, focal adhesion and proteoglycans in cancer were the main pathways in which those genes were enriched, which were closely linked to the progression and metastasis of cancer (Figure 4D). The PPI network of those 52 interGenes was constructed using the STRING database, which included 52 nodes and 62 edges, and FN1 had the highest degree in the network (Figure 4E).

Construction and Validation of Prognostic Signature

Based on the 52 interGenes, we screened a further six genes including COL8A2, MET, FN1, MPZL2, PDLIM4 and CLDN10 using LASSO analysis based on TGCA (Figure 5A and B). To assess the diagnostic role of the prognostic risk

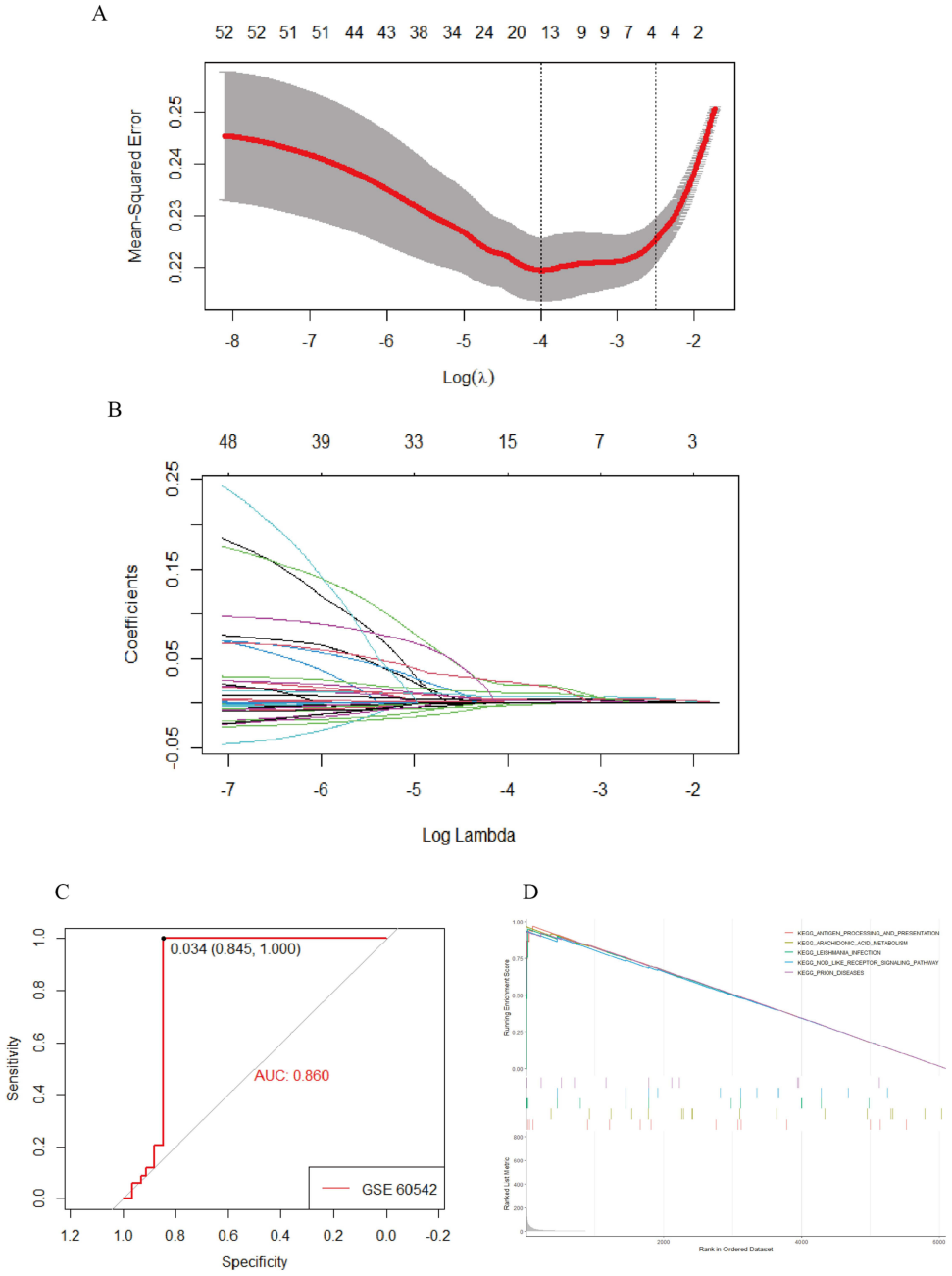

Figure 5 Construction of prognostic signature related to LNM in PTC. (**A**) Cross-validation of the parameter selection of LASSO regression. (**B**) The coefficient of LASSO regression. (**C**) ROC analysis of risk prognostic model based on GSE 60542. (**D**) GSEA analysis of TGCA THCA based on risk prognostic model.

https://doi.org/10.2147/IJGM.S502480

model, GSE 60542 was used to assess ROC and AUC values greater than 70% (86%), meaning that this prognostic risk model had good prognostic ability (Figure 5C). Using this prognostic risk model, the TCGA THCA database was grouped into high and low risk group, and five KEGG pathways (Antigen Processing and Presentation, Arachidonic Acid Metabolism, Nod-Like Receptor Signaling Pathway, Leishmania Infection And Prion Diseases) were enriched by GSEA analysis (Figure 5D).

Validating the Expression and Function of Hub Genes

To analyze the potential roles of those hub genes in PTC, the high and low expressions of hub genes in N0/N1 and different progress stages in the TCGA THCA database were compared. Results suggest that patients in N1 stage had a higher expression of all hub genes compared to N0 patients (Figure 6A) and the expression of all hub genes in stage I and stage II patients were also significantly higher than those in stage III and stage IV (Figure 6B), which were consistent with the results of WGCNA analysis and the relationship between N stage and LNM. Further to assess the differential expression of hub genes, GSE 60542 was used to compare the mRNA expression in the normal, N0 and N1 samples and the results indicated that all the hub genes were significantly highly expressed in tumor samples including N1 and N0 than normal tissue, also the expression of CLDN10, MET, MPZL2 and PDLIM4 in N1 were significantly higher than that in N0 (Figure 6C).

