Endocrine (2019) 66:573–584
https://doi.org/10.1007/s12020-019-02021-9

ORIGINAL ARTICLE

Check for updates

Identification of gene co-expression modules and hub genes associated with lymph node metastasis of papillary thyroid cancer

Tianyu Zhai[1] · Dilidaer Muhanhali[1] · Xi Jia[2] · Zhiyong Wu[3] · Zhenqin Cai[1] · Yan Ling[1]

Received: 2 May 2019 / Accepted: 12 July 2019 / Published online: 22 July 2019

Abstract

Papillary thyroid cancer (PTC) is the most prevalent histological type among thyroid cancers, and some patients are at a high risk for recurrent disease or even death. Identification for the potential biomarkers of PTC may contribute to early discovery of recurrence and treatment. In The Cancer Genome Atlas (TCGA) database, we obtained the information of RNA sequence data and clinical characteristics of PTC. Weighted gene co-expression network analysis (WGCNA) was performed to construct gene co-expression networks and investigate the relationship between modules and clinical traits. Finally, we constructed 16 co-expression modules in 10,428 genes, and three key modules (darkturquoise, lightyellow, and red) associated with tumor N grade were identified. The results of functional annotation indicated that the darkturquoise module was primarily enriched in the regulation of the extracellular matrix (ECM), collagen metabolism, and cell adhesion, the lightyellow module was primarily enriched in the mitochondrial function regulation and energy synthesis, and the red module was primarily enriched in the process of cell junction, apoptosis, and inflammatory response, suggesting their significant role in the progression of PTC. In addition, the hub genes in the three modules were identified and screened for differentially expressed genes (DEGs). Relapse-free survival analyses found that 11 genes (*KCNQ3*, *MET*, *FN1*, *ITGA3*, *RUNX1*, *ITGA2*, *PERP*, *GCSH*, *FAAH*, *NGFRAP1,* and *HSPA5*) may play a pivotal role in PTC relapse. In general, our research revealed the key co-expression modules and identified several prognostic biomarkers, which provides some new insights into the lymph node metastasis of PTC.

Keywords Papillary thyroid cancer · Weighted gene co-expression network analysis · Hub genes · Lymph node metastasis

Introduction

Thyroid cancer is a frequent malignant tumor of the endocrine system and its incidence has been increasing globally in recent decades [1–3]. The U.S. National Cancer Institute estimated that the number of new thyroid cancers and deaths is 53,990 and 2060 in 2018 (http://seer.cancer.gov/). Papillary thyroid cancer (PTC) is the most prevalent histological type and accounts for nearly 90% in thyroid carcinoma. Although PTC grows slowly with high 10-year survival rates, it commonly metastasizes to the cervical lymph nodes initially [4]. Thus, some patients are at high risk for recurrent disease or even death [5]. Currently, the explicit molecular mechanism of PTC has not been illuminated, and many published studies have focused on the role of solitary genes (such as *SIRT7*, *IGF-1R*, *ZNF677*, and *TBX3*) in the carcinogenesis of PTC [6–9]. But carcinogenesis is a complex mechanism, including high interactions between genes and similar gene expression patterns [10]. Therefore, it is urgent to explore the signature of the PTC genetic network, which can predict tumor stage or assess malignant potential, and provide personalized therapy for PTC patients.

Gene networks have been frequently used to analyze the origination and development of various cancers. Weighted

Supplementary information The online version of this article (https://doi.org/10.1007/s12020-019-02021-9) contains supplementary material, which is available to authorized users.

✉ Yan Ling
doctorlingyan@163.com

[1] Department of Endocrinology and Metabolism, Zhongshan Hospital, Fudan University, No.180 Fenglin Road, 200032 Shanghai, China

[2] Department of Endocrinology, Jinshan Hospital, Fudan University, No.1508 Longhang Road, 201500 Shanghai, China

[3] The Graduate School of Fujian Medical University, 350108 FuZhou, China

gene co-expression network analysis (WGCNA) is deemed as an efficient network-based approach, which can investigate the signature of gene networks in the pathogenesis of complicated diseases at system level [11]. WGCNA is an algorithm that constructs scale-free gene co-expression networks, which can not only classify different gene modules, but also figure out the relationships between clinical features and gene modules [12]. Some studies have successfully used WGCNA to explore the functional co-expression modules and the hub genes of different diseases, such as chronic kidney disease, pancreatic ductal adenocarcinoma, and Graves' disease [13–15].

In this study, we used WGCNA and other methods to explore public transcriptome data and clinical information of PTC patients. A few key gene modules associated with lymph node metastasis were identified, and several hub genes were found out. These genes may play a momentous role in pathogenesis or serve as clinical biomarkers of progression and recurrence of PTC.

Materials and methods

Data acquisition

We downloaded data information of PTC patients from The Cancer Genome Atlas (TCGA) database (https://cancergenome.nih.gov/), which is the largest publicly available dataset about cancer now. RNA-sequencing and relative clinical information of 502 patients were obtained. Among these 502 patients, relapse-related information of 498 patients is available. The official gene symbol of mRNA was obtained according to the Genome research project of ENCyclopedia of DNA Elements (GENCODE) (GRCh38) (v22) catalog (http://www.gencodegenes.org/). Median value was identified as the expression value of gene with multiple probes, and probes with multiple genes were removed. Genes with Fragments Per Kilobase Million (FPKM) <2 were filtered out, and the remaining 10,428 genes were included eventually.

Construction of co-expression module networks

First, all genes and samples in the expression profile passed the good genes or good samples test. Then, the "WGCNA" package of R software was run to construct scale-free network of all genes [12]. When the soft-thresholding power was equal to 6, correlation coefficient reached 0.8 and scale-free co-expression network was achieved. Then, cluster analysis was performed to divide genes with similar patterns into gene modules and gene module size was no <30. To further explore the gene modules, we defined the cut height threshold as 0.35 and merged a few modules.

Construction of module–trait relationships

The correlation between the phenotype and module eigengenes (MEs) was considered as the module–trait associations. Therefore, we calculated the correlation and identified the relevant modules, which are closely related to the clinical traits. For each gene modules, gene significance (GS) represents correlation level between expression pattern and phenotype, and module membership (MM) represents correlation level between expression pattern and MEs.

Enrichment analysis of key co-expression modules

Gene ontology (GO) enrichment analysis was performed by Database for Annotation, Visualization and Integrated Discovery (DAVID, http://david.abcc.ncifcrf.gov/), which is regarded as the most common functional annotation tool [16–18]. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways was used as reference to perform the pathway enrichment analysis through DAVID. Genes of the key co-expression modules were uploaded to investigate the potential functions, and $P < 0.001$ was regarded as the cutoff criteria.

Protein–protein interaction (PPI) analysis and identification of hub genes

We uploaded genes of the key co-expression modules to the STRING database [19] (http://www.string-db.org/) and HIPPIE database [20] (http://cbdm.uni-mainz.de/hippie/), then chose confidence >0.4 to perform the PPI network analysis and used Cytoscape software to visualize the results of STRING database [21].

Hub genes are usually deemed to be functionally critical and immensely interconnected with other genes in module, so we screened out hub genes through the three manners: (1) In the PPI analysis of STRING database, hub genes were defined as the 30 genes with highest connectivity degree. (2) In the PPI analysis of HIPPIE database, due to the lower connectivity degree of the genes in HIPPIE database, hub genes were defined as the ten genes with the highest connectivity degree. (3) In the module–trait analysis, genes with GS score >0.8 and MM score >0.2 or 0.15 were defined as hub genes that were notably correlated with the definite clinical trait. Hub genes derived from three manners [PPI network (STRING and HIPPIE database) and module–trait analysis] were all regarded as "final" hub genes in particular co-expression modules.

Screening differentially expressed genes (DEGs) in hub genes

We screened the DEGs in "final" hub genes between 58 PTC samples and its paired adjacent normal samples using

the "limma" (linear models for microarray data) R package [22]. With TPM value, cutoff criteria was chosen as |log2 fold change (FC)| >1 and FDR (false discovery rate) <0.05 to select DEGs for the further analysis.

Validation for DEGs

We obtained GSE60542 dataset from Gene Expression Omnibus database (https://www.ncbi.nlm.nih.gov/geo). GSE60542 dataset contains 27 normal thyroid samples and 28 PTC samples, which was used as the test set to validate the expression levels of DEGs. Mann–Whitney test was utilized to compare genes expression levels between groups.

