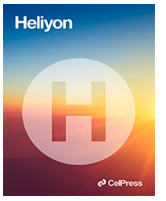
Heliyon 10 (2024) e27928

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Integrated transcriptome sequencing and weighted gene co-expression network analysis reveals key genes of papillary thyroid carcinomas

Lingfeng Pan [a], Lianbo Zhang [a], Jingyao Fu [b], Keyu Shen [b], Guang Zhang [b,*]

[a] Department of Plastic Surgery, China-Japan Union Hospital of Jilin University, Changchun, Jilin, 130033, China
[b] Department of Thyroid Surgery, China-Japan Union Hospital of Jilin University, Changchun, Jilin, 130033, China

ARTICLE INFO

Keywords:
Papillary thyroid carcinomas
WGCNA
MEXPREWSS analysis
Enrichment analysis
Differentially expressed genes

ABSTRACT

Objective: Papillary thyroid carcinoma (PTC) accounts for the majority of thyroid cancers and has a high recurrence rate. We aimed to screen key genes involved in PTC to provide novel insights into the mechanisms of PTC.

Methods: Seven microarray datasets of PTC were downloaded from gene expression omnibus database. Differentially expressed genes (DEGs) between PTC and normal samples were screened in the merged dataset. Then, protein-protein interaction (PPIs) functional modules analysis and weighted gene co-expression network analysis (WGCNA) were utilized to identify PTC-associated key genes. The identified key genes were then characterized from various aspects, including gene set enrichment analysis (GSEA) and the associations with immune infiltration, methylation levels and prognosis.

Results: A large numbers of DEGs were identified, and these DEGs are involved in several cancer pathways. Nine key genes (including down-regulated genes GNA14, AVPR1A, and WFS1, and up-regulated genes LAMB3, PLAU, MET, MFGE8, PRSS23, and SERPINA1) were identified. Patients in the AVPR1A and GNA14 high expression groups had better disease-free survival (DFS) than those in the low expression group. Key genes were mainly involved in P53 pathway, estrogen response, apoptosis, glycolysis, NOTCH signaling, epithelial mesenchymal transition, WNT_beta catenin signaling, and inflammatory response. The expression of key genes was associated with immune cell infiltration and corresponding methylation levels. The verification results of key gene proteins and mRNA expression levels using external validation datasets were consistent with our expectations, implying the involvements of key genes in PTC.

Conclusion: The key genes may serve as potential therapeutic targets for PTC. This study provides novel insights into the mechanisms underlying PTC development.

1. Introduction

Thyroid cancer is the ninth most common cancer worldwide and is three times more common in women than in men [1]. In the last few years, the incidence of thyroid cancer has continuously increased, mainly because of the increased detection rate of papillary thyroid carcinoma (PTC) [2], which is the most common thyroid malignancy, accounting for approximately 84% of all thyroid cancers

* Corresponding author.
E-mail address: zhangguang@jlu.edu.cn (G. Zhang).

https://doi.org/10.1016/j.heliyon.2024.e27928
Received 21 June 2023; Received in revised form 7 March 2024; Accepted 8 March 2024
Available online 15 March 2024

List of abbreviations

PTC	Papillary thyroid carcinomas
DEGs	Differentially expressed genes
DEMI	Differentially expressed miRNAs
WGCNA	weighted gene co-expression network
GSEA	gene set enrichment analysis
DFS	Disease free survival
FNAB	Fine needle aspiration biopsy
GEO	Gene expression omnibus
GO_BP	Gene ontology biological process
KEGG	Kyoto encyclopedia of genes and genomes
HPA	Human Protein Atlas
PPI	Protein-protein interactions
KLF7	Krüppel-like factor 7
AVPR1A	Arginine vasopressin receptor 1A
WFS1	Wolframin ER transmembrane glycoprotein
LAMB3	Laminin subunit beta 3
MFGE8	Milk fat globule EGF and factor V/VIII domain containing
PRSS23	Serine protease 23
SERPINA1	Serpin family a member 1

[3]. As an indolent tumor, there is still a risk that some cancer cells will metastasize to the lymph nodes around the thyroid gland [4], and lymph node metastasis is independently associated with an increased risk of recurrence and distant metastases [5,6]. Fine-needle aspiration biopsy (FNAB) and cytological assessment are the foundation for the diagnostic management of thyroid nodules; however, they cannot provide definitive diagnoses in 20–30% of patients, and further diagnostic surgery is needed to identify the characteristics of thyroid nodules [7]. Suitable molecular markers are helpful for guiding surgery and monitoring cases of uncertain FNAB [7]. Therefore, the identification of key genes involved in the pathogenesis of PTC is important for the improvement of clinical diagnosis and application of precision medicine in PTC [8].

In the last few years, a large number of molecular genetic studies have provided new insights into PTC, and molecular markers, such as RAS and BRAF(V600E) mutations and RET/PTC rearrangements, have improved the diagnosis and treatment of PTC [9]. For example, Kasko et al. showed that molecular marker-based stratification of thyroid lesions with indeterminate FNAB results contributes to the delivery of a more tailored therapy [9]. These findings highlight the necessity of identifying the key genes involved in PTC pathogenesis. Owing to the progress in RNA sequencing and microarray technology, gene expression profile analysis has widened our understanding not only in terms of the molecular mechanism of PTC but also in identifying molecular markers. Several studies have investigated key genes in PTC based on integrated bioinformatics analyses. However, these studies were conducted based on either a single dataset or a single level of analysis [10–12].

Weighted gene co-expression network analysis (WGCNA) is an algorithm that analyzes the correlation patterns among genes based on gene expression data and can help identify not only the potential modules of highly correlated genes but also the associations of these gene modules with various clinical traits [13]. WGCNA is widely used to identify diagnostic candidates for various diseases [14–16]. In this study, seven PTC datasets were used, and key genes were screened using differential expression analysis, topological analysis of the protein–protein interaction (PPI) network, and the WGCNA algorithm. Additionally, based on miRNA expression profiles and online databases, miRNAs targeting the identified key genes were predicted to uncover their possible regulatory networks. Most importantly, based on TCGA data, the identified key genes were characterized from various aspects, including associated pathways, immune associations with gene expression and gene copy numbers, associations with methylation levels, and prognostic values, which were not investigated in any of the previous studies.

2. Methods

This is an integrated bioinformatics analysis, and the workflow for this analysis is shown in Supplementary Fig. 1.

2.1. Data collection and preprocessing

Seven gene expression datasets were downloaded from the Gene Expression Omnibus (GEO) database [17], of which six (GSE33630, GSE53157, GSE54958, GSE151179, GSE73182, and GSE151180) were used as discovery datasets, and GSE64912 was used as the validation dataset. Among these discovery datasets, GSE73182 and GSE151180 were miRNA expression profiles and the rest were gene expression profiles. The standardized probe expression matrix were downloaded, and the probes were annotated based on the annotation files provided on the corresponding platform. The probes and gene symbols were individually matched. For different probes mapped to the same gene symbol, the average expression value of diverse probes was determined as the expression value of the

gene. For GSE64912, the normalized RPKM values were downloaded and log2 transformed. Detailed information on the dataset is provided in Supplementary Table 1. Four gene expression datasets and two miRNA expression datasets were merged separately, and the batch effect among diverse datasets was eliminated using the sva package [18]. Merged datasets were used for subsequent analyses.

2.2. Differentially expressed gene (DEG)and differentially expressed miRNA (DEMI) screening

The limma package [19] was used to identify the DEGs and DEMIs between the PTC and normal groups. The false discovery rate after Benjamini-Hochberg correction (<0.05) and $|\log2FC \text{ (fold change)}| > 1$ were set as the thresholds for screening DEGs and DEMIs.

2.3. Pathway and function enrichment analysis of DEGs

The biological functions of the DEGs were explored by functional enrichment. Specifically, the DAVID [20] tool was used to analyze the gene ontology biological process (GO_BP) [21] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [22] pathways involved in DEGs. To obtain comprehensive enrichment results, terms with p value $<$ of 0.05 [23] and enrichment count ≥ 2 were considered as significantly enriched terms.

2.4. PPI network analysis

The STRING database [24] was used to predict interactions among the proteins encoded by DEGs. DEGs were entered into the STRING database and the species were set to homo. PPI score was set to 0.9 (highest score). Cytoscape software [25] (version 3.4.0) was used to construct a PPI network based on the obtained interactions. CytoNCA [26] (version 2.1.6) was used to analyze the connectivity between nodes without weight. In addition, MCODE [27] (version 1.5.1) of the Cytoscape software was used to screen functional modules from the PPI network. The default parameters were as follows: degree cutoff = 2, node score cutoff = 0.2, k-core = 2 and maximum depth = 100. KEGG analysis was performed for genes in the PPI modules using clusterProfiler [28] (version 3.8.1) in R. P significance was set at $P < 0.05.3$. Modules enriched in multiple pathways were considered as key modules.