Furthermore, the mRNA expressions of those six hub genes were detected in K1, TPC-1 and B-CPAP cell lines using RT-qPCR. The relative expressions of MET, FN1, MPZL2 and PDLIM4 based on the expressions in the TPC-1 cell line indicated that all four genes were higher expressed in tumor cell lines apart from COL8A2 and CLDN10 which were hardly detected in each cell line (Figure 7A–D), in accordance with the low mRNA levels in samples from TCGA THCA. To determine how MET, FN1, MPZL2 and PDLIM4 impacted on the migration and invasion in PTC, transwell assay and wound healing assay were conducted in B-CPAP cell line and a knockdown of MET, FN1, MPZL2 and PDLIM4 by siRNAs could significantly reduce the migration and invasion ability in B-CPAP cell line, indicating that an abnormally high expression of those four genes could promote cancer cell metastasis (Figure 7E–H). To estimate the results from the cell functional assay, IHC staining was conducted (Figure 8A), and all six hub genes were significantly increased in the tumor tissues from PTC LNM patients compared to normal samples and the expression of MPZL2 and PDLIM4 were significantly increased in LNM tissues than in non-LNM tissues, which was in accordance with the mRNA expression based on TCGA and GSE 60542 (Figure 8B–G). Based on these findings, we concluded that MET, FN1, MPZL2 and PDLIM4 played important roles in the malignant progression of PTC. Moreover, KEGG enrichment analysis was conducted by GSEA to assess the biological functions of six hub genes from an integral perspective. From the results, we found that all hub genes were involved in two pathways, namely arachidonic acid metabolism and antigen processing and presentation pathway, as well as leishmania infection, long term depression, and a nod like receptor signaling pathway were also annotated for partial hub genes (Figure 9A–F). The detailed information of respective normalized enrichment scores (NES) and the false discovery rate (FDR) p-values are listed in SI Table 3.

Discussion

Lymphatic metastasis was commonly diagnosed in PTC and closely related to a worse prognosis, which might ultimately lead to metastasis.[15] While the 5-year survival time after thyroid cancer surgery was acceptable,[16] there was still a lack of comprehensive research on lymphatic metastasis in PTC. Therefore, new biomarkers for predicting PTC lymph node metastasis should be further identified. In this study, LNM-related genes from key modules were identified via WGCNA analysis on TCGA THCA and GEO datasets. Finally, six gene signatures including COL8A2, CLDN10, MET, FN1, MPZL2 and PDLIM4 were determined using LASSO Cox regression analysis, which were all significantly highly expressed in LNM tumors than in normal and non-LNM samples, and involved in the process of cell migration and invasion verified by cell functional experiments.

The alpha 2 chain of type VIII collagen is encoded by COL8A2, which possesses two non-triple helical domains at the N and C termini and a brief triple-helical domain. It was demonstrated that a knockdown of COL8A2 inhibited the proliferation, migration, and invasion of GBM cells via downregulating the expression of EMT-related proteins in GBM.[17] Claudin-10, encoded as CLDN10, was expressed broadly in the kidney, skin, and salivary glands and served as