Relapse-free survival analyses

There are 498 PTC patients in TCGA with available relapsed-related information. These patients were divided into two groups (High or Low) through the median expression value of each gene. Kaplan–Meier plots and log-rank test were applied to illustrate the relationship between groups and relapse-free survival.

Cell culture and RNA isolation

Two human PTC-derived cell lines [B-CPAP (harboring $BRAF^{V600E}$ mutation) and TPC-1 (harboring the *RET*/PTC1 rearrangement)] and a normal thyroid epithelial cell line (Nthy-ori 3-1) were purchased from FuHeng Biological Technology Co. Ltd (Shanghai, China). All cell lines were cultured in DMEM (HyClone, Logan, UT, USA) media with 10% fetal bovine serum at 37 °C in a humidified atmosphere containing 5% CO_2. The TRIzol Reagent (Takara, Kusatsu, Japan) and the Reverse Transcription Kit (Takara, Kusatsu, Japan) were used to extract total RNA and synthesize cDNA following the instructions of manufacturers.

Quantitative real-time polymerase chain reaction (qRT-PCR)

Eleven genes (*KCNQ3*, *MET*, *FN1*, *ITGA3*, *RUNX1*, *ITGA2*, *PERP*, *GCSH*, *FAAH*, *NGFRAP1*, and *HSPA5*) of interest were further examined in B-CPAP, TPC-1, and Nthy-ori 3-1 cell line by qRT-PCR, which was performed through the SYBR Green PCR Kit (Takara, Kusatsu, Japan). β-actin was used as an internal control and Nthy-ori 3-1 cell line was used as a calibrator sample for PTC-derived cell line, relative expressions were calculated using the $2^{-\Delta\Delta Ct}$ method. The primer sequences of this study were shown in Supplementary Table 1. Student's *t* test was utilized to compare genes expression levels PTC-derived cell lines and normal thyroid epithelial cell line. $P < 0.05$ was regarded as statistically significant.

Results

Clinical features of 502 PTC patients

The clinical characteristics of 502 PTC patients were provided by TCGA. The particular information was shown in Supplementary Table 2, including age, gender, TNM grade, stage, recurrence, status, histological type, mutation, etc.

Construction of co-expression module networks of PTC

We included 10,428 genes in 502 PTC patients, using the FPKM value to construct co-expression network through "WGCNA" package of R software. When the soft-thresholding power value was chosen as 6, the correlation coefficient threshold reached 0.8 and the topology network was closer to scale-free (Fig. 1a). As a result, we used 6 as the power value and successfully constructed 16 merged co-expression modules through the WGCNA analysis. These modules were shown in distinct colors (Fig. 1b) and contained varying numbers of genes (Table 1). Interactions analysis of each co-expression module shows that these main modules were separated from each other in network (Fig. 1c).

Module–trait analysis of PTC

We used clinical traits dataset of 502 PTC patients in the TCGA database and removed some worthless information in this study. Then, the correlations between MEs and clinical traits were determined (Fig. 2a). The result of analyses indicated that co-expression modules were correlated with some particular phenotypes. For example, the darkturquoise module was associated with disease stage, T grade, N grade, and status. Figure 2b displays the significance level of 16 co-expression modules associated with N grade. It is clear that three key modules (darkturquoise, lightyellow, and red) intensively correlated to N grade. Then, scatter diagrams of GS for N grade vs. MM in the three modules were plotted, respectively (Fig. 2c), which show significance of the genes in darkturquoise, lightyellow, and red module related to N grade. The *P*-value of 4.3×10^{-14} (darkturquoise), 3.9×10^{-173} (lightyellow), and 1.5×10^{-37} (red) indicated high correlations between the three modules and N grade.

Enrichment analysis of key co-expression modules

We uploaded genes of the key co-expression modules to the DAVID database for GO and KEGG analysis. In the darkturquoise module, the top five enriched biological processes in GO analysis are extracellular matrix

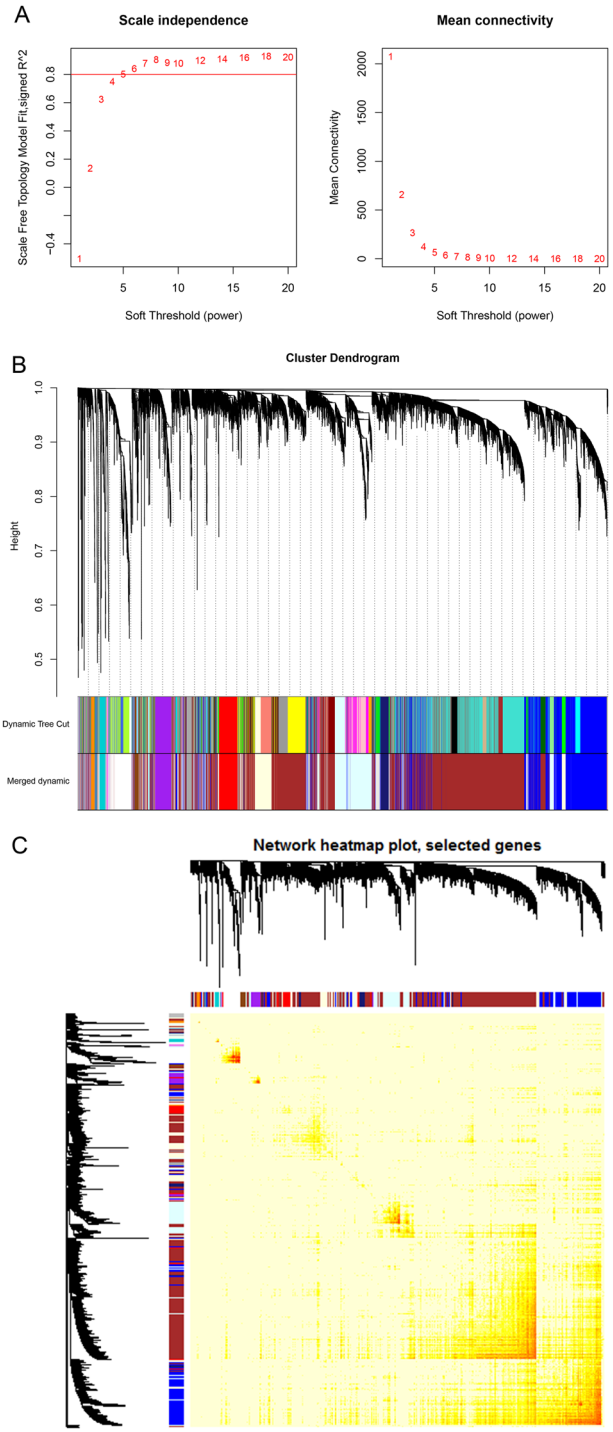


Fig. 1 WGCNA constructed gene co-expression networks and obtained 16 co-expression modules. **a** Calculation and selection of optimal soft-thresholding power. Influence of different powers on the scale independence (left) and influence of different powers on the mean connectivity (right). **b** Clustering dendrogram of 10,428 genes and 16 merged modules in 502 PTC patients of the TCGA database. **c** The heatmap plot of the interactions among co-expression modules. The darker yellow and red colors represent higher interconnectedness and lighter color represents lower interconnectedness

Table 1 The number of genes in 16 modules

Module	Gene number
Blue	1852
Brown	4504
Darkgray	103
Darkorange	75
Darkturquoise	125
Gray	85
Lightcyan	1144
Lightyellow	878
Midnightblue	216
Paleturquoise	38
Purple	324
Red	370
Saddlebrown	104
Steelblue	47
Violet	34
White	529

organization, collagen catabolic process, collagen fibril organization, cell adhesion, and skeletal system development, and the top five enriched pathways in KEGG analysis are ECM–receptor interaction, protein digestion and absorption, focal adhesion, *PI3K-Akt* signaling pathway, and amebiasis (Fig. 3a). In the lightyellow module, the top five enriched biological processes in GO analysis are mitochondrial respiratory chain complex I assembly, oxidation–reduction process, mitochondrial electron transport, ATP biosynthetic process, and tricarboxylic acid cycle, and the top five enriched pathways in KEGG analysis are oxidative phosphorylation, metabolic pathways, Parkinson's disease, carbon metabolism, and Huntington's disease (Fig. 3b). In the red module, the top five enriched biological processes in GO analysis are cell–cell adhesion, signal transduction, viral entry into the host cell, response to drug, and extracellular matrix organization, and the five enriched pathways in KEGG analysis are focal adhesion, cytokine–cytokine receptor interaction, proteoglycans in cancer, endocytosis, and apoptosis (Fig. 3c).