2.5. WGCNA

After performing variance analysis on the expression matrices of all genes, the top 20% of genes with inter-sample variability were analyzed using WGCNA (version 1.61) [13] in R to identify highly coordinated gene set modules. WGCNA is a comprehensive biological algorithm that has significant advantages for analyzing association patterns between genes. Module clustering of genes with similar expression patterns and correlation analysis between modules and clinical traits are the two highlights of WGCNA [29]. First, a soft threshold was selected to ensure that the constructed network conforms to a power-law distribution. Second, a scale-free weighted gene co-expression network was constructed using a blockwise module function. Third, module partition analysis was performed to identify gene co-expression modules that could group genes with similar expression patterns. To identify co-expressed gene modules, we chose minModuleSize = 30 and MEDissThres = 0.2 as thresholds. For important genes, GS $>$ 0.6 and MM $>$ 0.8 were set as the thresholds. The intersecting genes of the important genes obtained in WGCNA and the PPI module were defined as key genes related to the disease for subsequent analysis.

2.6. Gene set enrichment analysis (GSEA)

Based on the combined gene expression profile, after eliminating the batch effect, we performed GSEA using hallmark gene sets (h. all.v7.4. symbols. gmt) [30]. The Pearson correlation coefficients of the key genes and other genes were calculated and ranked. The significant enrichment threshold was set at a nominal p value < 0.05.

2.7. Relationship between key genes and immune cell infiltration abundance

Gene modules in the TIMER database [31] were used to show the relationship between the abundance of immune infiltration and expression of each key gene. The analysis parameters were set to default values. In addition, compared with normal samples, the effect of different gene copy numbers on immune infiltration was analyzed based on the data of GISTIC2.0. We also used the SCAN module in TIMER to explore the relationship between somatic copy number variation and immune infiltration of each key gene.

2.8. Key genes' protein expression and methylation analysis

The Human Protein Atlas (HPA) database [32] was used to determine the expression of key genes in TCP and normal tissues. MEXPRESS [33] is a data visualization tool used to simplify the visualization of TCGA expression, clinical data, and DNA methylation and to analyze the relationship between them. In this study, we used the online website MEXPREWSS to analyze the correlation between each key gene and its corresponding methylation site. The parameters used were the default settings on the website. Tumor histology samples for classical/uterine examination were selected for further study.

2.9. Predictive analysis of upstream miRNAs of key genes

Based on the miRWalk, Microt4, miRanda, PITA, and RNA22 databases, miRNAs upstream of key genes were predicted using mirwalk2.0 [34]. If a predicted miRNA-target relationship pair appeared at least three times in the prediction results of the six databases, it was considered that this miRNA regulated the corresponding target gene. The intersection of the predicted miRNA and DEMI was selected to further identify miRNA-target pairs related to PTC.

2.10. Verification of the expression of key genes and miRNAs

GSE64912 was used as the validation set for key gene verification. *P* value was obtained using a *t*-test. The expression of miRNAs between the two groups was verified using the UALCAN online tool. We also used the online tool gene expression profiling interactive analysis (GEPIA) to verify the expression of key genes and miRNAs. GEPIA is a web-based tool for mining and understanding gene functions in tumors based on RNA-Seq data from TCGA and GTEx [35]. |Log2FC| cutoff was set to 1, *p* value cutoff was 0.01, and it matched TCGA normal data.

2.11. Prognostic analysis of key genes and correlation analysis with clinical factors

Patients with PTC were divided into high and low expression groups according to the median expression value of the key genes. To explore the relationship between gene expression levels and prognosis of PTC patients, Kaplan-Meier (K-M) curves between the two groups were drawn using GEPIA2 [36]. The p value < 0.05 was set as significant. The expression of key genes in different clinical factor (tumor stage, tumor histology, and nodal metastasis status) groups of patients was determined using the online tool UALCAN [37]. We also downloaded the clinical information of PTC patients, including age, sex, TNM stage, tumor stage, and survival information, from TCGA database. Univariate Cox regression analysis was used to analyze the correlation between clinical factors, key genes, and disease free survival (DFS). Multivariate Cox regression analysis was used to screen for independent prognostic factors for PTC.

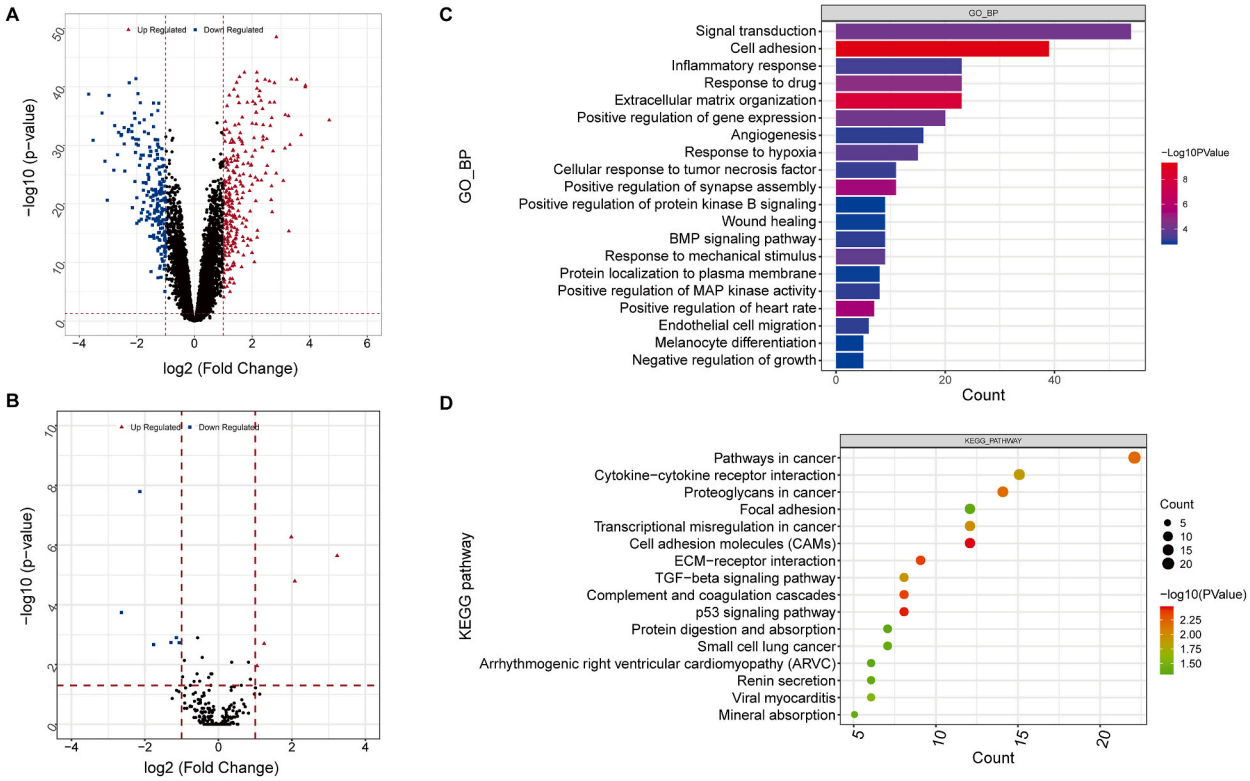


Fig. 1. Differential expression analysis and functional enrichment A-B,
Volcano plots of differentially expressed genes (A) and differentially expressed miRNAs (B); C-D, top 20 GO BP terms (C) and 16 KEGG pathways (D) enriched in differentially expressed genes.

3. Results

3.1. Identification of DEGs and DEMIs

Four mRNA expression datasets (GSE33630, GSE53157, GSE54958, and GSE151179) and two miRNA datasets (GSE73182 and GSE151180) were merged into larger datasets for subsequent analyses. After eliminating batch effects, there was no obvious separation among the diverse datasets, as shown in the principal component analysis (Supplementary Fig. 2). A total of 478 DEGs (264 upregulated and 214 downregulated) and 11 DEMIs (5 upregulated and 6 downregulated) were identified between PTC and normal samples (Fig. 1A–B).

3.2. Enrichment analysis of DEGs

In total, 478 DEGs were involved in 124 GO BP terms and 16 KEGG pathways. The top 20 GO BP terms and all KEGG pathways are shown in Fig. 1C–D. These DEGs were mainly involved in signal transduction, cell adhesion, inflammatory response, cancer pathways, cytokine-cytokine receptor interactions, and proteoglycans in cancer.