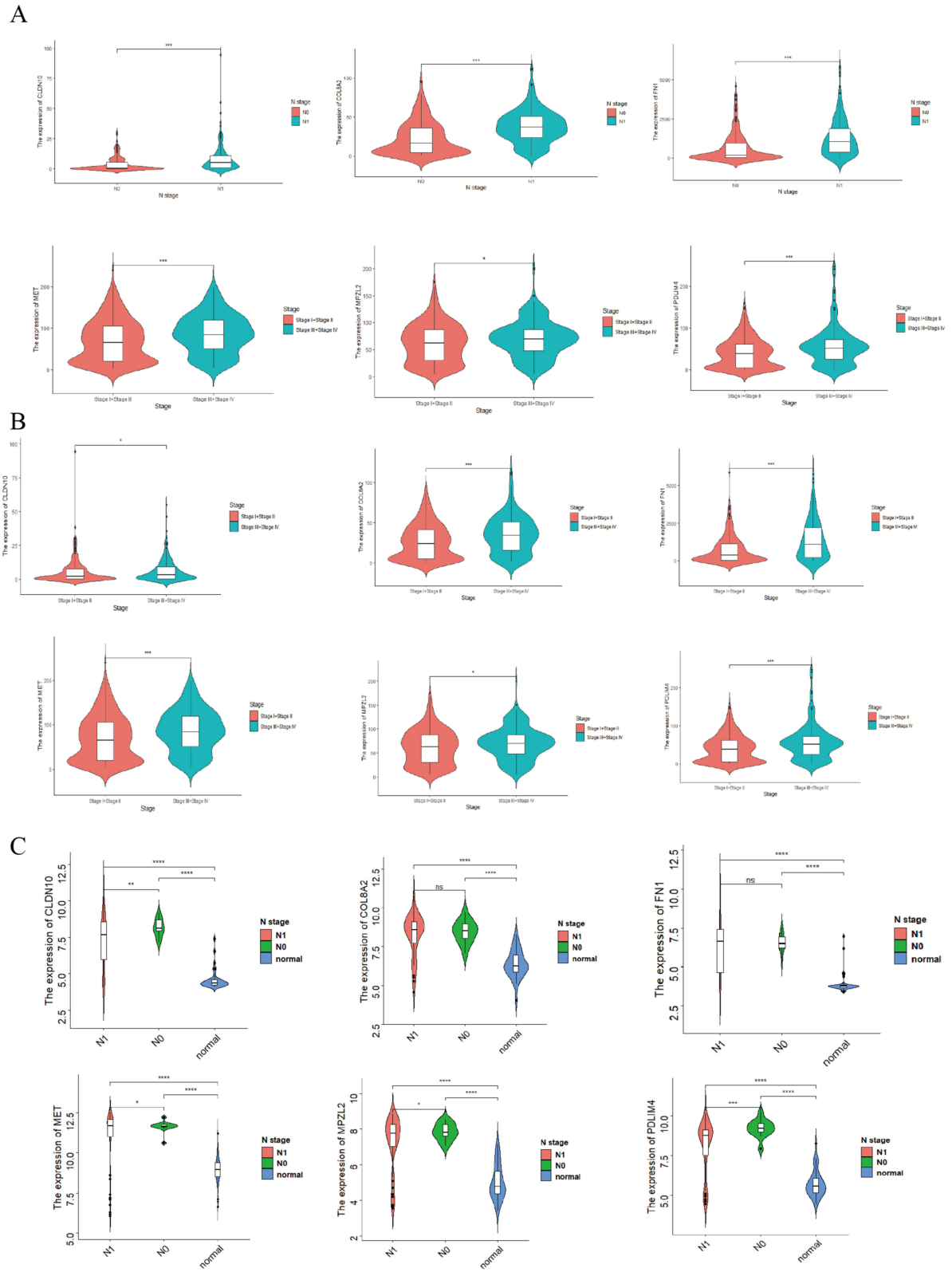

Figure 6 Validation of the expression of hub genes. (**A**) The expression of six hub genes between N0 and N1 patients from TCGA THCA. (**B**) The expression of six hub genes in different stages of patients from TCGA THCA. (**C**) The expression of six hub genes in normal, N0 and N1 tissues of patients from GSE 60542. ns P >0.05, *P < 0.05, **P < 0.01, ***P < 0.001, ****P<0.0001.

https://doi.org/10.2147/IJGM.S502480

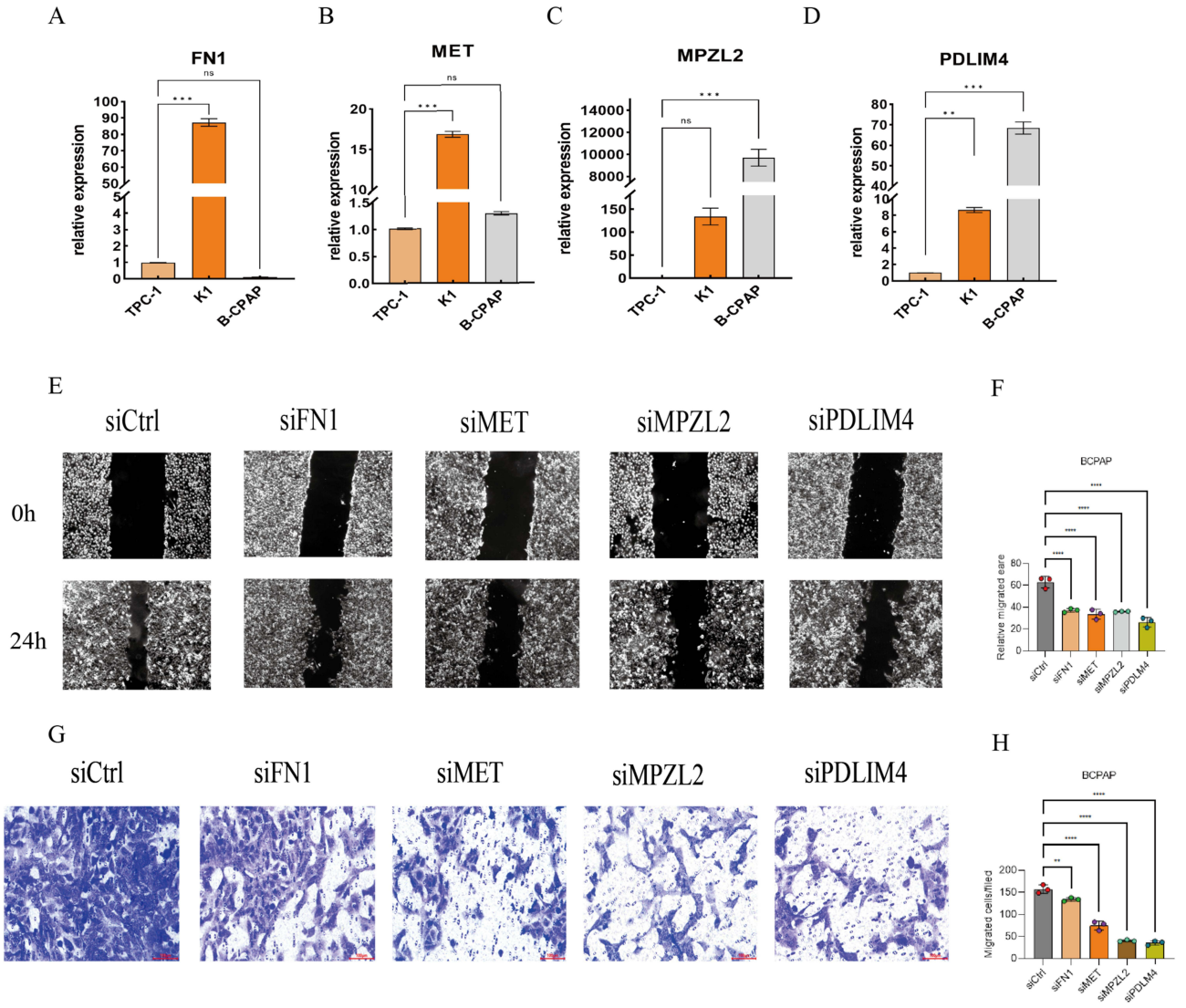

Figure 7 Validation of the functions that hub genes were involved in. (**A–D**) The relative mRNA expression levels of FN1, MET, MPZL2 and PDLIM4 in K1, TPC-1 and B-CPAP cell lines. (**E** and **F**) Results of wound healing experiments transfected with siCtrl and siRNA of FN1, MET, MPZL2 and PDLIM4. (**G** and **H**) Results of transwell assays transfected with siCtrl and siRNA of FN1, MET, MPZL2 and PDLIM4. ns P >0.05, **P < 0.01, ***P < 0.001, ****P<0.0001.

