

RESEARCH ARTICLE

Open Access



Identification of potential novel biomarkers to differentiate malignant thyroid nodules with cytological indeterminate

Dandan Wu¹, Shudong Hu², Yongzhong Hou¹, Yingying He^{1*} and Shubai Liu^{1*}

Abstract

Background: The fine-needle aspiration (FNA) biopsy was broadly applied to clinical diagnostics evaluation for thyroid carcinomas nodule, while companioning with higher uncertainty rate (15~30%) to identify malignancy for cytological indeterminate cases. It is requirement to discover novel molecular biomarkers to differentiate malignant thyroid nodule more precise.

Methods: We employed weighted gene co-expression network analysis (WGCNA) to discover genes significantly associated with malignant histopathology for cytological indeterminate nodules. In addition, identified significantly genes were validated through another independently investigations of thyroid carcinomas patient's samples via cBioportal and Geipa. The key function pathways of significant genes involving were blast through GenClip.

Results: Twenty-four signature genes were identified significantly related to thyroid nodules malignancy. Furthermore, five novel genes with missense mutation, *FN1* (R534P), *PROS1*((K200I), (Q571K)), *SCEL* (T320S), *SLC34A2*(T688M) and *TENM1* (S1131F), were highlighted as potential biomarkers to rule out nodules malignancy. It was identified that the key functional pathways involving in thyroid carcinomas.

Conclusion: These results will be helpful to better understand the mechanism of thyroid nodules malignant transformation and characterize the potentially biomarkers for thyroid carcinomas early diagnostics.

Keywords: Papillary thyroid carcinoma, Biomarker, Thyroid nodules, Biomarker, Fine-needle aspiration biopsy, WGCNA

Background

Thyroid cancer is a common malignant neoplasm in worldwide. Recently, the incidence rate of thyroid cancer is rapid raising in the world and becoming the potential threat for public health [1, 2]. It is important to develop early precise diagnostics method and interfere the thyroid neoplasm progress into malignant carcinoma. Up to now, the Fine-needle aspiration (FNA) biopsy is the most accurate and cost-effective tool for thyroid nodules

clinical evaluating. It has been strongly recommended by the American Thyroid Association as standardized clinical operation [3–5]. However, about 10~30% cases' cytological results are indeterminate, and being labelled as indeterminate or suspicious for malignancy. Among these cytological indeterminate cases, majority of patients underwent partial or complete thyroidectomy and checked by histological evaluation. Although the subsequent postsurgical evaluation results reveal only 6~30% cytological indeterminate cases identified as malignant, it made this clinical operation extremely low efficiency and non-specificity while with higher costs [6, 7].

* Correspondence: heyj@ujs.edu.cn; shubliu@gmail.com

¹Institute of Life Sciences, Jiangsu University, 301 Xuefu Road, JinKou District, Zhenjiang 212013, PR China

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Molecular biomarkers analysis is a powerful adjunct approach to traditional carcinomas pathological evaluation. Multiple molecular markers have been discovered and employed in developing precise diagnostics methods and novel strategies to properly treatment. These biomarkers are generated from gene sequence for gene mutations, gene rearrangements, RNA-based assays, gene expression profiling and immune-histochemistry [8, 9]. As endocrine neoplasm deriving from follicular or parafollicular thyroid cells, thyroid cancer has been reported associated with higher frequency (about 70%) somatic alternation or deletion of genes involving key signaling pathways, such as the mutation of *BRAF* and *RAS* [6, 10], *NTRK1* tyrosine kinases and key effectors mitogen-activated protein kinase (MAPK) signaling pathway [11]. With advanced understanding of thyroid tumor formation, the researches of generic mutation-based biomarkers discovery shifted from single mutation to molecular signatures genes or panels of multiple mutations [12]. According to the previously American Thyroid Association Management Guidelines, a 7-gene molecular biomarkers panel of genetic mutation and rearrangement (7-gene MT), including BRAFV600E, three isoforms of RAS point mutations and translocations of *PAX8/PPARc* and *RET/PTC* genes [6, 13], was recommended to evaluate the residual FNA sample with cytological indeterminate and estimate with high specificity (~ about 90%) [14–16]. Especially, the mutational testing of biomarker genes has been proposed to be a rule-in test with reported higher specificity in clinical practice. Recently, it was reported that the sensitivity of seven-genes mutational panel testing showed huge variation, from 44 to 100% [6, 17]. It is strong suggested that traditional gene mutation panels analysis may not reliably rule out nodules malignancy in some case population. In current, there is no definitively single optimal molecular test that 100% promised to rule-in or rule-out the malignancy in cytology-indeterminate cases [18]. It is necessary to discover novel potential molecular biomarkers to enhance sensitivity and specificity of mutational analysis and precise to rule-in the malignance for cytology indeterminate nodules.