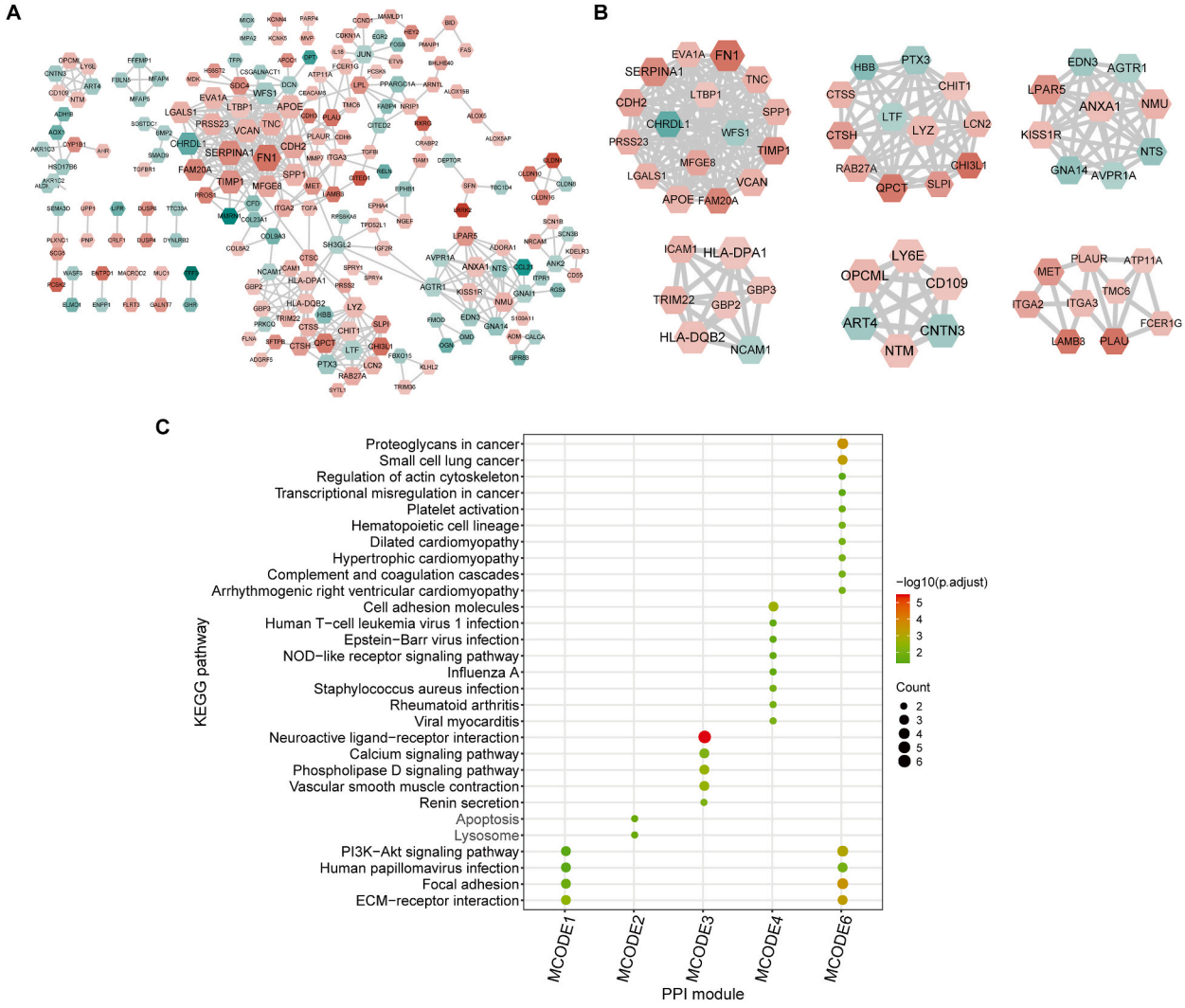


Fig. 2. Protein–protein interaction (PPI) analysis.
A, PPI network. B, PPI modules. C KEGG pathways enriched by genes in the modules. Green nodes represent downregulated genes and red nodes represent upregulated genes. The darker the color, the greater the absolute value of the difference multiples.

3.3. PPI construction and module analysis

The PPI network comprised 196 proteins and 449 interacting edges (Fig. 2A). Further module analysis yielded six modules with node degrees greater than five. MCODE1–MCODE6 contained 16, 12, 9, 7, 6, and 9 proteins, respectively (Fig. 2B). MCODE1 is mainly involved in the ECM-receptor interaction pathway, and MCODE2 is enriched in apoptosis. Most genes in MCODE3 are involved in neuroactive ligand-receptor interactions, and genes in MCODE3 are mainly involved in cell adhesion. Genes in MCODE6 were mainly

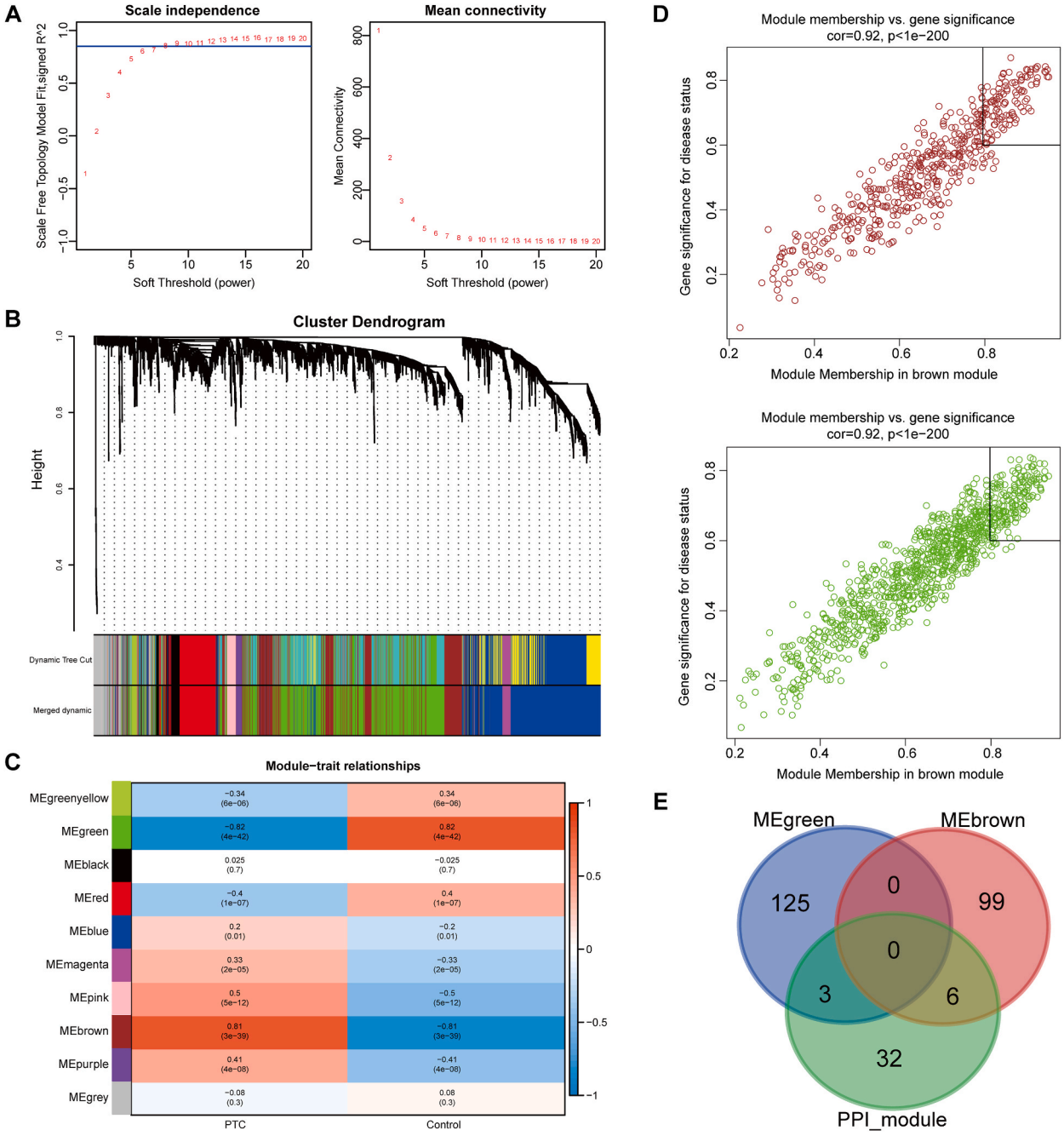


Fig. 3. WGCNA.
A, Calculation of soft threshold (power) in WGCNA by balancing mean connectivity and scale independence; B, Clustering dendrogram of the top 20% genes with a greater degree of variation. C Module–trait relationships. D, Gene significance for disease status (GS) and module memberships in the green module (MM) correlation scatter diagrams of brown and green module genes. The genes in the upper right corner represent the key genes of the module; E, Venn plot of common genes between the PPI and WGCNA modules.

encoded by proteoglycans in cancer and focal adhesion pathways (Fig. 2C). Module 5 was not significantly enriched in any of the pathways. Considering that MCODE5 had no enrichment results and MCODE2 had fewer enriched pathways, MCODE1, MCODE3, MCODE4, and MCODE6 were the focus of follow-up analysis. These four modules contained 41 genes.