an integral tight junction membrane-spanning protein in two alternatively spliced variant forms.[18] Overexpression of CLDN10 facilitated cell proliferation and cancer metastasis in osteosarcoma and liver cancer.[19,20] In this study, while the mRNA levels of COL8A2 and CLDN10 were comparatively and significantly higher in tumors than normal cells based on TCGA THCA and GEO datasets, we found that the expression of COL8A2 and CLDN10 in PTC cell lines such as K1, TPC-1 and B-CPAP sustained at a low level from the result of qPCR and IHC, which was consistent with the low mRNA expression in samples from TCGA database.

MET, also referred to as HGFR, encoded a transmembrane protein with four immunoglobulin-like domains that was essential for binding hepatocyte growth factor (HGF)[21] Through the HGF-MET interaction, MET regulated a variety of cellular functions, such as proliferation, survival, and motility. This process occurred through dimerization and possibly oligomerization of the MET receptor, which in turn led to trans-phosphorylation and activated multiple downstream pathways.[22–24] MET overexpression activated a carcinogenic pathway including aberrant paracrine or autocrine ligand production and promoted cancer progression such as lung cancer and head and neck cancers.[25,26] Sheng-Yao Cheng et al found MET was overexpressed in PTC and linked to immunoevasive characteristics, distant metastasis, treatment

resistance, and poor clinical outcomes.[27] Fibronectin 1 (FN1) encoded a glycoprotein belonging to the FN family which was expressed in the plasma of several cell types in a soluble dimeric form.[28] Previous studies proved FN1 regulated cell adhesion, proliferation, migration, and differentiation, which contributed to wound healing and embryonic development.-[29] Dysfunction of FN1 was involved in cancer progression and metastasis in melanoma and ovarian cancer as a potential therapeutic target.[30,31] Qi-Shun Geng et al proved overexpression of FN1 was correlated with CD276 and further impacted tumor immune infiltration in THCA.[32] In comparison to non-metastasis groups, IHC results showed that MET and FN1 were both more significantly positive in LNM samples, suggesting that those two genes might play a greater potential function throughout the process of lymph node metastasis in PTC.

PDLIM4, namely PDZ and LIM domain 4, was a PDZ and LIM domain– containing protein which normally maintained cell growth as a tumor suppressor.[33] Mostly, PDLIM4 was repressed via hypermethylation and significantly