PPI analysis and hub genes

In the PPI analysis of STRING database, hub genes were defined as the top 30 genes of the connectivity degree. Figure 4a–c shows either the PPI network of genes or the intensive correlations among hub genes in the three co-expression modules. In the PPI analysis of the HIPPIE database, hub genes were defined as the top ten genes of connectivity degree. In addition, 20 and 17 genes in the module–trait analysis with GS score >0.8 and MM score >0.2 were also defined as hub genes in lightyellow and red modules, 23

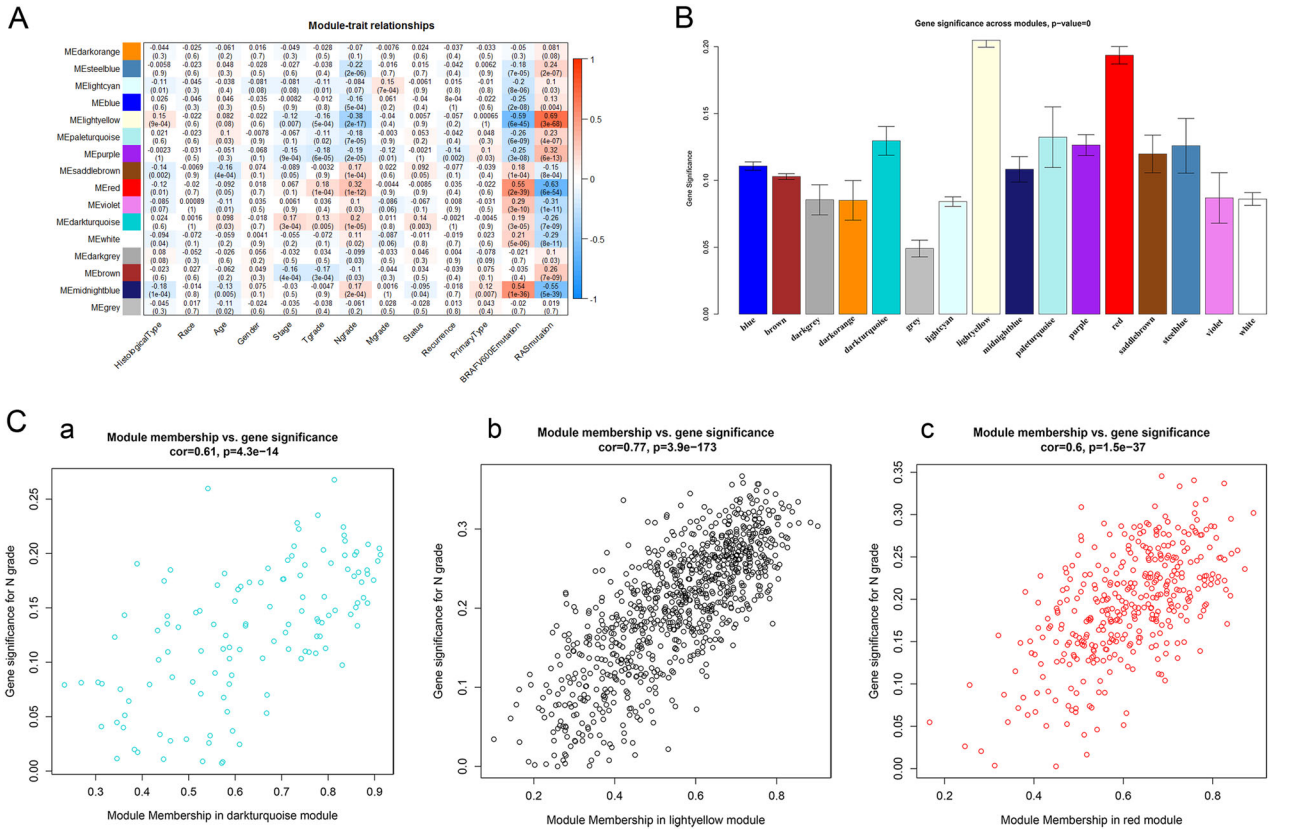


Fig. 2 Module–trait correlations analyses. **a** Relationships between MEs and clinical traits. Each cell contains the corresponding correlation coefficient and *P*-value. **b** The significance level of 16 co-expression modules associated with N grade. **c** Scatter diagrams of gene significance N grade vs. module membership in the three key modules ((a) darkturquoise, (b) lightyellow, and (c) red)

genes with GS score >0.8 and MM score >0.15 were also defined as hub genes in the darkturquoise module. Figure 4d indicates that these genes are also extensively associated with each other. Hub genes in each module derived from three manners were merged and regarded as the "final" hub genes for the further analysis (Supplementary Table 3).

DEGs between PTC samples and adjacent normal thyroid samples

Fifty eight PTC samples and its adjacent normal thyroid samples were used to identify DEGs. A total of 152 final hub genes were analyzed by "limma" R package and 25 DEGs (16 upregulated and 9 downregulated) were identified finally. Figure 5 shows the heatmap of these 25 DEGs.

Validation for DEGs

GSE60542 was used as the test gene set to perform the DEGs expression validation. Supplementary Fig 1 shows the validation results, and it is notable that the majority of 25 DEGs were consistent with previous results, except for *PTPRE*, *MTHFD1L,* and *LUM*. Detailed DEGs in the three key co-expression modules are shown in Table 2.

Relapse-free survival analyses

In the TCGA, tumor recurrence occurred in 46 of 498 PTC patients during the follow-up period. The high and low groups were determined based on the median expression value of each DEG. Supplementary Fig. 2 shows that several genes may play a significant role in the progression and recurrence of PTC patients, including *KCNQ3*, *MET*, *FN1*, *ITGA3*, *RUNX1*, *ITGA2*, *PERP*, *GCSH*, *FAAH*, *NGFRAP1*, and *HSPA5*.

Validation by qRT-PCR

We performed qRT-PCR to examine the relative mRNA expressions of those 11 genes that were significant in relapse-free survival analyses in PTC cell lines. Figure 6 shows that *KCNQ3*, *MET*, *FN1*, and *ITGA3* were upregulated, and *GCSH* was downregulated in B-CPAP and TPC-1 cells compared with Nthy-ori 3-1 cell, which was completely in accordance with the result of DEG identification. *RUNX1* and *PERP* were upregulated in the B-CPAP cell, and *ITGA2* was upregulated in the TPC-1 cell compared with the Nthy-ori 3-1 cell, while *FAAH, NGFRAP1*, and *HSPA5* were downregulated in the TPC-1 cell compared with the Nthy-ori 3-1 cell.