3.4. WGCNA and key gene identification

The top 20% genes with large variants were selected for WGCNA. To calculate the adjacency matrix, the optimal soft-threshold power was set at 10 (Fig. 3A). A cluster dendrogram with the adjacent matrix was constructed, and finally 10 color modules (magenta, pink, grey, purple, red, black, blue, brown, green, and green yellow) were identified. The grey module included genes that could not be included in any of the other modules (Fig. 3B). Next, we performed a correlation analysis between the color modules and traits. As shown in Fig. 3C, the four modules (magenta, pink, brown, and purple) were significantly positively correlated with PTC and negatively correlated with the control. The three modules (green-yellow, green, and red) were significantly negatively correlated with PTC and positively correlated with the control ($|R| > 0.3$). Modules green and brown, which were significantly positively correlated

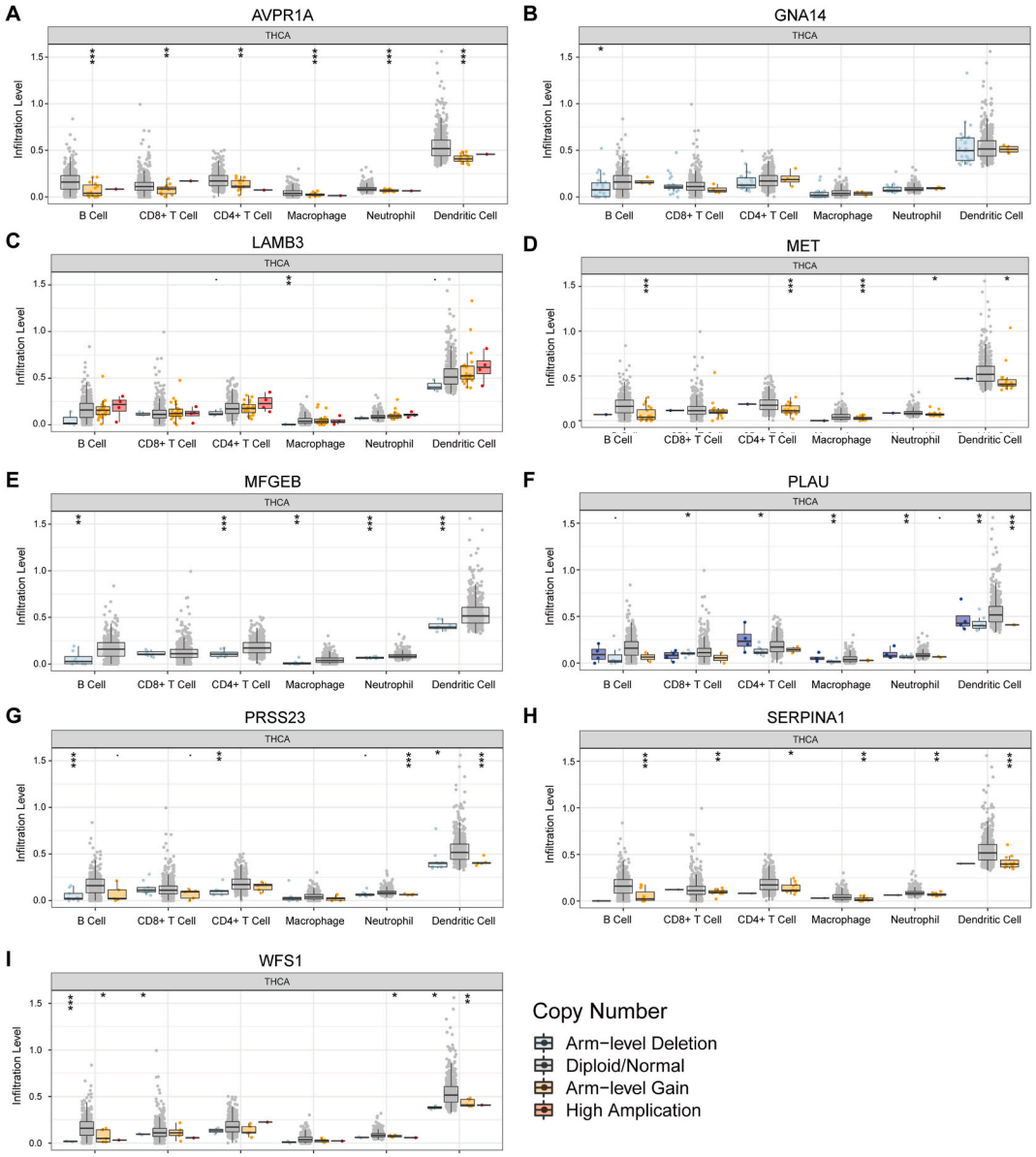


Fig. 4. Level of immune cell infiltration of key genes in different variation types.
Level of immune cell infiltration of AVPR1A (A), GNA14 (B), LAMB3 (C), MET (D), MFGEB (E), PLAU (F), PRSS23 (G), SERPINA1 (H) and WFS1 (I) in different variation types.

with PTC and control, respectively, were selected for correlation analysis with genes that were significantly related to disease status. As shown in Fig. 3D, both the brown and green modules were significantly positively correlated with genes that were significantly related to disease status (cor = 0.92, $p < 1e\text{-}200$). These results indicate that the genes in these modules may participate in PTC.

Furthermore, the common genes in the green and brown modules and genes in the four PPI modules were screened using Venn analysis (Fig. 3E). Finally, three downregulated genes (GNA14, AVPR1A, and WFS1) and six upregulated genes (LAMB3, PLAU, MET, MFGE8, PRSS23, and SERPINA1) were identified as key genes associated with PTC.

3.5. GSEA of key genes in cancer related pathways

GNA14, AVPR1A, WFS1, LAMB3, PLAU, MET, MFGE8, PRSS23, and SERPINA1 were significantly enriched in 0, 1, 4, 9, 14, 7, 13, 2, and 9 significantly cancer-related pathways, respectively. AVPR1A expression was significantly negatively correlated with the P53 pathway. WFS1 expression was significantly positively correlated with the reactive oxygen species pathway. Six up-regulated genes were related with P53 pathway, epithelial mesenchymal transition, estrogen response, apoptosis, glycolysis, NOTCH signaling, TGF_beta signaling, inflammatory response, and WNT_beta catenin signaling (Supplementary Table 2).

3.6. Association analysis between key genes and immune cell infiltration abundance

We calculated the correlation coefficient between immune cells infiltration and key genes to evaluate the correlation between the two (Supplementary Fig. 3). There was no significant correlation between the three downregulated genes (GNA14, AVPR1A, and WFS1) and the six immune cell types. The expression of LAMB3 and MET was significantly and positively correlated with that of B cells,

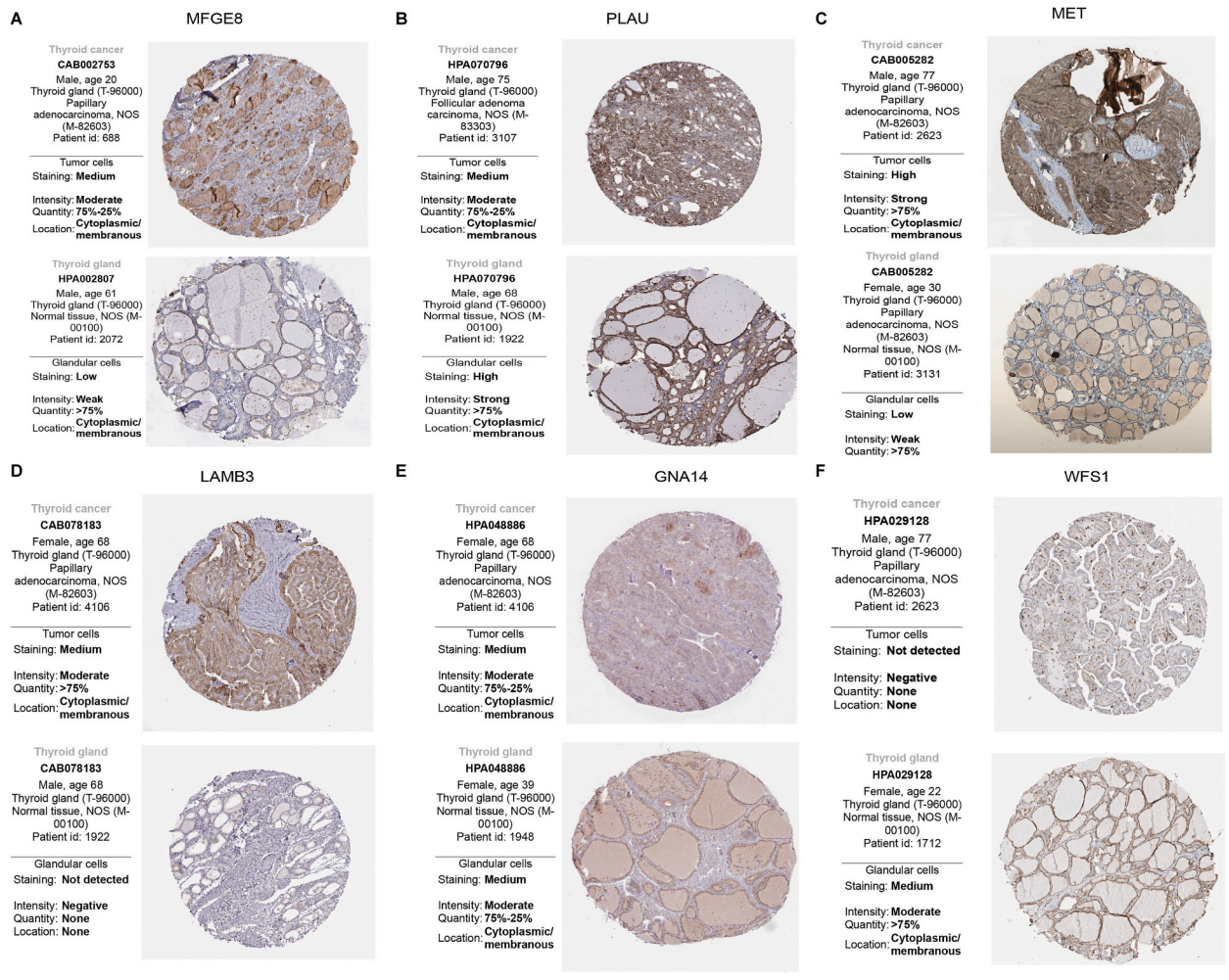


Fig. 5. Immunohistochemical staining of key gene proteins in normal tissues and papillary thyroid cancer tissues.
Immunohistochemical staining of MFGEB (A), PLAU (B), MET (C), LAMB3 (D), GNA14 (E) WFS1 (F) in normal tissues and papillary thyroid cancer tissues.