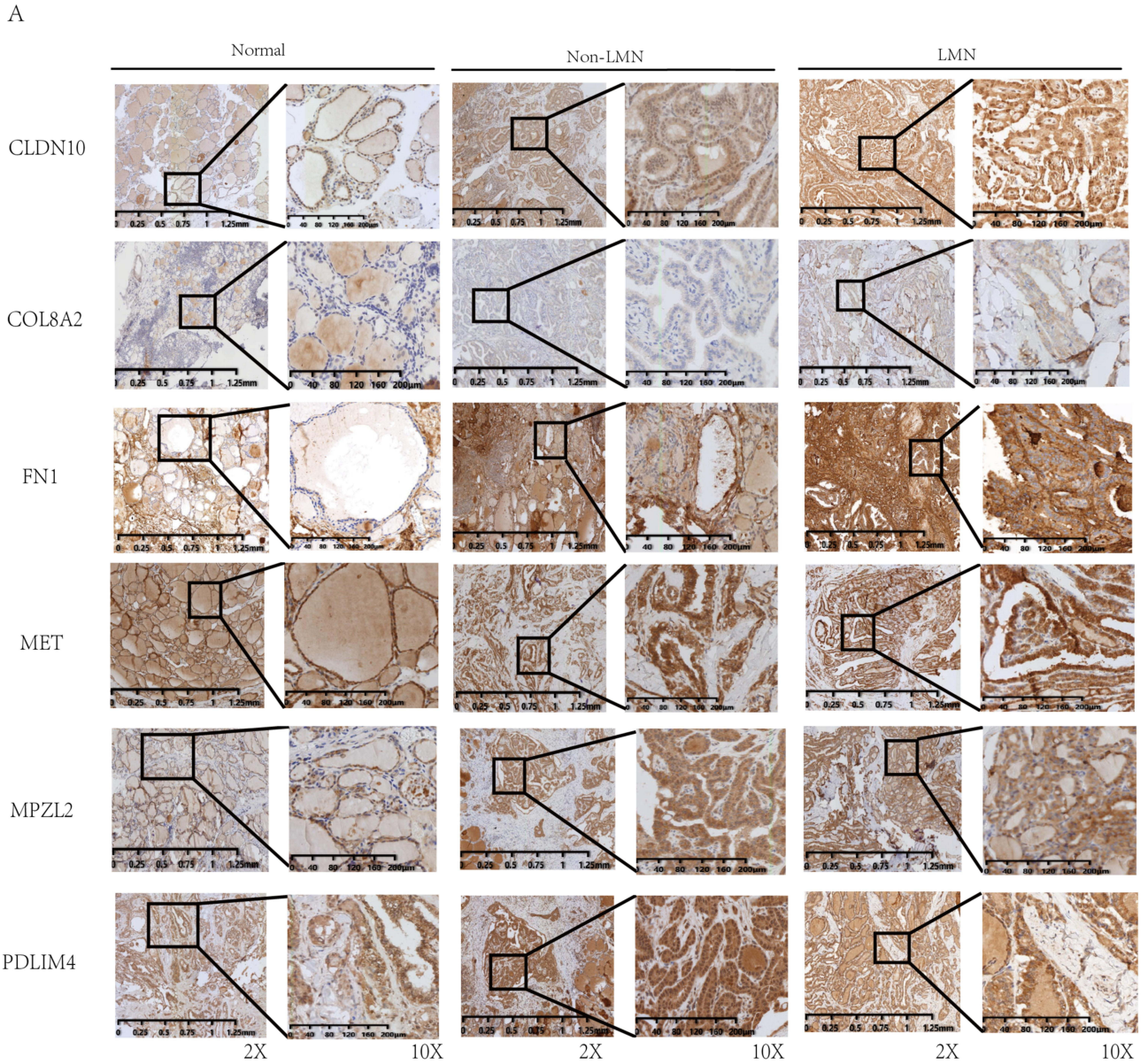

Figure 8 Continued.

https://doi.org/10.2147/IJGM.S502480

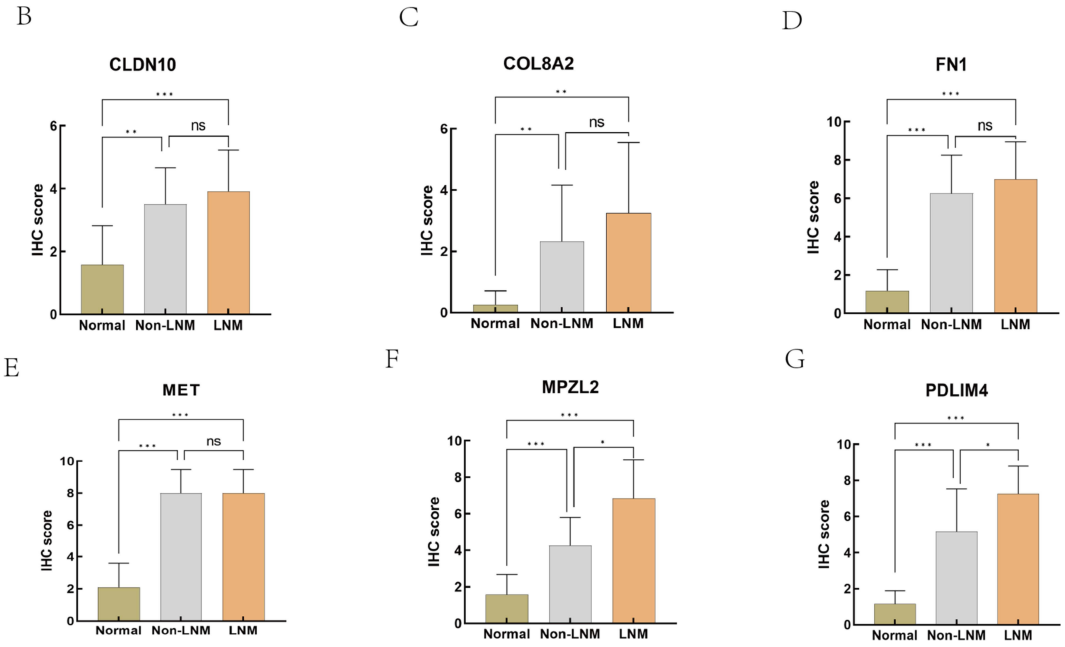

Figure 8 Immunohistochemical analysis of hub genes. (**A**) immunohistochemical staining of six hub genes in normal, non-LNM and LNM tissue samples. (**B-G**) Statistical analysis of IHC results of hub genes. ns P >0.05, *P < 0.05, **P < 0.01, ***P < 0.001.

associated with shorter overall survival in cancer tissues. Yanyan Jia et al demonstrated downregulated expression of PDLIM4 promoted migratory and invasion ovarian cancer cells by activating the STAT3 signal pathway, which contributed to lymph node metastasis in ovarian cancer.[34] However, the high level of PDLIM4 in the specific breast cancer cell lines (MDA-MB-231 and MDA-MB-436) had a higher cell migratory rate via increasing Src kinase activity as an oncogene,[35] demonstrating the controversial and diversity function of PDLIM4. In addition, PDLIM4 was proved to be highly expressed and underwent hypermethylation in THCA,[36] but the biological effect on lymph node metastasis in PTC was still unclear. Myelin protein zero-like 2, which was encoded by MPZL2, was an adhesion molecule that mediated epithelial cell-cell interactions and proved that its overexpression in AML and hepatocellular carcinoma was associated with poor prognosis and recurrence.[37,38] As the homology of myelin protein zero, MPZL2 was first reported to be related to hearing loss functionally via impacting homophilic cell–cell adhesion,[39] but its roles in lymph node metastasis in cancer were still unclear. In this study, we found that PDLIM4 and MPZL2 might link to lymph node metastasis of PTC in in vitro experiments, indicating that those two genes might facilitate the process of lymph node metastasis in PTC. However, the precise function of PDLIM4 and MPZL2 still needs further study.

The functions and mechanisms of hub genes were estimated by KEGG analysis using GSEA algorithm and arachidonic acid metabolism was significantly annotated for all hub genes. As one of the most important metabolic pathways, arachidonic acid metabolism produced many different biologically active metabolites as paracrine factors and signaling messengers which could modulate cardiovascular functions, vascular tone, and inflammatory responses.[40,41] It was proved that inhibiting the synthesis of 20-hydroxy-eicosatetraenoic acid (20-HETE) produced via arachidonic acid metabolism could suppress tumor cell metastasis and invasion.[42] Meanwhile, COX-2 inhibitors, which repressed the COX-2 enzyme activity that was closely related to arachidonic acid metabolism, was proved to inhibit tumor growth by suppressing inflammation and angiogenesis.[43] Therefore, anti-AA pathway therapy might be effective in treating lymph node metastasis of PTC.