Recently, lacking of long term clinical outcome tracking recording of using molecular markers, there are some controversies over the benefit and limitation of existing molecular markers testing [18]. To enhance the efficiency of thyroid carcinomas patient's diagnostic, treatment and health management, it is the trend to develop systemic diagnostic strategy and discover novel applicable and specific molecular biomarkers for early diagnostics through analyzing the genetic and expression profiling of thyroid nodules from FNA biopsy [19]. Among multiple computerization methodologies, the Weighted Gene Co-Expression Network Analysis

(WGCNA) is considered as one of the most useful approaches to discover gene co-expression network based functional feature through gene expression profiling analysis [20]. Recently, WGCNA is be widely applied to screening the signature genes significantly associated with clinical feature. It is powerful to discover candidate biomarkers for cancer early diagnostics, cancer-associated pathways or therapeutic targets for precise treatment in hepatocellular carcinoma [21], lung cancer [22], endometrial cancer [23] and melanoma [24].

In this study, we employed the WGCNA to analyze gene expression profile of thyroid nodules with cytological indeterminate and aimed to identify the highly connected hub and modules that genes significantly associated with histological malignant thyroid nodule. In addition, we will explore other independently clinical cases through the genetic database to verify the significantly signature genes with genetic changes and discover the key biological pathways significantly associated with malignant thyroid cancer by Genclip pathfinder.

Methods

Gene expression data source

In this study, the dataset applied for data analysis is available in the Gene Expression Omnibus (GEO) repository (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE34289>) in NCBI, and the platform entry number is GPL14961. This dataset came from the work of Erik et al. (2012) and contained 172 target genes expression data [25]. The samples related information and genes annotated probe-id were transformed into gene symbols and related functional annotations. The related clinical trait annotation was distracted from GSE34289 annotation dataset. Each gene expression value was normalized and performed with log2 transformation. The genes expression profiling of 364 thyroid nodules were split into four groups and 265 samples with cytology indeterminate were selected for further WGCNA analysis as workflow demonstration (Fig. 1).

Clinical case samples group sorting

According to the clinical cytological / histopathological traits, 364 thyroid nodules samples were split into four groups: Group one (180), cytology-Indeterminate/histopathology-benign; Group two (85), cytology-indeterminate/histopathology-malignant; Group three (44), cytology-benign/histopathology-benign; Group four (55), cytology-malignant /histopathology-benign (Additional file Table S1).

Construction of weighted gene co-expression network

The WGCNA package of R (version 1.63) was download and setup by following the protocol described previously [26]. The WGCNA package was used to perform various