$CD4^+$ T cells, macrophages, neutrophils, and dendritic cells ($p << 0.05$, $r > 0.3$). The expression of MFGE8, PLAU, PRSS23, and SERPINA1 was significantly positively correlated with that of B cells, $CD4^+$ T cells, neutrophils, and dendritic cells ($p << 0.05$, $r > 0.3$).

Furthermore, we analyzed the level of immune cell infiltration of key genes with different mutation types (Fig. 4A–I). Except for GNA14 and LAMB3, there were significant differences in the level of immune cell infiltration under different mutation states for the other key genes ($p << 0.05$).

3.7. Protein expression and methylation level of key genes

The HPA database was used to retrieve immunohistochemical staining results for key gene proteins in normal thyroid and PTC tissues. As shown in Fig. 5A–F, the staining of MFGE8, PLAU, MET, and LAMB3 in cancer tissues was deeper, indicating a higher protein expression of these four genes in cancer tissues. In contrast, the staining of GNA14 and WFS1 in cancer tissues was lower than that in normal tissues, indicating relatively low protein expression of these two genes in cancer tissues. Immunohistochemical staining data for the other three genes were not retrieved from the HPA database.

It is generally considered that DNA methylation within the CpG island region of a gene's promoter region correlates with gene expression, whereas DNA methylation within the gene body is associated with chromosomal integrity [38]. The correlation between the methylation site level of the key gene and the expression level of the corresponding gene was analyzed using MEXPREWSS (Fig. 6A–I). The results indicated that the expression levels of most key genes were significantly negatively correlated with DNA methylation ($p < 0.05$).

3.8. Construction of miRNA-target network

A total of 4323 miRNA-target relationship pairs were predicted for the nine key genes, and the miRNA-target relationship pairs were further screened by taking the intersection with the 11 DEMIs above. These 11 DEMIs included six downregulated (hsa-miR-451a,

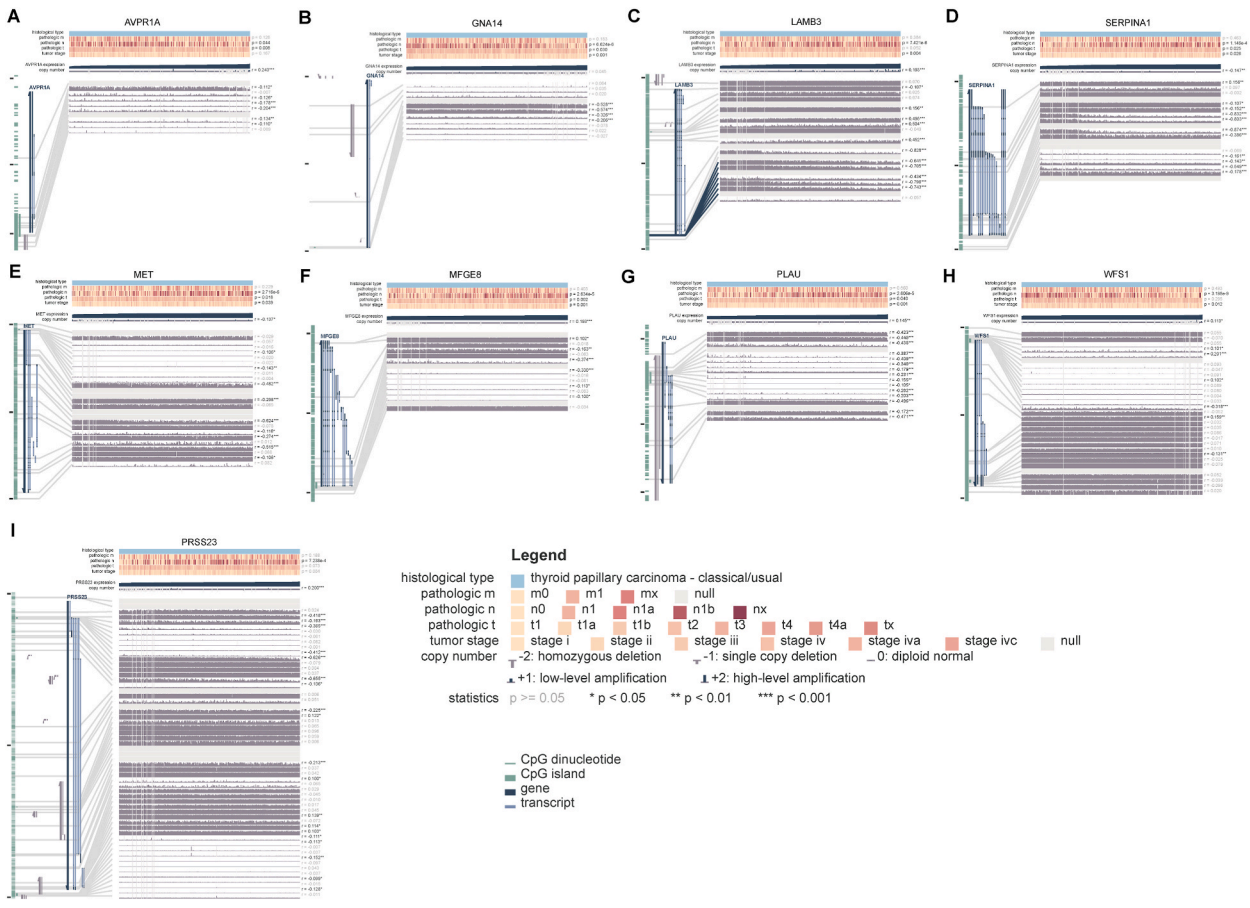


Fig. 6. Correlation between DNA methylation and expression level of key genes.
Correlation between DNA methylation and expression level of AVPR1A (A), GNA14 (B), LAMB3 (C), SERPINA1 (D), MET (E), MFGEB (F), PLAU (G), WFS1 (H) and PRSS23 (I).

hsa-miR-7-5p, hsa-miR-144–3p, hsa-miR-199b-5p, hsa-miR-204–5p, and hsa-miR-100–5p) and five upregulated (hsa-miR-551b-3p, hsa-miR-21–5p, hsa-miR-222–3p, hsa-miR-221–3p, and hsa-miR-146b-5p) genes. The expression of these 11 DEMIs was validated using the UALCAN database and consistent results were obtained (Fig. 7A).

Eleven miRNA-target relationship pairs were screened, involving six key genes and six miRNAs (Fig. 7B). A regulatory relationship existed between hsa-miR-7-5p and PRSS23, SERPINA1, and MFGE8. hsa-miR-204–5p regulated the expression of PRSS23, MFGE8, and WFS1. MET and PLAU were regulated by hsa-miR-144–3p and hsa-miR-199b-5p, respectively. hsa-miR-221–3p regulated WFS1 expression and hsa-miR-222–3p targeted PRSS23 and WFS1.

3.9. Verification of the nine key genes using validation dataset and TCGA data

The GSE64912 dataset was used to verify the expression of nine key genes (Fig. 7C). Consistent with the predicted results, the expression levels of GNA14, AVPR1V, and WFS1 in the tumor tissues were significantly lower than those in the control group ($p \leq 0.05$), whereas the expression levels of LAMB3, PLAU, MET, MFGE8, PRSS23, and SERPINA1 in the tumor tissues were significantly higher than those in the control group ($p < < 0.01$).

In addition, the expression of nine key genes was verified in TCGA database. As shown in Fig. 7D, the expression levels of these nine genes were significantly different between the papillary carcinoma and normal samples. Among these, GNA14, AVPR1A, and WFS1 were significantly upregulated in papillary carcinoma samples, whereas LAMB3, PLAU, MET, MFGE8, PRSS23, and SERPINA1 were significantly downregulated in papillary carcinoma samples, consistent with our previous prediction.