Lymph node metastasis in papillary thyroid carcinoma always predicted a further deterioration of disease progression. It was vital to search suitable diagnostic biomarkers for the early prediction of lymph node metastasis in PTC, which would effectively extend patients' lives and improve their quality of life. The six gene signatures, including COL8A2,

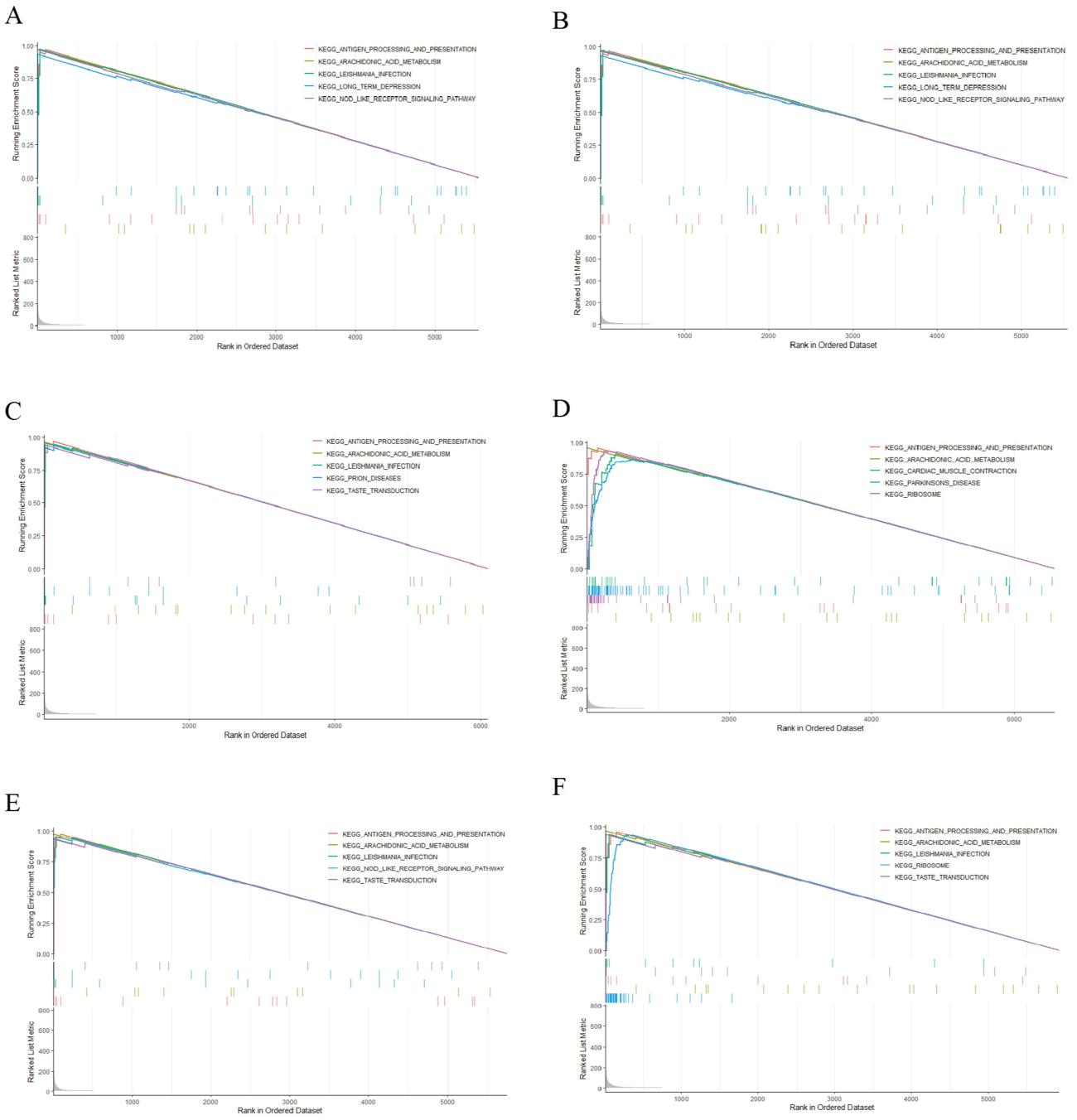

Figure 9 GSEA for CLDN10, COL8A2, FN1, MET, MPZL2 and PDLIM4. (**A** and **B**) KEGG enrichment analysis for CLDN10 and COL8A2. (**C** and **D**) KEGG enrichment analysis for FN1 and MET. (**E** and **F**) KEGG enrichment analysis for MPZL2 and PDLIM4.

CLDN10, MET, FN1, MPZL2 and PDLIM4, could provide references and bases for doctors to determine whether to perform a thyroidectomy in a timely manner by detecting their expression and establishing a predictive model, thereby avoiding misdiagnosis. Meanwhile, the discovery of PDLIM4 and MPZL2, as potential therapeutic targets, provides an important direction for further elucidating the biological mechanisms of lymph node metastasis in PTC. However, this study also had some limitations. Firstly, the three GEO datasets used in this study were relatively outdated and lacked more detailed clinical information, which might introduce bias. Secondly, the number of samples for immunohistochemical analysis was too small considering specific clinical features, and more samples were needed for further investigation

https://doi.org/10.2147/IJGM.S502480

of the potential of predicting lymph node metastasis in PTC, though there were statistical differences from the result of IHC in this study. Thirdly, PDLIM4 and MPZL2 were reported for the first time to be associated with lymph node metastasis in PTC, the specific molecular mechanisms require further research into their inhibiting cancer cell growth by knocking down the expression of PDLIM4 and MPZL2 in vitro in this study.