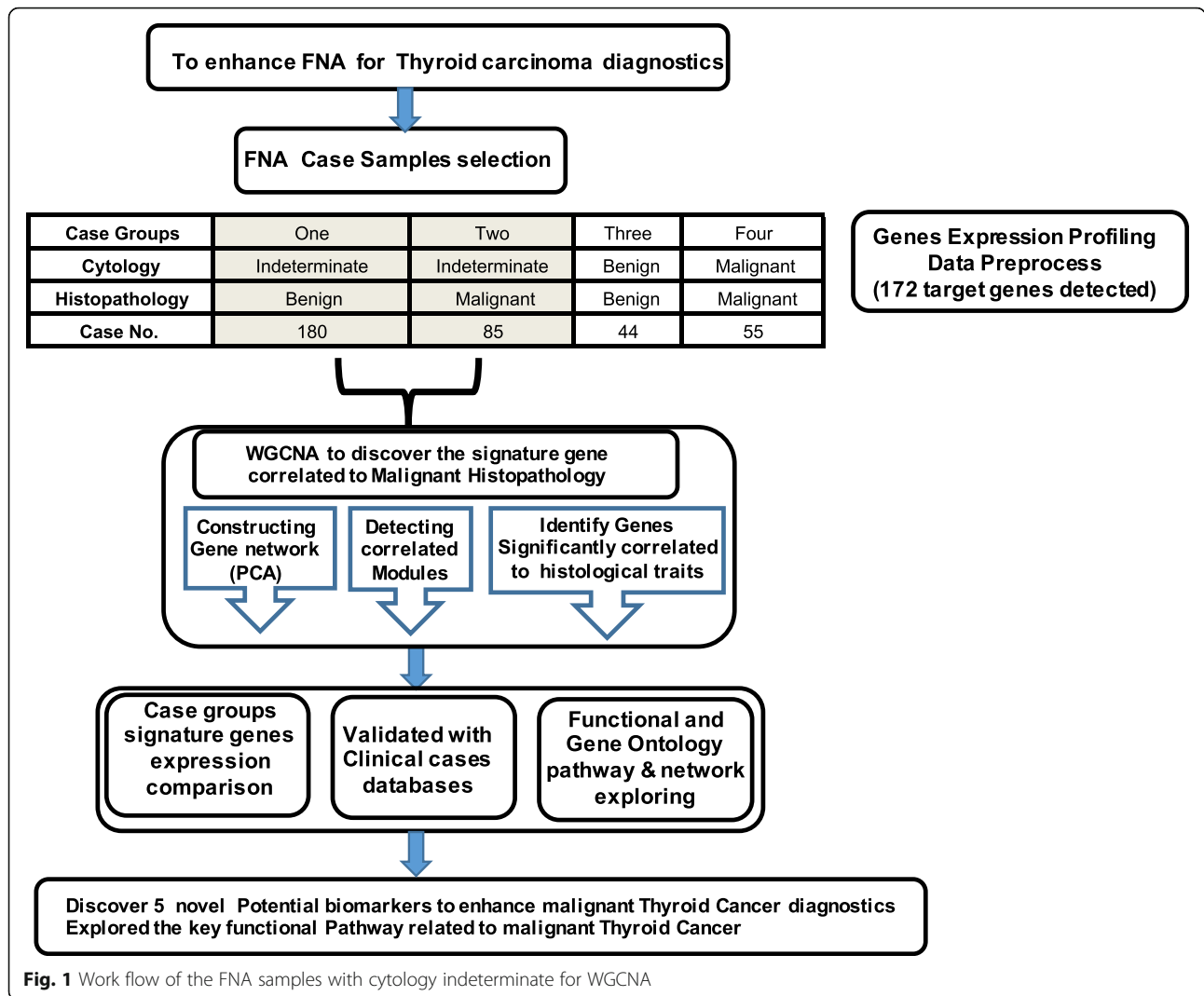


Fig. 1 Work flow of the FNA samples with cytology indeterminate for WGCNA

functions in weighted correlation network analysis, including constructing network, detecting module, calculating topological properties, simulating data, visualization, and interfacing with external software [26]. First of all, we have checked data to exclude the sample with excessive missing values and identify outlier microarray samples. After data preprocessing, we applied the principal component analysis (PCA) to double check the data quality. We observed that tumor and normal samples were separated in the PCA plot (Additional file Figure S1), and then we performed hierarchical clustering on the samples to further detect potential outliers. The total 265 samples were used for next step analysis (Fig. 1). We chose the soft threshold $\beta = 7$ to construct the co-expression network as the R^2 reached the peak for the first time when $\beta = 7$. The plot of $\log_{10}(p(k))$ versus $\log_{10}(k)$ (Additional file Figure S2) indicates that the network is close to a scale-free network by using $\beta = 7$, where k is the whole network connectivity