3.10. Prognostic value of key genes

The expression levels of key genes in the subgroups were divided according to clinical factors (Supplementary Fig. 4). The expression of two downregulated genes (AVPR1A and GNA14) in the late stage (stages III–IV) of the tumor was lower than that in the early stage (stages I–II), and the other seven upregulated genes showed the opposite trend. Except for AVPR1A, GNA14, and WFS1, all other key genes showed high expression levels in follicular thyroid papillary carcinoma. Furthermore, the expression of two downregulated genes in the non-metastatic group (N0) was higher than that in the lymph node metastasis group (N1), whereas the

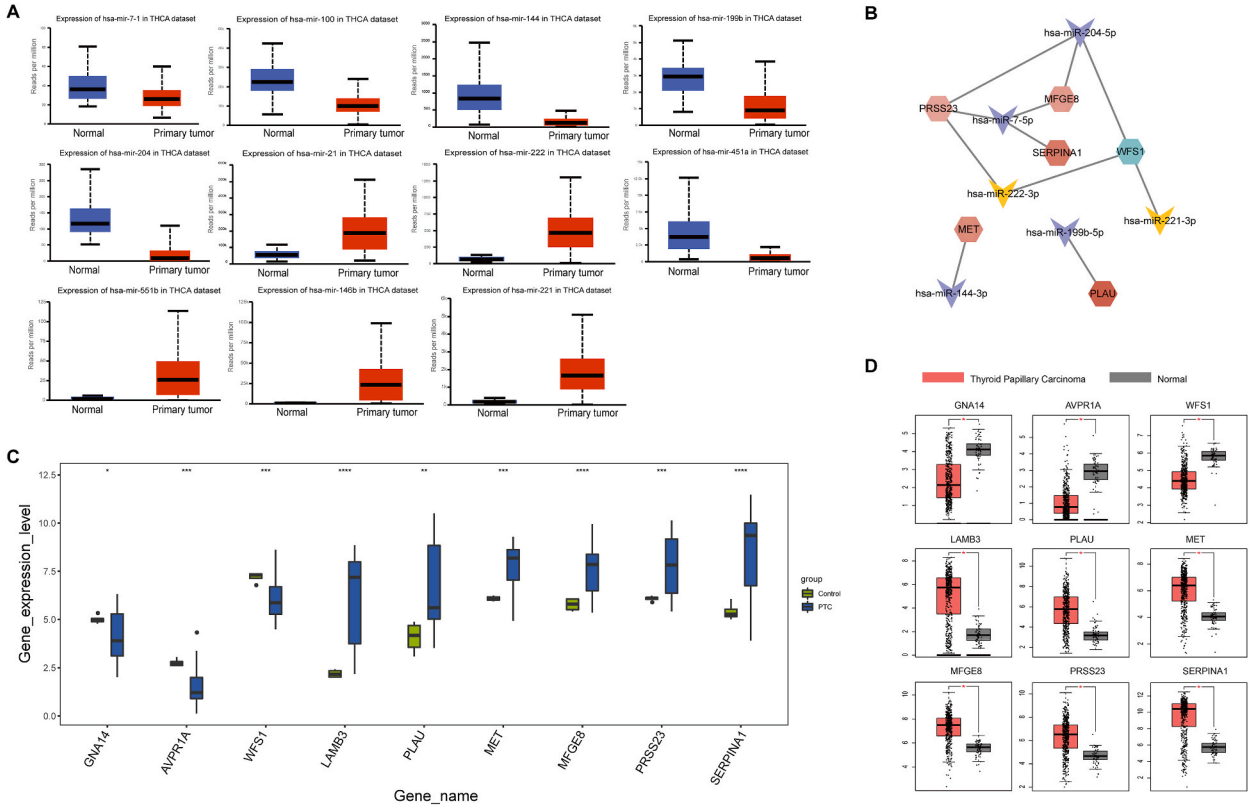


Fig. 7. miRNA-target analysis. A, verification of the expression of miRNAs in the TCGA data using UALCAN; B, miRNA-target interaction network. C, verification of the expression of key genes in the GSE64912; D, verification of the expression of key genes in the TCGA data using GEPIA. Red hexagons indicate upregulated genes, and green hexagons indicate downregulated genes. The yellow arrow indicates upregulated miRNA, and the purple arrow indicates downregulated miRNA.

expression of six upregulated genes showed the opposite trend. These results imply that the expression of these nine key genes is correlated with clinical factors. Therefore, we further investigated the prognostic values of these genes. The results indicated that patients in the AVPR1A and GNA14 high expression groups had better DFS than those in the low expression group (Fig. 8A–B). To identify independent prognostic factors, clinical factors and key genes were analyzed using univariate Cox regression analysis (Fig. 8C). The results indicated that pathologicT, pathologicN, and GNA14 levels were significantly correlated with DFS ($p < 0.05$). Further multivariate Cox regression analysis revealed that only the pathological T stage was an independent prognostic factor (Fig. 8D) ($p < 0.05$).

4. Discussion

As the most common type of thyroid cancer, PTC has a better prognosis than other malignant tumors, but there are still about 30% of patients with PTC who will relapse after treatment [39]. This study aimed to provide new insights into the pathogenesis and treatment of PTC by studying the roles of key genes in PTC.

By integrating the data from multiple datasets and eliminating batch effects, 478 DEGs were identified. The 478 DEGs were mainly involved in signal transduction, cell adhesion, and some cancer pathways. A previous study has shown that common mutations in PTC (BRAF mutations and RET/PTC1 rearrangements) have different sensitivities to MEK1/2, suggesting different signal transduction pathways [40]. Many studies have shown that in malignant tumors, tumor progression is related to changes in cell adhesion, which can affect the biological characteristics of the tumors [41]. This result indicated that the screened DEGs may participate in the occurrence and development of PTC.

Ten modules were screened for genes with a large degree of variation using WGCNA. Further analysis found that the green module had a significant negative correlation with PTC (r = −0.82, p = 4e-42) and a significant positive correlation with control (r = 0.82, p = 4e-42). The brown module showed opposite results to those of the green module. By integrating the brown and green modules with the PPI network modules, we identified nine key genes. GNA14 encodes the G protein subunit alpha 14, and its mutation can upregulate the MAPK pathway to make cell growth factors independent [42]. A previous study reported that a high expression level of GNA14 is positively related to a better prognosis of thyroid cancer S [43], which is consistent with our findings. However, in another study,

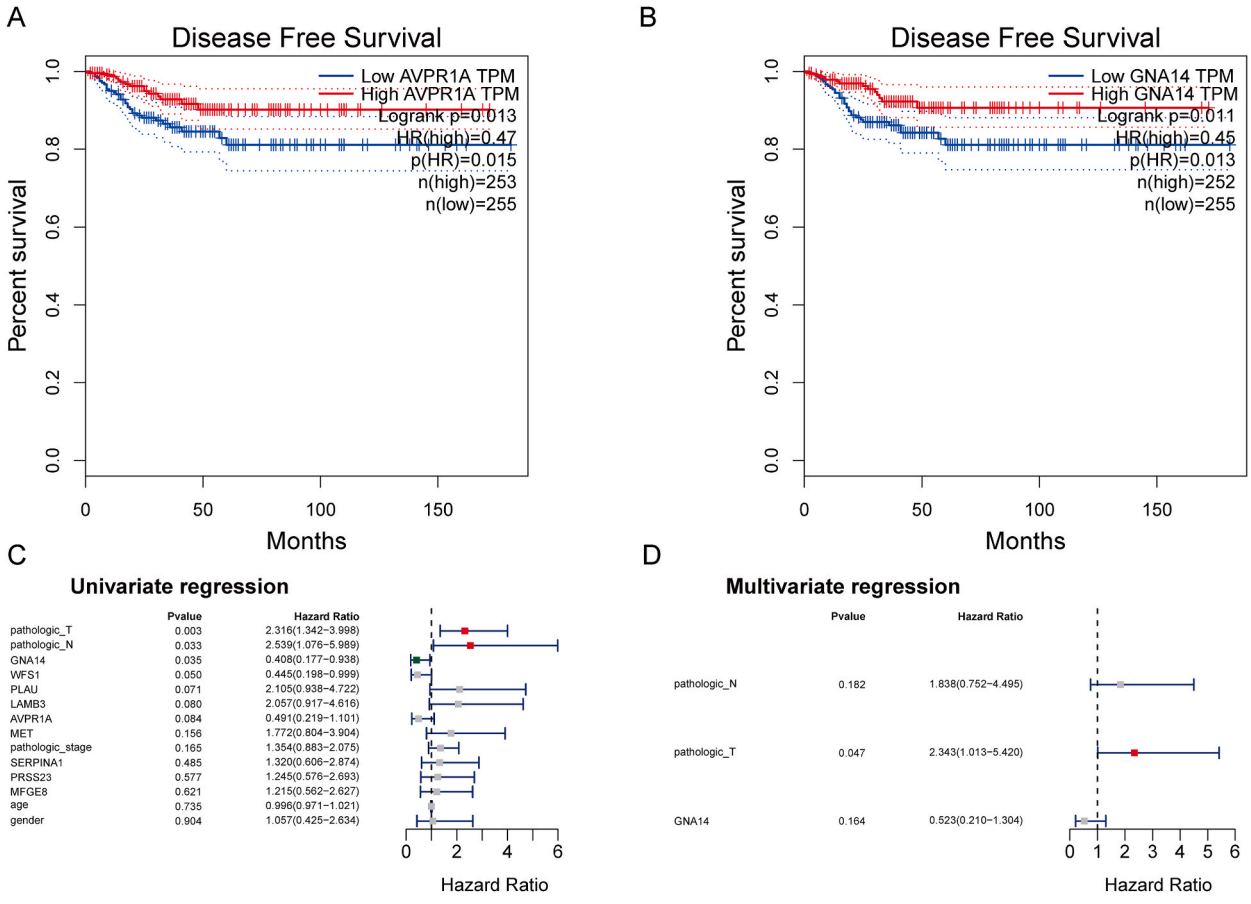


Fig. 8. Prognostic analysis. A, KM curves of disease free survival (DFS) for the expression of AVPR1A. B, KM curves of DFS for the expression of GNA14. C, univariate Cox proportional hazards analysis. D, multivariate Cox proportional hazard analysis.