Conclusions

In summary, six hub genes (COL8A2, MET, FN1, MPZL2, PDLIM4 and CLDN10) were identified as biomarkers by WGCNA and LASSO analysis and the high expressions of hub genes were confirmed to be associated with lymph node metastasis in PTC via cell biology experiments and IHC. Meanwhile, the abnormality of arachidonic acid metabolism might impact the process of lymph node metastasis, which needs further study. Though detailed biological functions remain to be investigated, the finding of the six hub genes offers new insights of diagnosis and treatment on lymph node metastasis in PTC.

Acknowledgments

This paper has been uploaded to ResearchSquare as a preprint: https://www.researchsquare.com/article/rs-4397638/v1.

Disclosure

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Mulita F, Anjum F. *Thyroid Adenoma*. Treasure Island (FL): StatPearls; 2024.
2. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res*. 2014;74(11):2913–2921. doi:10.1158/0008-5472.CAN-14-0155
3. Du L, Zhao Z, Zheng R, et al. Epidemiology of thyroid cancer: incidence and mortality in China, 2015. *Front Oncol*. 2020;10:1702. doi:10.3389/fonc.2020.01702
4. Podnos YD, Smith D, Wagman LD, Ellenhorn JD. The implication of lymph node metastasis on survival in patients with well-differentiated thyroid cancer. *Am Surg*. 2005;71(9):731–734. doi:10.1177/000313480507100907
5. Scheumann GF, Gimm O, Wegener G, Hundeshagen H, Dralle H. Prognostic significance and surgical management of locoregional lymph node metastases in papillary thyroid cancer. *World J Surg*. 1994;18(4):559–67;discussion67–8. doi:10.1007/BF00353765
6. Welch HG, Doherty GM. Saving thyroids - overtreatment of small papillary cancers. *N Engl J Med*. 2018;379(4):310–312. doi:10.1056/NEJMp1804426
7. Wang K, Li H, Zhao J, et al. Potential diagnostic of lymph node metastasis and prognostic values of TM4SFs in papillary thyroid carcinoma patients. *Front Cell Dev Biol*. 2022;10:1001954. doi:10.3389/fcell.2022.1001954
8. Chen W, Li G, Li Z, Zhu J, Wei T, Lei J. Evaluation of plasma exosomal miRNAs as potential diagnostic biomarkers of lymph node metastasis in papillary thyroid carcinoma. *Endocrine*. 2022;75(3):846–855. doi:10.1007/s12020-021-02949-x
9. Yu Y, Guo X, Chai J, et al. Identification of key immune genes related to lymphatic metastasis of papillary thyroid cancer via bioinformatics analysis and experimental validation. *Front Oncol*. 2023;13:1181325. doi:10.3389/fonc.2023.1181325
10. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf*. 2008;9:559. doi:10.1186/1471-2105-9-559
11. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27–30. doi:10.1093/nar/28.1.27
12. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–287. doi:10.1089/omi.2011.0118
13. Szklarczyk D, Gable AL, Nastou KC, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*. 2021;49(D1):D605–D12. doi:10.1093/nar/gkaa1074
14. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–2504. doi:10.1101/gr.1239303
15. Wada N, Duh QY, Sugino K, et al. Lymph node metastasis from 259 papillary thyroid microcarcinomas: frequency, pattern of occurrence and recurrence, and optimal strategy for neck dissection. *Ann Surg*. 2003;237(3):399–407. doi:10.1097/01.SLA.0000055273.58908.19
16. Eichhorn W, Tabler H, Lippold R, Lochmann M, Schreckenberger M, Bartenstein P. Prognostic factors determining long-term survival in well-differentiated thyroid cancer: an analysis of four hundred eighty-four patients undergoing therapy and aftercare at the same institution. *Thyroid*. 2003;13(10):949–958. doi:10.1089/105072503322511355
17. Cheng YX, Xiao L, Yang YL, et al. Collagen type VIII alpha 2 chain (COL8A2), an important component of the basement membrane of the corneal endothelium, facilitates the malignant development of glioblastoma cells via inducing EMT. *J Bioenerg Biomembr*. 2021;53(1):49–59. doi:10.1007/s10863-020-09865-1
18. Hadj-Rabia S, Brideau G, Al-Sarraj Y, et al. Multiplex epithelium dysfunction due to CLDN10 mutation: the HELIX syndrome. *Genet Med*. 2018;20(2):190–201. doi:10.1038/gim.2017.71