and $p(k)$ is the corresponding frequency distribution (Additional file Table S2). When $\beta = 7$, the R^2 is 0.98, ensuring that the network was close to the scale-free network. After the soft thresholding power β was determined, the Topological Overlap Matrix (TOM) and $\text{dissTOM} = 1 - \text{TOM}$ were obtained (Additional file Figure S3). After the modules were identified, the T-test was used to calculate the significant p -value of candidate genes, and the gene significance (GS) was defined as mediated p -value of each gene ($\text{GS} = \lg P$). Then, the module significance (MS) were defined as the average GS of all the genes involved in the module. In general, the module with the highest MS among all the selected modules will be considered as the one associated with disease. In addition, we also calculated the relevance between the clinical feature (histopathology) of modules and phenotypes to identify the most relevant module. The hierarchical clustering analysis was used to identify gene modules and color to indicate modules,

which is a cluster of densely interconnected genes in terms of co-expression (Additional file Figure S4). For genes that are not assigned to any of the modules, WGCNA places them into a grey module as not co-expressed (Additional file Table S3). The module eigengene (ME) of a module is defined as the first principal component of the module and represents the overall expression level of the module. To identify modules that were significantly associated with the traits of histology, age and gender status, we correlated the MEs (i.e. the first principle component of a module) [27] with clinical traits and searched the most significant associations. A hierarchical clustering of MEs was performed to study the correlations among the modules. we used the linear mixed-effects model (eq. (4)) for testing the association of a module to the histology determinate tumor status [26].

Exploring the clinical cancer cases databases

Genes significance associated with histology feature of malignant thyroid nodule were blast in the cBioportal Cancer Genomics dataset with independent cases and verified the association of thyroid carcinoma's patient's cases to public [28, 29]. In addition, we blast Gene Expression Profiling Interactive Analysis databases (<http://gepia.cancer-pku.cn>) to validate the expression of these five biomarker candidates.

Validations of signature genes expression

The validations of significant genes were performed by comparison of expression level among the thyroid nodule case groups and blast in TCGA (The Cancer Genome Atlas) database with independent cases. The case group with cytology indeterminate/histopathology benign was used as the benchmark. The individual gene expression in each group are presented as means \pm standard error of the mean (SEM) that represent distribution of group cases. The expression level comparison was used the fold change ratio to quantitatively analyze. The Significance of differences for the values were determined using the student t-test with the Prism software (GraphPad Software, Inc. San Diego, CA). A P value < 0.001 was setup as significant difference standard.

GO and pathway enrichment analysis

We utilized GenCLiP 2.0 tool to collect the correlated Gene Ontology (GO) functional clustering and pathway enrichment analyses for the genes significance in blue module, which is powerful to discovery the abnormal pathway or key components related to certain diseases [30]. The P value < 0.05 was setup as the significantly cut-off criterion.