GNA14 was found to stimulate the expression of KLF7 in endometrial cancer (UCEC) cells, while Krüppel-like factor 7 (KLF7) promoted the proliferation and migration of UCEC cells and inhibited apoptosis [44]. Therefore, the specific mechanism of GNA14 action requires further experimental validation. Arginine vasopressin receptor 1A (AVPR1A) is a receptor for arginine vasopressin. Arginine vasopressin is a hormone that regulates water balance in mammals and is associated with social behavior [45]. This study found that patients with PTC and high AVPR1A expression had better DFS than those with low expression group. Further experiments are needed to confirm the role of AVPR1A in PTC. Wolframin ER transmembrane glycoprotein (WFS1) is associated with diabetes, deafness, optic nerve atrophy, and mental retardation [46]. Laminin subunit beta 3 (LAMB3) is a risk factor of PTC. LAMB3 upregulation can increase the proliferation and metastasis of PTC [47]. It has also been reported that metastatic tumor behavior in PTC could be mediated by LAMB3 based on c-MET/Akt signals [48]. In this study, the expression level of LAMB3 positively correlated with the infiltration levels of macrophages, neutrophils, B cells, $CD4^+$ T cells, and dendritic cells. Plasminogen activator urokinase (PLAU) is a type of protease. PLAU activates the MAPK and Jak-STAT signaling pathways. This activation is achieved indirectly by direct binding of PLAU to uPAR or by activating plasmin [49]. Researchers have demonstrated that PLAU plays an important role in the early process of tumor cell dissemination, which is the initial deliberation of cancer cells from primary sites [49]. Wang et al. showed that the inhibition of PLAU expression suppressed the migratory and invasive abilities of cervical cancer cells by downregulating MMP2 expression [50]. Studies have reported that PLAU can be used as a biomarker for PTC therapy [51]. MET encodes the product of the proto-oncogene MET and is a member of the receptor tyrosine kinase family of proteins. Fusion of MET and multiple genes (RET, NTRK3, ALK, NTRK1, and BRAF) has been found in children with PTC [52]. Milk fat globule EGF and factor V/VIII domain-containing (MFGE8) oncogenes are potential targets for ovarian cancer treatment [53]. Xue et al. reported that MFGE8 was overexpressed in PTC tissues [54], which is consistent with the results of our study. Serine protease 23 (PRSS23) is upregulated in TPC [55]. Serpin family member 1 (SERPINA1) is a potential diagnostic marker of PTC [56]. We also screened for miRNAs that could regulate key genes. These genes are widely reported to be associated with a variety of cancers [57,58].

Prognostic analysis of the key genes showed that only AVPR1A and GNA14 were significantly associated with DFS, and none of the nine key genes were found to be independent prognostic factors. These results suggest that these nine genes are only suitable diagnostic factors rather than prognostic factors. Our study has several advantages. First, we integrated multiple datasets, and the results from big data were more reliable. Second. we eliminated batch effects from the consolidated data. Finally, key genes were verified using both the proteome and transcriptome. This study had some limitations. Because our results were predicted using datasets from public databases, more *in vitro* and *in vivo* experiments are needed in the future.

In conclusion, nine key genes and related miRNAs were screened, and multiple analyses indicated that they may be therapeutic targets for PTC. This study provides novel insights into the mechanisms underlying PTC development.

Funding

This research was supported by the National Natural Science Foundation of China (No. 82072184).

CRediT authorship contribution statement

Lingfeng Pan: Writing – original draft, Formal analysis, Data curation, Conceptualization. **Lianbo Zhang:** Writing – review & editing, Methodology, Formal analysis. **Jingyao Fu:** Resources, Investigation. **Keyu Shen:** Writing – review & editing, Resources. **Guang Zhang:** Writing – review & editing, Project administration, Formal analysis.

Declaration of competing interest

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e27928.

References

[1] H. Sung, J. Ferlay, R. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J Clin 71 (2021) 209–249.
[2] C.M. Kitahara, J.A. Sosa, The changing incidence of thyroid cancer, Nat. Rev. Endocrinol. 12 (2016) 646–653.
[3] J. Yu, Y. Deng, T. Liu, J. Zhou, X. Jia, T. Xiao, S. Zhou, J. Li, Y. Guo, Y. Wang, Lymph node metastasis prediction of papillary thyroid carcinoma based on transfer learning radiomics, Nat. Commun. 11 (2020) 1–10.
[4] M. Luster, C. Aktolun, I. Amendoeira, M. Barczyński, K.C. Bible, L.H. Duntas, R. Elisei, D. Handkiewicz-Junak, M. Hoffmann, B. Jarząb, L. Leenhardt, T. J. Musholt, K. Newbold, I.J. Nixon, J. Smit, M. Sobrinho-Simões, J.A. Sosa, R.M. Tuttle, F.A. Verburg, L. Wartofsky, D. Führer, European perspective on 2015 American thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: proceedings of an interactive international symposium, Thyroid 29 (2019) 7–26.
[5] X. Luo, A. Wu, Analysis of risk factors for postoperative recurrence of thyroid cancer, J buon 24 (2019) 813–818.