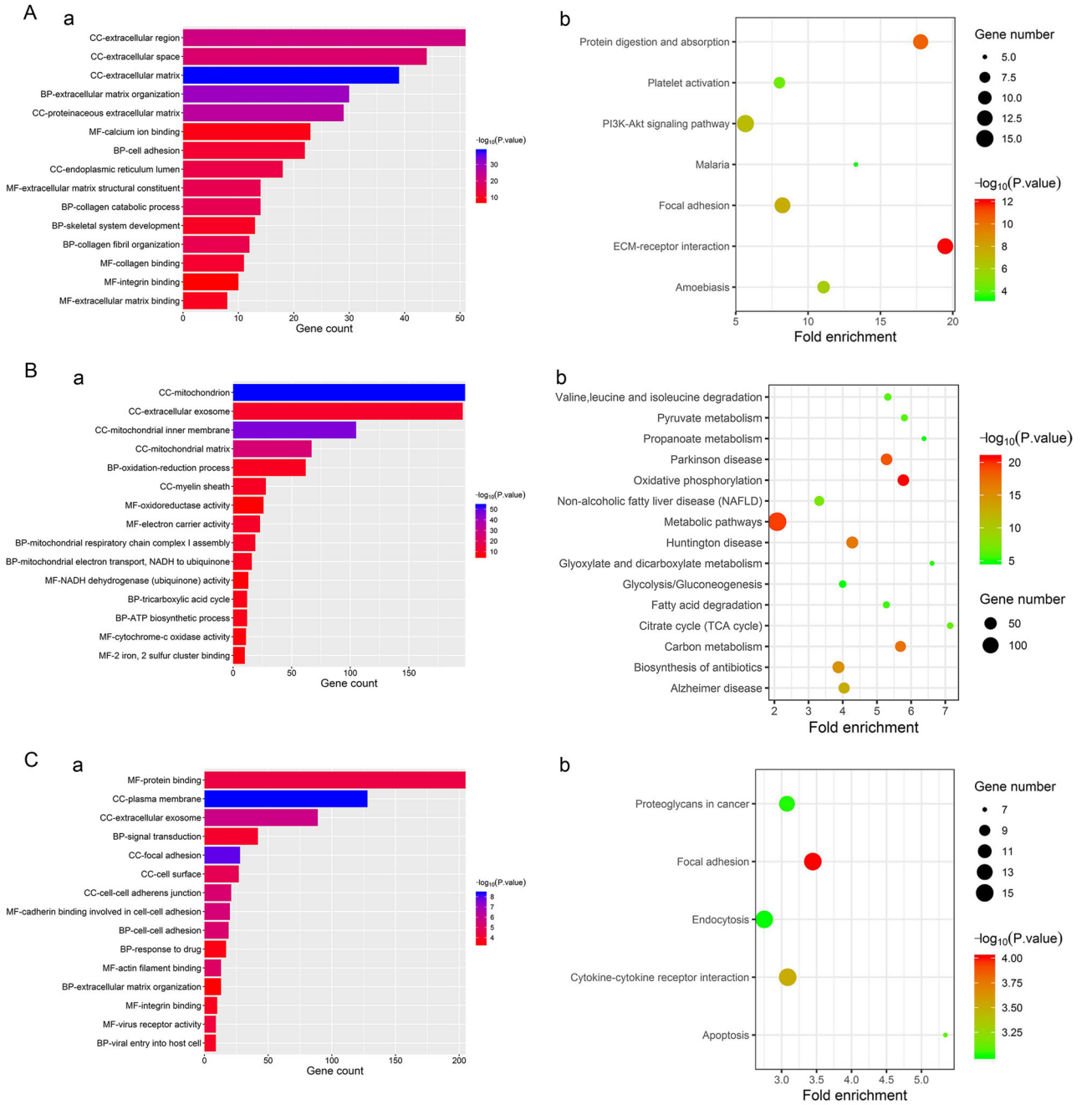
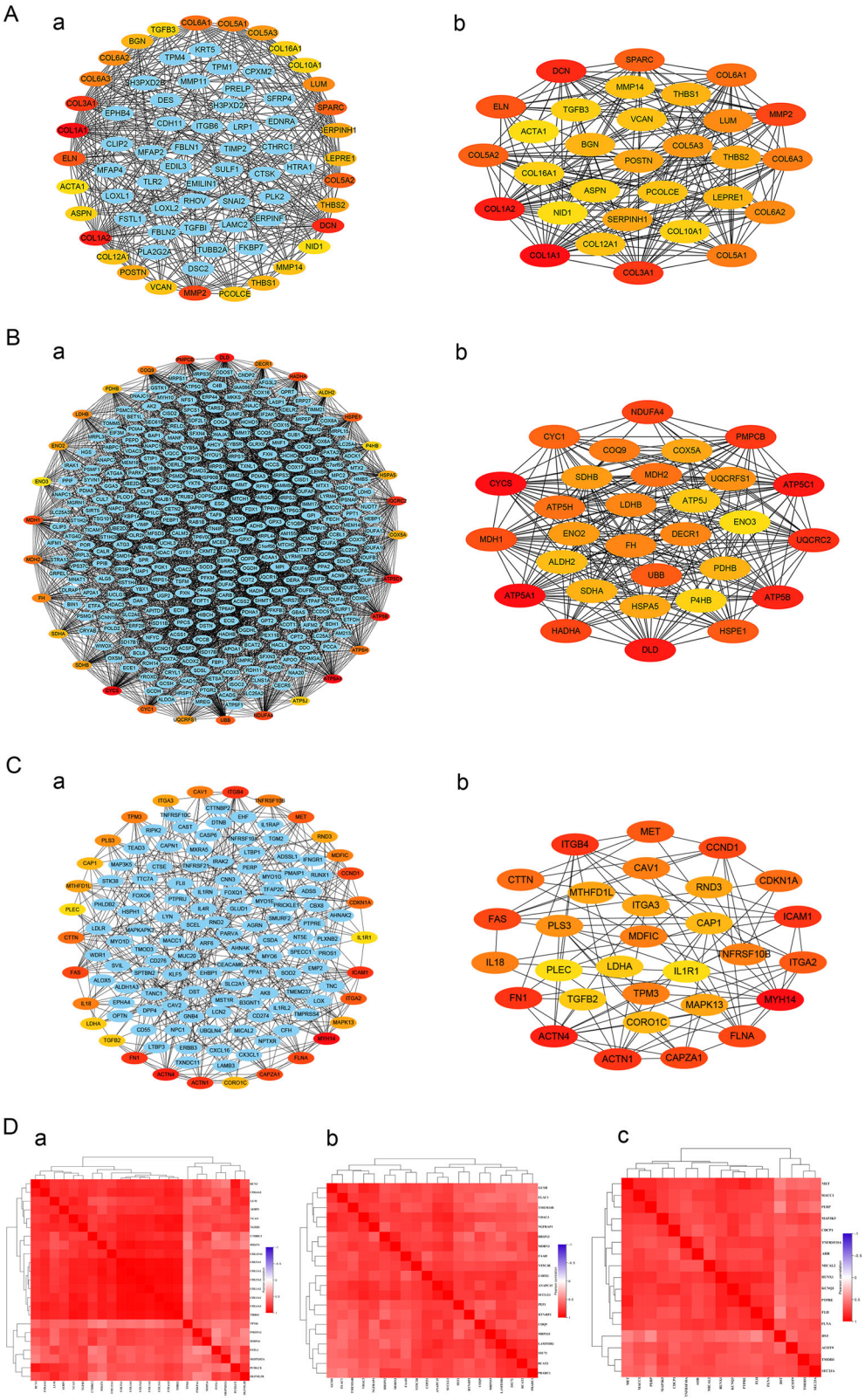


Fig. 3 Plot of the enriched GO and KEGG terms in key co-expression modules. **a** (a) The plot of the top five enriched BP (biological process), MF (molecular function), and CC (cellular component) terms in the darkturquoise module. (b) The plot of the enriched KEGG pathway in the darkturquoise module. **b** (a) The plot of the top five enriched BP (biological process), MF (molecular function), and CC (cellular component) terms in the lightyellow module. (b) The plot of the top 15 enriched KEGG pathways in the lightyellow module. **c** (a) The plot of the top five enriched BP (biological process), MF (molecular function), and CC (cellular component) terms in the red module. (b) The plot of the enriched KEGG pathway in the red module

Discussion

PTC is a common and complex cancer with great heterogeneity in terms of morphological features and prognosis [23]. Most papillary cancers tend to be biologically inactive, but there are a small percentage of patients who show the higher invasiveness and aggressive clinical features [24]. To better elucidate the molecular pathogenesis of PTC progression and recurrence, we used the bioinformatics approach—WGCNA to investigate the signature of the PTC genetic network.

In the present study, gene co-expression networks were constructed and 16 co-expression modules in total were obtained from 10,428 genes. Among them, we identified three key co-expression modules (darkturquoise,



lightyellow, and red) related to clinical trait (N grade). In addition to the three modules, a few other modules also function in PTC, such as the lightcyan, purple, and midnightblue module. Therefore, our study suggested that sophisticated gene networks function jointly in the origination and development of PTC.

◀ **Fig. 4** Visualization of the key modules and hub genes. **a** PPI analysis visualization of the darkturquoise module. (a) All genes in the darkturquoise module. Red and yellow nodes represent the hub genes. (b) Thirty hub genes in the darkturquoise module. **b** PPI analysis visualization of the lightyellow module. (a) All genes in the lightyellow module. Red and yellow nodes represent the hub genes. (b) Thirty hub genes in the lightyellow module. **c** PPI analysis visualization of the red module. (a) All genes in the red module. Red and yellow nodes represent the hub genes. (b) Thirty hub genes in the red module. **d** Genes with high gene significance for N grade were intensively correlated to each other. The darker color represents the higher correlation between genes. (a) Genes in the darkturquoise module with GS score >0.8 and MM score >0.15. (b) Genes in the lightyellow module with GS score >0.8 and MM score >0.2. (c) Genes in the red module with GS score >0.8 and MM score >0.2

GO analysis is a compelling analytical approach, and is generally used to divide biological terms into function-related groups [25]. Therefore, genes in the three modules were functionally annotated by performing GO analysis. Annotation results show that the darkturquoise module was primarily enriched in the adjustment of the extracellular matrix, collagen metabolism, and cell adhesion. The lightyellow module was primarily enriched in the mitochondrial function regulation and energy synthesis. The red module was primarily enriched in the process of cell junction, apoptosis, and inflammatory response. These results demonstrated that the three modules may lead to the invasiveness and metastasis of tumor cells, and multiple biological processes are involved in the PTC progression at the molecular and cellular level.