Results

Identified gene modules correlation with histological traits

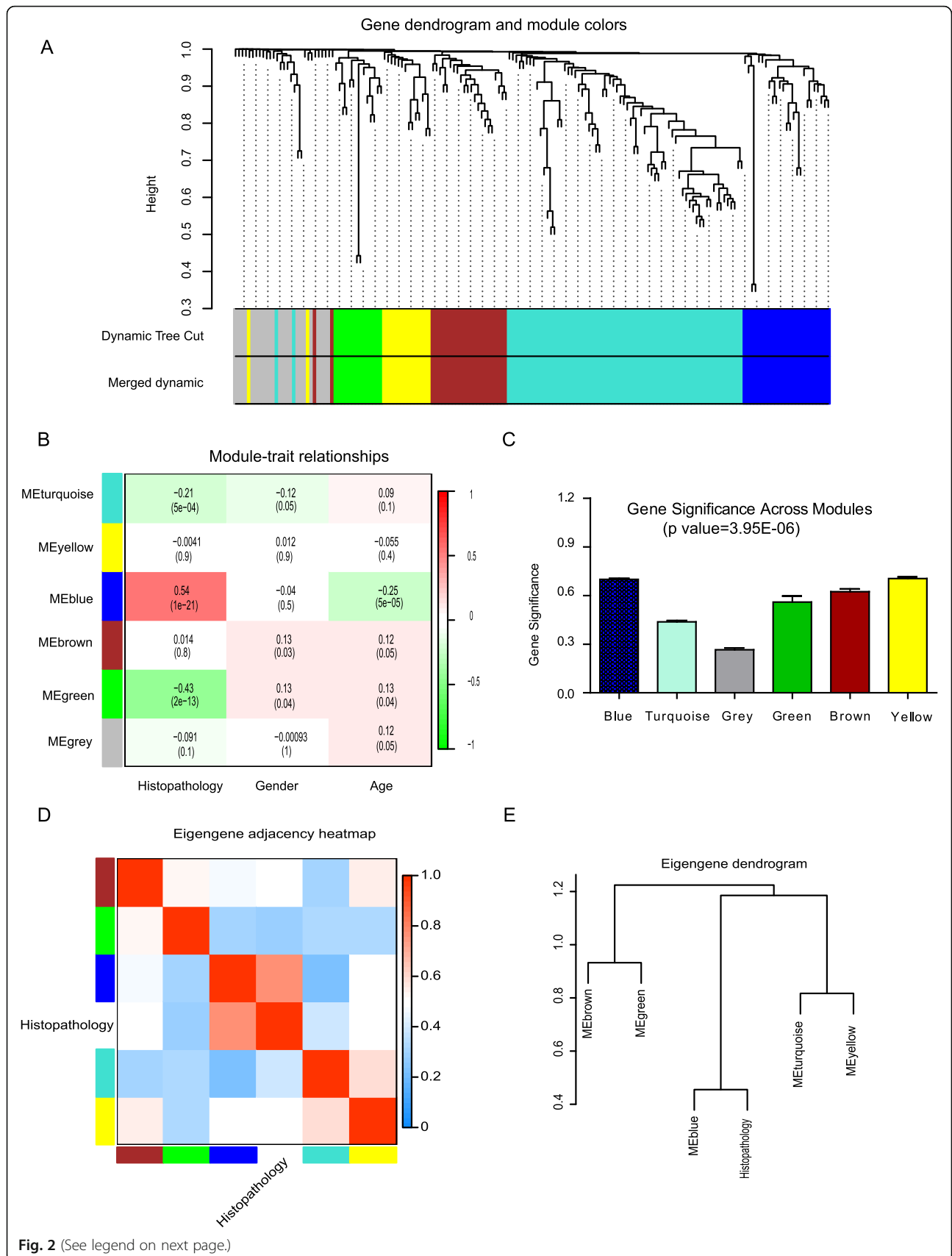
In this study, we applied WGCNA to investigate the relationship between gene expression profiling of FNA thyroid nodules with cytology indeterminate (265 cases, group one and group two, Additional file Table S1) and clinical traits-histopathology, age and gender. After using a dynamic tree cutting algorithm, we identified 6 distinct co-expression modules (Fig. 2a), including Blue (24), Turquoise (66), Green (14), Brown (23), Yellow (15) and Grey (29) modules containing with varied different number genes. There are three MEs, Blue, Green and Turquoise, highly significantly correlated to histopathology trait based on the hierarchical clustering analysis (Fig. 2b), and Blue is positive correlated with histopathology trait. Through calculation of the linear mixed-effects model, the turquoise module (t-value = -0.21 , P value = 0.004), blue module (t-value = 0.54 , P value = $1e-21$) and the green module (t-value = -0.43 , P value = $2e-13$) are identified significantly associated with malignant thyroid nodule status (Fig. 2c). The blue module, containing 24 genes (Additional file Table S4), is the most significant module (P value = $1e-21$) associated with thyroid nodule malignant histopathology feature, while green and turquoise module are negative correlated with malignant feature and do not discuss. The 29 uncorrelated genes were assigned into a grey module, which was ignored in the following analysis (Fig. 2b, and Additional file Table S3).