19. Zhang X, Wang X, Wang A, Li Q, Zhou M, Li T. CLDN10 promotes a malignant phenotype of osteosarcoma cells via JAK1/Stat1 signaling. *J Cell Commun Signal*. 2019;13(3):395–405. doi:10.1007/s12079-019-00509-7
20. Sun H, Cui C, Xiao F, et al. miR-486 regulates metastasis and chemosensitivity in hepatocellular carcinoma by targeting CLDN10 and CITRON. *Hepatol Res*. 2015;45(13):1312–1322. doi:10.1111/hepr.12500
21. Cai K, Zhang X, Bai XC. Cryo-electron microscopic analysis of single-pass transmembrane receptors. *Chem Rev*. 2022;122(17):13952–13988. doi:10.1021/acs.chemrev.1c01035
22. Mood K, Saucier C, Bong YS, Lee HS, Park M, Daar IO. Gab1 is required for cell cycle transition, cell proliferation, and transformation induced by an oncogenic met receptor. *mol Biol Cell*. 2006;17(9):3717–3728. doi:10.1091/mbc.e06-03-0244
23. Orian-Rousseau V, Morrison H, Matzke A, et al. Hepatocyte growth factor-induced Ras activation requires ERM proteins linked to both CD44v6 and F-actin. *mol Biol Cell*. 2007;18(1):76–83. doi:10.1091/mbc.e06-08-0674
24. Rajadurai CV, Havrylov S, Zaoui K, et al. Met receptor tyrosine kinase signals through a cortactin-Gab1 scaffold complex, to mediate invadopodia. *J Cell Sci*. 2012;125(Pt 12):2940–2953. doi:10.1242/jcs.100834
25. Lengyel E, Sawada K, Salgia R. Tyrosine kinase mutations in human cancer. *Curr Mol Med*. 2007;7(1):77–84. doi:10.2174/156652407779940486
26. Birchmeier C, Birchmeier W, Gherardi E, Vande Woude GF. Met, metastasis, motility and more. *Nat Rev mol Cell Biol*. 2003;4(12):915–925. doi:10.1038/nrm1261
27. Cheng SY, Wu ATH, Batiha GE, et al. Identification of DPP4/CTNNB1/MET as a theranostic signature of thyroid cancer and evaluation of the therapeutic potential of Sitagliptin. *Biology*. 2022;11(2):324. doi:10.3390/biology11020324
28. Gao W, Liu Y, Qin R, Liu D, Feng Q. Silence of fibronectin 1 increases cisplatin sensitivity of non-small cell lung cancer cell line. *Biochem Biophys Res Commun*. 2016;476(1):35–41. doi:10.1016/j.bbrc.2016.05.081
29. Pankov R, Yamada KM. Fibronectin at a glance. *J Cell Sci*. 2002;115(Pt 20):3861–3863. doi:10.1242/jcs.00059
30. Li B, Shen W, Peng H, et al. Fibronectin 1 promotes melanoma proliferation and metastasis by inhibiting apoptosis and regulating EMT. *Onco Targets Ther*. 2019;12:3207–3221. doi:10.2147/OTT.S195703
31. Lou X, Han X, Jin C, et al. SOX2 targets fibronectin 1 to promote cell migration and invasion in ovarian cancer: new molecular leads for therapeutic intervention. *OMICS*. 2013;17(10):510–518. doi:10.1089/omi.2013.0058
32. Geng QS, Huang T, Li LF, Shen ZB, Xue WH, Zhao J. Over-expression and prognostic significance of FN1, correlating with immune infiltrates in thyroid cancer. *Front Med*. 2021;8:812278. doi:10.3389/fmed.2021.812278
33. Vallenius T, Scharm B, Vesikansa A, Luukko K, Schafer R, Makela TP. The PDZ-LIM protein RIL modulates actin stress fiber turnover and enhances the association of alpha-actinin with F-actin. *Exp Cell Res*. 2004;293(1):117–128. doi:10.1016/j.yexcr.2003.09.004
34. Jia Y, Shi H, Cao Y, Feng W, Li M, Li X. PDZ and LIM domain protein 4 suppresses the growth and invasion of ovarian cancer cells via inactivation of STAT3 signaling. *Life Sci*. 2019;233:116715. doi:10.1016/j.lfs.2019.116715
35. Kravchenko DS, Ivanova AE, Podshivalova ES, Chumakov SP. PDLIM4/RIL-mediated regulation of Src and malignant properties of breast cancer cells. *Oncotarget*. 2020;11(1):22–30. doi:10.18632/oncotarget.27410
36. Jiang X, Xu Z, Jiang S, et al. PDZ and LIM domain-encoding genes: their role in cancer development. *Cancers*. 2023;15(20):5042.
37. Yu P, Lan H, Song X, Pan Z. High expression of the SH3TC2-DT/SH3TC2 gene pair associated with FLT3 mutation and poor survival in acute myeloid leukemia: an integrated TCGA analysis. *Front Oncol*. 2020;10:829. doi:10.3389/fonc.2020.00829
38. Ni Q, Chen Z, Zheng Q, et al. Epithelial V-like antigen 1 promotes hepatocellular carcinoma growth and metastasis via the ERBB-PI3K-AKT pathway. *Cancer Sci*. 2020;111(5):1500–1513. doi:10.1111/cas.14331
39. Bademci G, Abad C, Incesulu A, et al. MPZL2 is a novel gene associated with autosomal recessive nonsyndromic moderate hearing loss. *Hum Genet*. 2018;137(6–7):479–486. doi:10.1007/s00439-018-1901-4
40. Roman RJ. P-450 metabolites of arachidonic acid in the control of cardiovascular function. *Physiol Rev*. 2002;82(1):131–185. doi:10.1152/physrev.00021.2001
41. Zhao H, Qi G, Han Y, et al. 20-hydroxyeicosatetraenoic acid is a key mediator of angiotensin II-induced apoptosis in cardiac myocytes. *J Cardiovasc Pharmacol*. 2015;66(1):86–95. doi:10.1097/FJC.0000000000000248
42. Borin TF, Shankar A, Angara K, et al. HET0016 decreases lung metastasis from breast cancer in immune-competent mouse model. *PLoS One*. 2017;12(6):e0178830. doi:10.1371/journal.pone.0178830
43. Yang H, Rothenberger E, Zhao T, et al. Regulation of inflammation in cancer by dietary eicosanoids. *Pharmacol Ther*. 2023;248:108455. doi:10.1016/j.pharmthera.2023.108455

International Journal of General Medicine

Dovepress
Taylor & Francis Group

Publish your work in this journal

The International Journal of General Medicine is an international, peer-reviewed open-access journal that focuses on general and internal medicine, pathogenesis, epidemiology, diagnosis, monitoring and treatment protocols. The journal is characterized by the rapid reporting of reviews, original research and clinical studies across all disease areas. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit http://www.dovepress.com/testimonials.php to read real quotes from published authors.

Submit your manuscript here: https://www.dovepress.com/international-journal-of-general-medicine-journal