The final hub genes were obtained through GS vs. MM and PPI network. Twenty two DEGs were screened out finally. Five DEGs of the darkturquoise module are *FBLN1*, *POSTN*, *COL1A1*, *FBLN2,* and *DCN*. Five DEGs, *ALDH2*, *GCSH*, *FAAH*, *NGFRAP1*, and *HSPA5*, in the lightyellow module, were identified out. The DEGs of the red module are *KCNQ3*, *CCND1*, *MET*, *FN1*, *ITGA3*, *ITGA2*, *ITGB4*, *RUNX1*, *PERP*, *ICAM1*, *CDKN1A*, and *ST8SIA4*. In previous studies, six DEGs (*POSTN*, *COL1A1*, *ALDH2*, *CCND1*, *CDKN1A*, and *MET*) have been confirmed to be more or less associated with PTC. Consistent with our results, Puppin et al. identified that *POSTN* mRNA levels showed a significant increase in PTC, and were positively associated with higher tumor stage, extrathyroidal invasion, and distant metastasis [26]. Jolly et al. reported that collagen type I alpha 1 chain (*COL1A1*) was upregulated in PTC and promoted thyroid cancer progression through fibroblast recruitment and remodeling of the extracellular matrix [27]. *ALDH2* participate in the major oxidative pathway of alcohol metabolism and alcohol consumption is a recognized risk factor for head and neck cancer (HNC) [28]. Another research confirmed that genetic polymorphisms of *ADH1B* and *ALDH2* may modify the risk between alcohol drinking and HNC risk [29]. *CCND1* has been determined to be overexpressed in PTC tissues [30, 31], and controls the development of PTC [32]. *CDKN1A*, also known as *P21*, was also identified to be over-expressed in PTC in relation to the degree of progression [33]. *MET* is a member of the receptor tyrosine kinase family and controls apoptosis by activating of *PI3K* and *AKT* [34–36]. Koo et al. found that the *HGF/c-Met* axis is correlated with subclinical lymph node metastasis of the central neck in PTC and induce cell migration and invasion though *AKT* activation [37, 38].

Moreover, patient relapse-free survival analyses show that some DEGs could effectively stratify the high and low recurrence risk of PTC patients in TCGA. In particular, 11 genes (*KCNQ3*, *MET*, *FN1*, *ITGA3*, *RUNX1*, *ITGA2*, *PERP*, *GCSH*, *FAAH*, *NGFRAP1*, and *HSPA5*) may serve as prognostic biomarkers for PTC. Except for *MET*, the mechanism of the remaining ten genes in PTC is still unclear. *FN1* is an epithelial-to-mesenchymal transition-related gene and may stimulate cancer invasion and survival through the *PI3K/AKT* signaling pathway [39], another research shows that fibronectin mediates cancer cell association and directional migration [40]. *ITGA2* and *ITGA3* are components of integrin that function as cell surface adhesion molecules, several studies identified that patients with high *ITGA2* or *ITGA3* expression had poorer overall survival and relapse-free survival than patients with low *ITGA2* or *ITGA3* expression [41–43]. *RUNX1* is considered to be a main regulator of hematopoiesis and may play a dual function role in the carcinogenesis. For instance, *RUNX1* acts as an oncogene in prostate cancer and gastric cancer [44, 45], but plays a protective role in breast cancer [46]. *PERP* is a *p53/p63*-regulated gene and encodes a desmosomal protein. Some studies identified that *PERP* promotes cancer cell apoptosis, desmosome construction, and thus plays a pivotal role in tumor suppression [47, 48], which is inconsistent with our findings. The study of *NGFRAP1* is limited, and it showed a major upregulation in the nasopharyngeal carcinoma, and might induce the cisplatin-resistant phenotype [49]. *HSPA5* protein regulates the unfolded protein response, which is activated in various cancers, including breast cancer and head and neck cancer, and it is considered critical for tumor progression [50, 51]. *FAAH* encodes an enzyme, which is responsible for fatty acid amide hydrolysis; an inhibitor of *FAAH* could reduce tumorigenesis and progression of colorectal and prostate cancer, and represents a target of anticancer drugs [52, 53]. However, one gene may function as different roles in various cancers. In our study, *NGFRAP1*, *HSPA5*, and *FAAH* were downregulated in PTC samples, and patients with low expression have shorter relapse-free survival. *KCNQ3* encodes a protein that functions in the regulation of neuronal excitability and *GCSH* encodes a mitochondrial protein, which make up the glycine cleavage system, there is still a lack of research on the mechanism of these two genes in cancer.

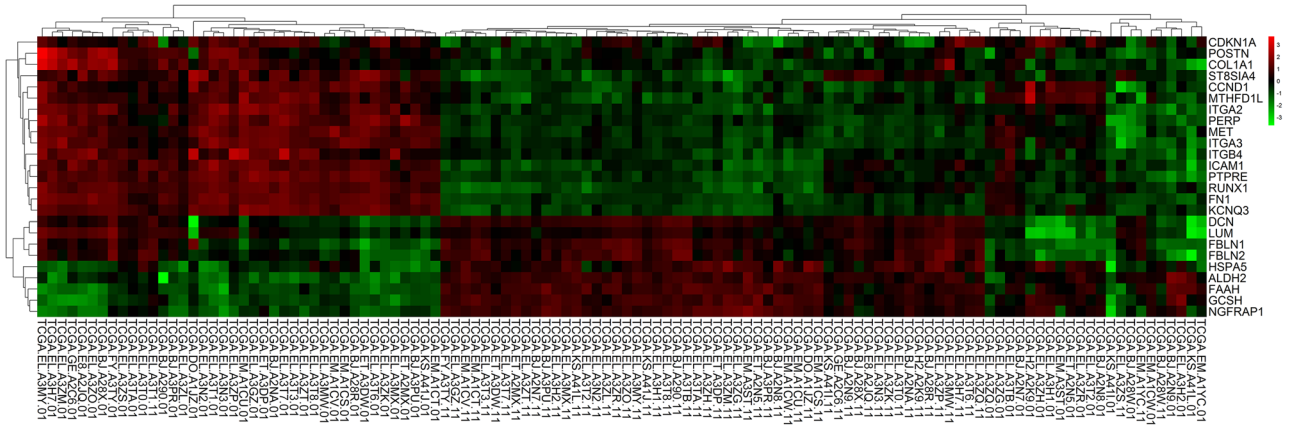


Fig. 5 The heatmap on expressions of the 25 DEGs for PTC (the *X* axis represents samples, while the *Y* axis represents the DEGs)

Table 2 DEGs in the key co-expression modules

Module	Gene symbol	Official full name	*P*-value	Fold change
Darkturquoise	DCN	Decorin	5.77E−15	−3.61
	FBLN1	Fibulin 1	3.78E−16	−2.99
	POSTN	Periostin	1.34E−05	1.47
	COL1A1	Collagen type I alpha 1	1.03E−04	1.36
	FBLN2	Fibulin 2	1.24E−11	−1.48
Lightyellow	ALDH2	Aldehyde dehydrogenase 2 family member	7.34E−17	−1.21
	GCSH	Glycine cleavage system protein H	3.78E−16	1.40
	NGFRAP1	Brain-expressed X-linked 3	8.53E−16	−1.01
	FAAH	Fatty acid amide hydrolase	2.83E−15	−1.00
	HSPA5	Heat-shock protein family A (Hsp70) member 5	2.80E−11	−1.07
Red	KCNQ3	Potassium voltage-gated channel subfamily Q member 3	1.10E−22	3.73
	CCND1	Cyclin D1	8.21E–18	1.52
	MET	MET proto-oncogene, receptor tyrosine kinase	8.06E–18	2.10
	FN1	Fibronectin 1	4.99E–17	4.26
	ITGA3	Integrin subunit alpha 3	1.19E–13	1.25
	ST8SIA4	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 4	1.31E–08	1.32
	RUNX1	Runt-related transcription factor 1	3.78E–16	2.32
	ITGA2	Integrin subunit alpha 2	2.16E–13	2.23
	PERP	TP53 apoptosis effector	6.31E–12	1.24
	ICAM1	Intercellular adhesion molecule 1	4.57E–11	1.80
	ITGB4	Integrin subunit beta 4	1.82E–09	1.36
	CDKN1A	Cyclin-dependent kinase inhibitor 1A	1.43E–06	1.17