[6] A. Machens, K. Lorenz, F. Weber, H. Dralle, Risk patterns of distant metastases in follicular, papillary and medullary thyroid cancer, Horm. Metab. Res. 54 (2022) 7–11.
[7] M. Xing, B.R. Haugen, M. Schlumberger, Progress in molecular-based management of differentiated thyroid cancer, Lancet 381 (2013) 1058–1069.
[8] R. Niciporuka, J. Nazarovs, A. Ozolins, Z. Narbuts, E. Miklasevics, J. Gardovskis, Can We Predict Differentiated Thyroid Cancer Behavior? Role of Genetic and Molecular Markers, vol. 57, Medicina (Kaunas), 2021.
[9] M. Kasko, M. Grigerova, A. Alemayehu, K. Zavodna, V. Kasko, J. Podoba, Fine-needle biopsy of thyroid nodules and the contribution of molecular analysis of BRAF and RAS mutations, Bratisl. Lek. Listy 124 (2023) 869–872.
[10] R. Fan, L. Dong, P. Li, X. Wang, X. Chen, Integrated bioinformatics analysis and screening of hub genes in papillary thyroid carcinoma, PLoS One 16 (2021) e0251962.
[11] Y. Wan, X. Zhang, H. Leng, W. Yin, W. Zeng, C. Zhang, Identifying hub genes of papillary thyroid carcinoma in the TCGA and GEO database using bioinformatics analysis, PeerJ 8 (2020) e9120.
[12] G. Xue, X. Lin, J.F. Wu, D. Pei, D.M. Wang, J. Zhang, W.J. Zhang, Identification of key genes of papillary thyroid carcinoma by integrated bioinformatics analysis, Biosci. Rep. 40 (2020).
[13] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinf. 9 (2008) 559.
[14] S. Wang, D. Sun, C. Liu, Y. Guo, J. Ma, R.L. Ge, S. Cui, Weighted gene co-expression network analysis reveals the hub genes associated with pulmonary hypertension, Exp. Biol. Med. 248 (2023) 217–231.
[15] X. Wang, A.K. Bajpai, Q. Gu, D.G. Ashbrook, A. Starlard-Davenport, L. Lu, Weighted gene co-expression network analysis identifies key hub genes and pathways in acute myeloid leukemia, Front. Genet. 14 (2023) 1009462.
[16] Z. Wang, X. Chen, C. Li, W. Tang, Application of weighted gene co-expression network analysis to identify novel key genes in diabetic nephropathy, J Diabetes Investig 13 (2022) 112–124.
[17] T. Barrett, D.B. Troup, S.E. Wilhite, P. Ledoux, D. Rudnev, C. Evangelista, I.F. Kim, A. Soboleva, M. Tomashevsky, R. Edgar, NCBI GEO: mining tens of millions of expression profiles–database and tools update, Nucleic Acids Res. 35 (2007) D760–D765.
[18] J.T. Leek, W.E. Johnson, H.S. Parker, A.E. Jaffe, J.D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments, Bioinformatics 28 (2012) 882–883.
[19] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, G.K. Smyth, Limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res. 43 (2015) e47.
[20] W. Huang da, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat. Protoc. 4 (2009) 44–57.
[21] 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 (2000) 25–29.
[22] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (2000) 27–30.
[23] C. Fan, C. Zhao, F. Wang, S. Li, J. Wang, Significance of PTEN mutation in cellular process, prognosis, and drug selection in clear cell renal cell carcinoma, Front. Oncol. 9 (2019).
[24] D. Szklarczyk, J.H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N.T. Doncheva, A. Roth, P. Bork, L.J. Jensen, C. von Mering, The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible, Nucleic Acids Res. 45 (2017) D362–d368.
[25] 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 (2003) 2498–2504.
[26] Y. Tang, M. Li, J. Wang, Y. Pan, F.X. Wu, CytoNCA: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks, Biosystems 127 (2015) 67–72.
[27] G.D. Bader, C.W. Hogue, An automated method for finding molecular complexes in large protein interaction networks, BMC Bioinf. 4 (2003) 2.
[28] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS 16 (2012) 284–287.
[29] Z. Yang, Q. Zi, K. Xu, C. Wang, Q. Chi, Development of a macrophages-related 4-gene signature and nomogram for the overall survival prediction of hepatocellular carcinoma based on WGCNA and LASSO algorithm, Int. Immunopharm. 90 (2021) 107238.
[30] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 15545–15550.
[31] T. Li, J. Fan, B. Wang, N. Traugh, Q. Chen, J.S. Liu, B. Li, X.S. Liu, TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells, Cancer Res. 77 (2017) e108–e110.
[32] F. Pontén, J.M. Schwenk, A. Asplund, P.H. Edqvist, The Human Protein Atlas as a proteomic resource for biomarker discovery, J. Intern. Med. 270 (2011) 428–446.
[33] A. Koch, T. De Meyer, J. Jeschke, W. Van Criekinge, MEXPRESS: visualizing expression, DNA methylation and clinical TCGA data, BMC Genom. 16 (2015) 636.
[34] H. Dweep, N. Gretz, miRWalk2.0: a comprehensive atlas of microRNA-target interactions, Nat. Methods 12 (2015) 697.
[35] Z. Tang, C. Li, B. Kang, G. Gao, C. Li, Z. Zhang, GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses, Nucleic Acids Res. 45 (2017) W98–w102.
[36] Z. Tang, B. Kang, C. Li, T. Chen, Z. Zhang, GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis, Nucleic Acids Res. 47 (2019) W556–w560.
[37] D.S. Chandrashekar, B. Bashel, S.A.H. Balasubramanya, C.J. Creighton, I. Ponce-Rodriguez, B. Chakravarthi, S. Varambally, UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses, Neoplasia 19 (2017) 649–658.
[38] M. Milella, I. Falcone, F. Conciatori, U. Cesta Incani, A. Del Curatolo, N. Inzerilli, C.M. Nuzzo, V. Vaccaro, S. Vari, F. Cognetti, L. Ciuffreda, PTEN: multiple functions in human malignant tumors, Front. Oncol. 5 (2015) 24.
[39] R. Qin, C. Li, X. Wang, Z. Zhong, C. Sun, Identification and validation of an immune-related prognostic signature and key gene in papillary thyroid carcinoma, Cancer Cell Int. 21 (2021) 378.
[40] Y.C. Henderson, R. Toro-Serra, Y. Chen, J. Ryu, M.J. Frederick, G. Zhou, G.E. Gallick, S.Y. Lai, G.L. Clayman, Src inhibitors in suppression of papillary thyroid carcinoma growth, Head Neck 36 (2014) 375–384.
[41] B. Górka, J. Skubis-Zegadło, M. Mikula, K. Bardadin, E. Paliczka, B. Czarnocka, NrCAM, a neuronal system cell-adhesion molecule, is induced in papillary thyroid carcinomas, Br. J. Cancer 97 (2007) 531–538.
[42] Y.H. Lim, A. Bacchiocchi, J. Qiu, R. Straub, A. Bruckner, L. Bercovitch, D. Narayan, J. McNiff, C. Ko, L. Robinson-Bostom, R. Antaya, R. Halaban, K.A. Choate, GNA14 somatic mutation causes congenital and sporadic vascular tumors by MAPK activation, Am. J. Hum. Genet. 99 (2016) 443–450.
[43] Y. Shen, S. Dong, J. Liu, L. Zhang, J. Zhang, H. Zhou, W. Dong, Identification of Potential Biomarkers for Thyroid Cancer Using Bioinformatics Strategy: a Study Based on GEO Datasets, BioMed research international, 2020, 2020.
[44] J. Wang, F. Teng, H. Chai, C. Zhang, X. Liang, Y. Yang, GNA14 stimulation of KLF7 promotes malignant growth of endometrial cancer through upregulation of HAS2, BMC Cancer 21 (2021) 456.
[45] O. Vollebregt, E. Koyama, C.C. Zai, S.A. Shaikh, A.J. Lisoway, J.L. Kennedy, J.H. Beitchman, Evidence for association of vasopressin receptor 1A promoter region repeat with childhood onset aggression, J. Psychiatr. Res. 140 (2021) 522–528.
[46] I.N. Bespalova, G. Van Camp, S. J.H. Bom, D.J. Brown, K. Cryns, A.T. DeWan, A.E. Erson, K. Flothmann, H.P.M. Kunst, P. Kurnool, T.A. Sivakumaran, C.W.R. J. Cremers, S.M. Leal, M. Burmeister, M.M. Lesperance, Mutations in the Wolfram syndrome 1 gene (WFS1) are a common cause of low frequency sensorineural hearing loss, Hum. Mol. Genet. 10 (2001) 2501–2508.
[47] Y. Wang, Y. Jin, A. Bhandari, Z. Yao, F. Yang, Y. Pan, Z. Zheng, S. Lv, O. Wang, Upregulated LAMB3 increases proliferation and metastasis in thyroid cancer, OncoTargets Ther. 11 (2018) 37.

[48] 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 (2018) 1–10.
[49] C. Ai, J. Zhang, S. Lian, J. Ma, B. Győrffy, Z. Qian, Y. Han, Q. Feng, FOXM1 functions collaboratively with PLAU to promote gastric cancer progression, J. Cancer 11 (2020) 788–794.
[50] X. Wang, Z. Jiang, J. An, X. Mao, F. Lin, P. Sun, Effect of a synthetic inhibitor of urokinase plasminogen activator on the migration and invasion of human cervical cancer cells in vitro, Mol. Med. Rep. 17 (2018) 4273–4280.
[51] J. Qiu, W. Zhang, Q. Xia, F. Liu, L. Li, S. Zhao, X. Gao, C. Zang, R. Ge, Y. Sun, RNA sequencing identifies crucial genes in papillary thyroid carcinoma (PTC) progression, Exp. Mol. Pathol. 100 (2016) 151–159.
[52] B. Pekova, V. Sykorova, S. Dvorakova, E. Vaclavikova, J. Moravcova, R. Katra, J. Astl, P. Vlcek, D. Kodetova, J. Vcelak, B. Bendlova, RET, NTRK, ALK, BRAF, and MET fusions in a large cohort of pediatric papillary thyroid carcinomas, Thyroid 30 (2020) 1771–1780.
[53] L. Tibaldi, S. Leyman, A. Nicolas, S. Notebaert, M. Dewulf, T.H. Ngo, C. Zuany-Amorim, N. Amzallag, I. Bernard-Pierrot, X. Sastre-Garau, C. Théry, New blocking antibodies impede adhesion, migration and survival of ovarian cancer cells, highlighting MFGE8 as a potential therapeutic target of human ovarian carcinoma, PLoS One 8 (2013) e72708.
[54] G. Xue, X. Lin, J.-F. Wu, D. Pei, D.-M. Wang, J. Zhang, W.-J. Zhang, Identification of key genes of papillary thyroid carcinoma by integrated bioinformatics analysis, Biosci. Rep. 40 (2020).
[55] K.-W. Chung, S.W. Kim, S.W. Kim, Gene expression profiling of papillary thyroid carcinomas in Korean patients by oligonucleotide microarrays, J. Korean Surg. Soc. 82 (2012) 271–280.
[56] K. Vierlinger, M.H. Mansfeld, O. Koperek, C. Nöhammer, K. Kaserer, F. Leisch, Identification of SERPINA1 as single marker for papillary thyroid carcinoma through microarray meta analysis and quantification of its discriminatory power in independent validation, BMC Med. Genom. 4 (2011) 1–9.
[57] K. Jiang, G. Li, W. Chen, L. Song, T. Wei, Z. Li, R. Gong, J. Lei, H. Shi, J. Zhu, Plasma exosomal miR-146b-5p and miR-222-3p are potential biomarkers for lymph node metastasis in papillary thyroid carcinomas, OncoTargets Ther. 13 (2020) 1311.
[58] L. Liu, J. Wang, X. Li, J. Ma, C. Shi, H. Zhu, Q. Xi, J. Zhang, X. Zhao, M. Gu, MiR-204-5p suppresses cell proliferation by inhibiting IGFBP5 in papillary thyroid carcinoma, Biochem. Biophys. Res. Commun. 457 (2015) 621–626.