Enriched genes significance related to histological feature

Compared the MS among the modules (Fig. 2c), the results showed that the Blue module is the highest relevance and positive correlated to histopathology malignant status (cor = 0.77 , P value = $6.8e-06$). For each gene contained in a module, we plotted the scatter figure of multiple module memberships (MM) against the GS (Additional file Figure S5A-E). In the WGCNA, the module membership (MM): $MM(i) = \text{cor}(xi, ME)$ is defined to measure the importance of the gene within the module. The greater absolute value of $MM(i)$, gene i is more important the in the module. The GS in the blue module is highly correlated with MM, indicating that Gene is significantly associated with malignant histological feature (Fig. 2d, P value = $6.8e-06$). The genes significance is also the important element of the Blue module (P value = $3.95E-06$, Fig. 2e) and listed (Additional file Table S4). The heatmap plot is depicted of topological overlap in the gene network (Additional file Figure S6).

Validated significant genes through cBioPortal database

Compared with cytology-indeterminate/histopathology-benign group as the benchmark, the 23 signature genes



(See figure on previous page.)

Fig. 2 Gene dendrogram and module cluster for Histopathological feature. **a** Clustering dendrogram of genes, with dissimilarity based on topological overlap, is merged with assigned module colors and the original module colors. **b** The correlation of Module-clinical traits. Each row corresponds to a module; each column corresponds to a clinical trait feature. Each cell contains the test statistic value and its corresponding p value from the linear mixed-effects model. Network of eigengene represents the relationships among the modules and the histological traits. **c** There are total 6 Module memberships vs. gene significance cluster for histopathology trait. Module membership vs gene significance is correlating to thyroid nodule histopathological status. Panel **d** shows a hierarchical clustering dendrogram of the eigengenes in which the dissimilarity of eigengenes (EI, EJ) is given by $1 - \text{cor}(EI, EJ)$. The heatmap in panel **(e)** shows the eigengene adjacency ($AIJ = (1 + \text{cor}(EI, EJ))/2$)

showed significant higher expression (Fold change > 1.0, P value < 0.001) in the cytology-indeterminate/histopathology-malignant group cases, while only PPP2R2B with lower expression (P value = 0.0039, Fig. 3a). In the negative case group, with double cytology/histopathology benign, although 12 genes were lower expression ($0.92 < FC < 0.98$, P value > 0.001), 10 genes were close to equal expression (*CC2D2B*, *CFH*, *CLDN16*, *FBXO2*, *GABRB2*, *KRT19*, *PPP2R2B*, *ST3GAL5*, *PROS1*, *SLC34A2*, $0.98 < FC < 1.01$, P value > 0.05), and *FNI* ($FC = 1.0881$, P value = 0.0086) and *GOS2* ($FC = 1.0881$, P value = 0.0266) indicated higher expression, there are no significantly different expression than control. In the positive case group, with double cytology/histopathology malignant, except *PPP2R2B* with lower ($FC = 0.9227$, P value = 0.0042) and *CC2D2B* with higher expression ($FC = 1.0725$, P value = 0.002), the other 22 genes were significantly higher expression ($FC > 1.0$, P value < 0.0001) than benchmark. Furthermore, the identified 5 potential biomarkers were significantly higher expression ($FC > 1.15$, P value < 0.0001) in both cytology-indeterminate/histopathology-malignant group and double cytology/histopathology malignant group (Fig. 3a&b).

Moreover, we put these 24 genes into cBioPortal Cancer Genomics database for validation and inquired with 915 patients' datasets in 3 independent studies as Papillary Thyroid Carcinoma (TCGA, Cell 2014), Thyroid Carcinoma (TCGA Provisional) and Poorly-Differentiated and Anaplastic Thyroid Cancer (MSKCC, JCI 2016). The exploring results indicated that 15 of 24 genes significance were altered in 37 (4.0%) of 915 queried cases/patients as listed in Oncoprint table (Fig. 4a). The matched genes are listed as *CC2D2B*, *CFH*, *CITED1*, *FNI*, *GOS2*, *GABRB2*, *KRT19*, *TENM1*, *PPP2R2B*, *PROS1*, *RXRG*, *SCEL*, *SERGEF*, *SLC34A2* and *STK32A*. The genetic alternation types included missense mutation, amplification and deep deletion (Fig. 4a). Sixteen genes that associated with detail information of copy-number alterations were identified, including the alternation type, altered samples number and percent of patient's cases (Additional file Table S5). The exploring results contains 86 gene pairs with mutually exclusive alterations (none significant), and 167 gene pairs with co-occurrent alterations (non-significant) and 6 genes pairs with significant alternation (P value < 0.05). The 6 genes pairs are identified as *CFH* & *GOS2*, *CFH* &