The differential mRNA expressions of these 11 genes between PTC cell lines and normal thyroid epithelial cell line were further confirmed by qRT-PCR. Gene expression profiles are associated with genetic background and gene expression may vary with specific mutations. One study identified that both *ITGA2* and *ITGA3* have a higher expression level in PTC tissue compared with normal thyroid tissue, but *ITGA2* expression was not upregulated in *BRAF*-positive samples [54]. Similarly, *ITGA3* was upregulated in B-CPAP and TPC-1 cell lines, but *ITGA2* was not upregulated in the B-CPAP cell line harboring the *BRAF* mutation. *FAAH*, *NGFRAP1*, and *HSPA5* were upregulated in the TPC-1 cell, not in the B-CPAP cell, while *RUNX1* and *PERP* were upregulated in the B-CPAP cell, not in the TPC-1 cell, which may be also associated with the different mutations in these two cell lines.

In the TCGA, the remaining six DEGs (*ST8SIA4*, *DCN*, *ITGB4*, *ICAM1*, *FBLN1*, and *FBLN2*) did not show a significant influence on the recurrence risk of PTC patients. The majority of the 22 DEGs have been proved to be associated with several other cancers, such as prostate cancer, ovarian cancer and breast cancer, but the molecular pathogenesis of these genes in PTC are still ambiguous. A further analysis and validation of the candidate PTC biomarkers reported here is necessary, especially those that have not yet been associated with PTC.

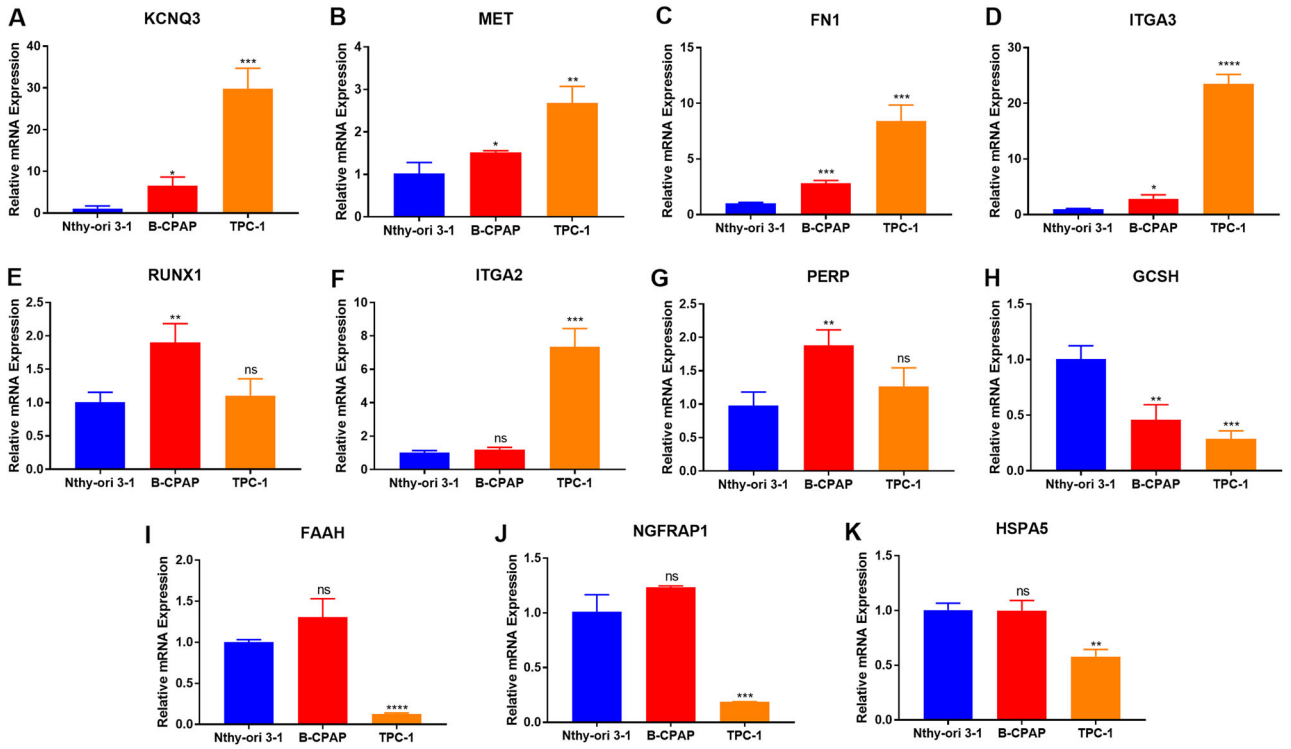


Fig. 6 Validation of 11 genes (*KCNQ3*, *MET*, *FN1*, *ITGA3*, *RUNX1*, *ITGA2*, *PERP*, *GCSH*, *FAAH*, *NGFRAP1*, and *HSPA5*) by qRT-PCR. **a**–**k** The relative expressions of *KCNQ3*, *MET*, *FN1*, *ITGA3*, *RUNX1*, *ITGA2*, *PERP*, *GCSH*, *FAAH*, *NGFRAP1*, and *HSPA5* were normalized to β-actin. The results were expressed as mean ± SD of three independent experiments. *P* values were determined based on comparison with Nthy-ori 3-1. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; $****P < 0.0001$; ns not significant

In summary, this study adopted WGCNA analysis to construct gene co-expression networks, identifying the key co-expression modules with their functional enrichment pathways and hub genes correlated with the progression and prognosis of PTC. Our findings provide new insights into lymph node metastasis of PTC, although the explicit molecular mechanisms need further analysis and validation.

Acknowledgements We thank all the authors for their contributions.

Funding This study was funded by the National Natural Science Foundation of China (grant number 81402207).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

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

References

1. W. Chen, R. Zheng, T. Zuo, H. Zeng, S. Zhang, J. He, National cancer incidence and mortality in China, 2012. Chin. J. Cancer Res. **28**(1), 1–11 (2016). https://doi.org/10.3978/j.issn.1000-9604.2016.02.08
2. C.M. Kitahara, J.A. Sosa, The changing incidence of thyroid cancer. Nat. Rev. Endocrinol. **12**(11), 646–653 (2016). https://doi.org/10.1038/nrendo.2016.110
3. Y. Wang, W. Wang, Increasing Incidence of Thyroid Cancer in Shanghai, China, 1983-2007. Asia Pac. J. Public Health **27**(2), NP223–NP229 (2015). https://doi.org/10.1177/1010539512436874
4. D.S. McLeod, Current concepts and future directions in differentiated thyroid cancer. Clin. Biochem Rev. **31**(1), 9–19 (2010)
5. M.J. Schlumberger, Papillary and follicular thyroid carcinoma. N. Engl. J. Med **338**(5), 297–306 (1998). https://doi.org/10.1056/nejm199801293380506
6. H. Li, Z. Tian, Y. Qu, Q. Yang, H. Guan, B. Shi, M. Ji, P. Hou, SIRT7 promotes thyroid tumorigenesis through phosphorylation and activation of Akt and p70S6K1 via DBC1/SIRT1 axis. Oncogene 1–13 (2018). https://doi.org/10.1038/s41388-018-0434-6
7. X. Li, X. Ruan, P. Zhang, Y. Yu, M. Gao, S. Yuan, Z. Zhao, J. Yang, L. Zhao, TBX3 promotes proliferation of papillary thyroid carcinoma cells through facilitating PRC2-mediated p57KIP2 repression. Oncogene **37**(21), 2773–2792 (2018). https://doi.org/10.1038/s41388-017-0090-2
8. Y. Li, Q. Yang, H. Guan, B. Shi, M. Ji, P. Hou, ZNF677 suppresses Akt phosphorylation and tumorigenesis in thyroid cancer. Cancer Res. 5216–5228 (2018). https://doi.org/10.1158/0008-5472.can-18-0003
9. S. Ock, J. Ahn, S.H. Lee, H. M. Kim, H. Kang, Y.-K. Kim, Y.-K., H. Kook, W.J. Park, S. Kim, S. Kimura, C.K. Jung, M. Shong, M. Holzenberger, E. Dale Abel, T.J. Lee, B.Y. Cho, H.-S. Kim, J. Kim, Thyrocyte-specific deletion of insulin and IGF-1 receptors induces papillary thyroid carcinoma-like lesions through EGFR pathway activation. Int. J. Cancer 1–13 (2018). https://doi.org/10.1002/ijc.31779

10. S. Tavazoie, J.D. Hughes, M.J. Campbell, R.J. Cho, G.M. Church, Systematic determination of genetic network architecture. Nat. Genet. **22**(3), 281–285 (1999). https://doi.org/10.1038/10343
11. A.P. Presson, E.M. Sobel, J.C. Papp, C.J. Suarez, T. Whistler, M. S. Rajeevan, S.D. Vernon, S. Horvath, Integrated weighted gene co-expression network analysis with an application to chronic fatigue syndrome. BMC Syst. Biol. **2**, 95–116 (2008). https://doi.org/10.1186/1752-0509-2-95
12. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. **9**, 559–572 (2008). https://doi.org/10.1186/1471-2105-9-559
13. M. Giulietti, G. Occhipinti, G. Principato, F. Piva, Identification of candidate miRNA biomarkers for pancreatic ductal adenocarcinoma by weighted gene co-expression network analysis. Cell. Oncol. **40**(2), 181–192 (2017). https://doi.org/10.1007/s13402-017-0315-y
14. X. Shao, B. Wang, K. Mu, L. Li, Q. Li, W. He, Q. Yao, X. Jia, J.-a.Zhang, Key gene co-expression modules and functional pathways involved in the pathogenesis of Graves' disease. Mol. Cell. Endocrinol. **474**, 252–259 (2018). https://doi.org/10.1016/j.mce.2018.03.015
15. N. Shved, G. Warsow, F. Eichinger, D. Hoogewijs, S. Brandt, P. Wild, M. Kretzler, C.D. Cohen, M.T. Lindenmeyer, Transcriptome-based network analysis reveals renal cell type-specific dysregulation of hypoxia-associated transcripts. Sci. Rep. **7**(1), 8576–8593 (2017). https://doi.org/10.1038/s41598-017-08492-y
16. G.J. Dennis, B.T. Sherman, D.A. Hosack, J. Yang, W. Gao, H.C. Lane, R.A. Lempicki, DAVID: database for annotation, visualization, and integrated discovery. Genome Biol. **4**(5), P3–P14 (2003)
17. M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat. Genet **25**(1), 25–29 (2000). https://doi.org/10.1038/75556
18. M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, M. Tanabe, KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. **40**(D1), D109–D114 (2011). https://doi.org/10.1093/nar/gkr988
19. C.V. Mering, STRING: a database of predicted functional associations between proteins. Nucleic Acids Res. **31**(1), 258–261 (2003). https://doi.org/10.1093/nar/gkg034
20. G. Alanis-Lobato, M.A. Andrade-Navarro, M.H. Schaefer, HIPPIE v2.0: enhancing meaningfulness and reliability of protein–protein interaction networks. Nucleic Acids Res. **45**(D1), D408–D414 (2017). https://doi.org/10.1093/nar/gkw985
21. P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. **13**(11), 2498–2504 (2003). https://doi.org/10.1101/gr.1239303
22. J.M. Wettenhall, G.K. Smyth, limmaGUI: a graphical user interface for linear modeling of microarray data. Bioinformatics **20** (18), 3705–3706 (2004). https://doi.org/10.1093/bioinformatics/bth449
23. Y.E. Nikiforov, M.N. Nikiforova, Molecular genetics and diagnosis of thyroid cancer. Nat. Rev. Endocrin. **7**(10), 569–580 (2011). https://doi.org/10.1038/nrendo.2011.142
24. V.A. LiVolsi, Papillary thyroid carcinoma: an update. Mod. Pathol. **24**(Suppl 2), S1–S9 (2011). https://doi.org/10.1038/modpathol.2010.129
25. K. Rue-Albrecht, P.A. McGettigan, B. Hernández, N.C. Nalpas, D.A. Magee, A.C. Parnell, S.V. Gordon, D.E. MacHugh, GOexpress: an R/Bioconductor package for the identification and visualisation of robust gene ontology signatures through supervised learning of gene expression data. BMC Bioinform. **17**(1), 126–138 (2016). https://doi.org/10.1186/s12859-016-0971-3
26. C. Puppin, D. Fabbro, M. Dima, C. Di Loreto, E. Puxeddu, S. Filetti, D. Russo, G. Damante, High periostin expression correlates with aggressiveness in papillary thyroid carcinomas. J. Endocrinol. **197**(2), 401–408 (2008). https://doi.org/10.1677/joe-07-0618
27. L.A. Jolly, S. Novitskiy, P. Owens, N. Massoll, N. Cheng, W. Fang, H.L. Moses, A.T. Franco, Fibroblast-mediated collagen remodeling within the tumor microenvironment facilitates progression of thyroid cancers driven by BrafV600E and Pten loss. Cancer Res. **76**(7), 1804–1813 (2016). https://doi.org/10.1158/0008-5472.can-15-2351
28. P. Boffetta, M. Hashibe, C. La Vecchia, W. Zatonski, J. Rehm, The burden of cancer attributable to alcohol drinking. Int. J. Cancer **119**(4), 884–887 (2006). https://doi.org/10.1002/ijc.21903
29. S.-T. Tsai, T.-Y. Wong, C.-Y. Ou, S.-Y. Fang, K.-C. Chen, J.-R. Hsiao, C.-C. Huang, W.-T. Lee, H.-I. Lo, J.-S. Huang, J.-L. Wu, C.-J. Yen, W.-T. Hsueh, Y.-H. Wu, M.-W. Yang, F.-C. Lin, J.-Y. Chang, K.-Y. Chang, S.-Y. Wu, H.-C. Liao, C.-L. Lin, Y.-H. Wang, Y.-L. Weng, H.-C. Yang, J.S. Chang, The interplay between alcohol consumption, oral hygiene, ALDH2 and ADH1B in the risk of head and neck cancer. Int. J. Cancer **135**(10), 2424–2436 (2014). https://doi.org/10.1002/ijc.28885
30. I. Bieche, B. Franc, D. Vidaud, M. Vidaud, R. Lidereau, Analyses of MYC, ERBB2, and CCND1 genes in benign and malignant thyroid follicular cell tumors by real-time polymerase chain reaction. Thyroid **11**(2), 147–152 (2001). https://doi.org/10.1089/105072501300042802
31. G.-Y. Liu, Q. Luo, B. Xiong, C. Pan, P. Yin, H.-F. Liao, W.-C. Zhuang, H.-Z. Gao, Tissue array for Tp53, C-myc, CCND1 gene over-expression in different tumors. World J. Gastroenterol. **14** (47), 7199–7207 (2008). https://doi.org/10.3748/wjg.14.7199
32. J. Sun, R. Shi, S. Zhao, X. Li, S. Lu, H. Bu, X. Ma, C. Su, E2F8, a direct target of miR-144, promotes papillary thyroid cancer progression via regulating cell cycle. J. Exp. Clin. Cancer Res. **36**(1), 1–14 (2017). https://doi.org/10.1186/s13046-017-0504-6
33. P.A. Siironen, Ivi, S. Nordling, J. Louhimo, R. Haapiainen, C. Haglund, Immunohistochemical expression of Bcl-2, Ki-67, and p21 in patients with papillary thyroid cancer. Tumor Biol. **26**(1), 50–56 (2005). https://doi.org/10.1159/000084340
34. G.H. Xiao, M. Jeffers, A. Bellacosa, Y. Mitsuuchi, W.G. Vande, J.R. Testa, Anti-apoptotic signaling by hepatocyte growth factor/Met via the phosphatidylinositol 3-kinase/Akt and mitogen-activated protein kinase pathways. Proc. Natl Acad. Sci. USA **98**(1), 247–252 (2001). https://doi.org/10.1073/pnas.011532898
35. S. Fan, Y.X. Ma, J.A. Wang, R.Q. Yuan, Q. Meng, Y. Cao, J.J. Laterra, I.D. Goldberg, E.M. Rosen, The cytokine hepatocyte growth factor/scatter factor inhibits apoptosis and enhances DNA repair by a common mechanism involving signaling through phosphatidyl inositol 3' kinase. Oncogene **19**(18), 2212–2223 (2000). https://doi.org/10.1038/sj.onc.1203566
36. P.W. Derksen, D.J. de Gorter, H.P. Meijer, R.J. Bende, M. van Dijk, H.M. Lokhorst, A.C. Bloem, M. Spaargaren, S.T. Pals, The hepatocyte growth factor/Met pathway controls proliferation and apoptosis in multiple myeloma. Leukemia **17**(4), 764–774 (2003). https://doi.org/10.1038/sj.leu.2402875
37. S.-N. Jung, H.S. Lim, L. Liu, J.W. Chang, Y.C. Lim, K.S. Rha, B. S. Koo, LAMB3 mediates metastatic tumor behavior in papillary thyroid cancer by regulating c-MET/Akt signals. Sci. Rep. **8**(1), 2718–2728 (2018). https://doi.org/10.1038/s41598-018-21216-0
38. B.S. Koo, J.M. Kim, S.T. Seo, Y.H. Yoon, K.R. Kwon, S.H. Kim, H.W. Kwon, W.J. Bae, Y.C. Lim, Upregulation of HGF and c-MET is associated with subclinical central lymph node metastasis