RXRG, *GOS2* & *RXRG*, *PPP2R2B* & *STK32A*, *PROS1* & *SCEL* and *CITED1* & *TENM1* (Additional file Table S6). It is summarized the detail of information about inquired genes genetic alternation (Additional file Table S7). The queried results discovered 5 genes significantly associated with missense mutation, *FNI* (R534P), *PROS1*((K200I), (Q571K)), *SCEL* (T320S), *SLC34A2*(T688M) and *TENM1* (S1131F), plus the key information about mutation type, protein change sites and mutation occurrence in patient's case number (Fig. 4b). Furthermore, these 5 genes are also significantly higher expression in thyroid cancer cases (P value < 0.01) explored in TCGA database (Fig. 3c).

Functional and gene ontology pathway enrichment analysis

The key functional pathway enrichment analysis was performed for the significant genes in Blue module. The significantly enriched pathways mainly concentrated in cell adhesion, extracellular matrix and low density lipoprotein metabolic, also included membrane-associated biological processes and cellular components (Table 1). There are 8 genes in Blue module and 4 clusters of significance enriched KEGG pathways identified by Genclip. The most significant top 2 cluster pathways are resorted to associated with Thyroid Cancer, small/non-small lung cancer or other cancers (Table 1, KEGG Pathway Analysis, cluster1 & 2, Additional file Figure S6). The other 2 clusters pathways mainly involved in the cell adhesion, cell junction interaction & organization, platelet activation & degranulation, and leukocyte transendothelial migration (Table 1). The GO analysis identified 2 significant clusters functional associated with the responses to lipid, hormone, steroid hormone and organic cyclic compound (Table 1, GO Analysis, Additional file Figure S6A). According to previously research reports, 12 genes were involving in constructed a co-citation network. Through literature profiles analysis, the significant genes in blue module are mainly clustered in functions related to type 2 diabetes, cell adhesion, extracellular matrix and low density lipoprotein (Table 1, GO Analysis).

Discussion

In this study, to discover novel biomarkers to accelerate the precise clinical diagnostics for thyroid nodule cases