in papillary thyroid microcarcinoma. Ann. Surg. Oncol. **21**(7), 2310–2317 (2014). https://doi.org/10.1245/s10434-014-3553-5
39. Y.-X. Liao, Z.-P. Zhang, J. Zhao, J.-P. Liu, Effects of fibronectin 1 on cell proliferation, senescence and apoptosis of human glioma cells through the PI3K/AKT signaling pathway. Cell. Physiol. Biochem. **48**(3), 1382–1396 (2018). https://doi.org/10.1159/000492096
40. B. Erdogan, M. Ao, L.M. White, A.L. Means, B.M. Brewer, L. Yang, M.K. Washington, C. Shi, O.E. Franco, A.M. Weaver, S. W. Hayward, D. Li, D.J. Webb, Cancer-associated fibroblasts promote directional cancer cell migration by aligning fibronectin. J. Cell Biol. **216**(11), 3799–3816 (2017). https://doi.org/10.1083/jcb.201704053
41. J. Dong, R. Wang, G. Ren, X. Li, J. Wang, Y. Sun, J. Liang, Y. Nie, K. Wu, B. Feng, Y. Shang, D. Fan, HMGA2-FOXL2 axis regulates metastases and epithelial-to-mesenchymal transition of chemoresistant gastric cancer. Clin. Cancer Res. **23**(13), 3461–3473 (2017). https://doi.org/10.1158/1078-0432.CCR-16-2180.
42. X.-R. Tang, X. Wen, Q.-M. He, Y.-Q. Li, X.-Y. Ren, X.-J. Yang, J. Zhang, Y.-Q. Wang, J. Ma, N. Liu, MicroRNA-101 inhibits invasion and angiogenesis through targeting ITGA3 and its systemic delivery inhibits lung metastasis in nasopharyngeal carcinoma. Cell Death Dis. **8**(1), e2566–e2566 (2017). https://doi.org/10.1038/cddis.2016.486
43. T. Sakaguchi, H. Yoshino, M. Yonemori, K. Miyamoto, S. Sugita, R. Matsushita, T. Itesako, S. Tatarano, M. Nakagawa, H. Enokida, Regulation of ITGA3 by the dual-stranded microRNA-199 family as a potential prognostic marker in bladder cancer. Br. J. Cancer **116**(8), 1077–1087 (2017). https://doi.org/10.1038/bjc.2017.43
44. G. Zhang, G. Han, X. Zhang, Q. Yu, Z. Li, Z. Li, J. Li, Long non-coding RNA FENDRR reduces prostate cancer malignancy by competitively binding miR-18a-5p with RUNX1. Biomarkers **23**(5), 435–445 (2018). https://doi.org/10.1080/1354750x.2018.1443509
45. Y. Mitsuda, K. Morita, G. Kashiwazaki, J. Taniguchi, T. Bando, M. Obara, M. Hirata, T.R. Kataoka, M. Muto, Y. Kaneda, T. Nakahata, P.P. Liu, S. Adachi, H. Sugiyama, Y. Kamikubo, RUNX1 positively regulates the ErbB2/HER2 signaling pathway through modulating SOS1 expression in gastric cancer cells. Sci. Rep. **8**(1), 6423–6436 (2018). https://doi.org/10.1038/s41598-018-24969-w
46. M. Kulkarni, T.Z. Tan, S.N. Syed, J.M. Lamar, P. Bansal, J. Cui, Y. Qiao, Y. Ito, RUNX1 and RUNX3 protect against YAP-mediated EMT, stem-ness and shorter survival outcomes in breast cancer. Oncotarget **9**(18), 14175–14192 (2018). https://doi.org/10.18632/oncotarget.24419
47. I.A. Khan, B.H. Yoo, O. Masson, S. Baron, D. Corkery, G. Dellaire, L.D. Attardi, K.V. Rosen, ErbB2-dependent down-regulation of a pro-apoptotic protein Perp is required for oncogenic transformation of breast epithelial cells. Oncogene **35**(44), 5759–5769 (2016). https://doi.org/10.1038/onc.2016.109
48. V.G. Beaudry, D. Jiang, R.L. Dusek, E.J. Park, S. Knezevich, K. Ridd, H. Vogel, B.C. Bastian, L.D. Attardi, Loss of the p53/p63 regulated desmosomal protein Perp promotes tumorigenesis. PLoS Genet. **6**(10), e1001168–e1001185 (2010). https://doi.org/10.1371/journal.pgen.1001168
49. W. Gao, J.Z. Li, S.Q. Chen, C.Y. Chu, J.Y. Chan, T.S. Wong, BEX3 contributes to cisplatin chemoresistance in nasopharyngeal carcinoma. Cancer Med. **6**(2), 439–451 (2017). https://doi.org/10.1002/cam4.982
50. S.Y. Kim, H.J. Kim, H.J. Kim, D.H. Kim, J.H. Han, H.K. Byeon, K. Lee, C.H. Kim, HSPA5 negatively regulates lysosomal activity through ubiquitination of MUL1 in head and neck cancer. Autophagy **14**(3), 385–403 (2018). https://doi.org/10.1080/15548627.2017.1414126
51. H.A. Chen, Y.W. Chang, C.F. Tseng, C.F. Chiu, C.C. Hong, W. Wang, M.Y. Wang, M. Hsiao, J.T. Ma, C.H. Chen, S.S. Jiang, C. H. Wu, M.C. Hung, M.T. Huang, J.L. Su, E1A-mediated inhibition of HSPA5 suppresses cell migration and invasion in triple-negative breast cancer. Ann. Surg. Oncol. **22**(3), 889–898 (2015). https://doi.org/10.1245/s10434-014-4061-3
52. D. Fiore, M.C. Proto, S. Pisanti, P. Picardi, A.C. Pagano Zottola, S. Butini, S. Gemma, A. Casagni, C. Laezza, M. Vitale, A. Ligresti, V. Di Marzo, D.M. Zisterer, S. Nathwani, D.C. Williams, G. Campiani, P. Gazzerro, M. Bifulco, Antitumor effect of pyrrolo-1,5-benzoxazepine-15 and its synergistic effect with Oxaliplatin and 5-FU in colorectal cancer cells. Cancer Biol. Ther. **17**(8), 849–858 (2016). https://doi.org/10.1080/15384047.2015.1078028
53. M.P. Endsley, R. Thill, I. Choudhry, C.L. Williams, A. Kajdacsy-Balla, W.B. Campbell, K. Nithipatikom, Expression and function of fatty acid amide hydrolase in prostate cancer. Int. J. Cancer **123**(6), 1318–1326 (2008). https://doi.org/10.1002/ijc.23674
54. G. Chernaya, N. Mikhno, T. Khabalova, S. Svyatchenko, L. Mostovich, S. Shevchenko, L. Gulyaeva, The expression profile of integrin receptors and osteopontin in thyroid malignancies varies depending on the tumor progression rate and presence of BRAF V600E mutation. Surg. Oncol. **27**(4), 702–708 (2018). https://doi.org/10.1016/j.suronc.2018.09.007