A

The comparison of significant genes expression level among different case groups

Case Groups	One			Two			Three			Four								
Cytology	Indeterminate			Indeterminate			Benign			Malignant								
Histopathology	Benign			Malignant			Benign			Malignant								
GeneSymbol	Mean	SEM	No.	Mean	SEM	No.	FC (M/B)	p value	Mean	SEM	No.	FC (B/B)	p value	Mean	SEM	No.	FC (M/B)	p value
CC2D2B	4.5471	0.0509	180	4.9695	0.0991	85	1.0929	< 0.0001	4.5078	0.0696	44	0.9914	0.7180	4.8765	0.0929	55	1.0725	0.002
CCND1	8.7840	0.0532	180	9.5306	0.1062	85	1.0850	< 0.0001	8.4014	0.0939	44	0.9564	0.0013	10.2243	0.0925	55	1.1640	< 0.0001
CFH	6.4653	0.0806	180	7.1735	0.1893	85	1.1095	< 0.0001	6.4615	0.1517	44	0.9994	0.9831	9.0629	0.1889	55	1.4018	< 0.0001
CITED1	7.1338	0.0981	180	8.6891	0.2053	85	1.2180	< 0.0001	6.7193	0.0929	44	0.9419	0.0433	9.7515	0.2111	55	1.3669	< 0.0001
CLDN1	7.6796	0.1184	180	9.4430	0.2029	85	1.2296	< 0.0001	7.3308	0.1857	44	0.9546	0.1755	11.1833	0.1335	55	1.4562	< 0.0001
CLDN16	6.4054	0.0821	180	8.2403	0.2322	85	1.2865	< 0.0001	6.3907	0.1392	44	0.9977	0.9349	10.1718	0.2151	55	1.5880	< 0.0001
FBXO2	7.0883	0.0325	180	7.5309	0.0638	85	1.0624	< 0.0001	7.0122	0.0454	44	0.9893	0.2748	7.9077	0.0839	55	1.1156	< 0.0001
FN1	8.8063	0.1290	180	10.1406	0.2211	85	1.1515	< 0.0001	9.5825	0.2684	44	1.0881	0.0086	12.4590	0.1306	55	1.4148	< 0.0001
FRMD3	6.9935	0.0742	180	7.7732	0.1510	85	1.1115	< 0.0001	6.7098	0.0795	44	0.9594	0.0691	8.1860	0.1337	55	1.1705	< 0.0001
G0S2	7.8689	0.0803	180	8.6001	0.1390	85	1.0929	< 0.0001	8.2743	0.1643	44	1.0515	0.0266	9.2176	0.1089	55	1.1714	< 0.0001
GABRB2	4.8529	0.0418	180	6.5880	0.2423	85	1.3575	< 0.0001	4.7676	0.0477	44	0.9824	0.3327	9.1577	0.2228	55	1.8870	< 0.0001
KRT19	7.9836	0.0464	180	8.8815	0.1368	85	1.1125	< 0.0001	7.9101	0.0672	44	0.9908	0.4615	10.4713	0.1432	55	1.3116	< 0.0001
LIPH	4.6926	0.0640	180	6.3388	0.1831	85	1.3508	< 0.0001	4.4334	0.0783	44	0.9448	0.0566	7.9835	0.1626	55	1.7013	< 0.0001
MDK	7.4542	0.0344	180	8.1890	0.0768	85	1.0986	< 0.0001	7.2495	0.0539	44	0.9725	0.0065	8.9670	0.0871	55	1.2030	< 0.0001
NDUFC2	6.5858	0.0630	180	7.0248	0.1058	85	1.0667	0.0002	6.4155	0.0819	44	0.9741	0.2045	6.9706	0.1242	55	1.0584	0.0042
PPP2R2B	6.5754	0.0303	180	6.4025	0.0579	85	0.9737	0.0039	6.4608	0.0612	44	0.9826	0.0950	6.0673	0.0393	55	0.9227	< 0.0001
PROS1	7.8668	0.0744	180	9.1498	0.1852	85	1.1631	< 0.0001	7.9400	0.1236	44	1.0093	0.6528	10.9580	0.1274	55	1.3929	< 0.0001
RXRG	5.7742	0.0552	180	6.9801	0.1706	85	1.2088	< 0.0001	5.5145	0.0352	44	0.9550	0.0227	7.9765	0.1641	55	1.3814	< 0.0001
SCEL	5.0470	0.0761	180	6.1447	0.1549	85	1.2175	< 0.0001	4.6304	0.0914	44	0.9174	0.0101	7.5012	0.1951	55	1.4863	< 0.0001
SERGEF	8.8427	0.0562	180	9.3538	0.0836	85	1.0578	< 0.0001	8.6541	0.0684	44	0.9787	0.1135	9.4299	0.1032	55	1.0664	< 0.0001
SLC34A2	7.4009	0.1153	180	9.1150	0.2673	85	1.2316	< 0.0001	7.4303	0.2147	44	1.0040	0.9088	11.6827	0.2289	55	1.5786	< 0.0001
ST3GAL5	9.3110	0.0485	180	9.9544	0.0879	85	1.0691	< 0.0001	9.1612	0.0649	44	0.9839	0.1484	10.4560	0.0805	55	1.1230	< 0.0001
STK32A	6.5912	0.0677	180	7.3380	0.1027	85	1.1133	< 0.0001	6.1852	0.0722	44	0.9384	0.0046	7.8668	0.1127	55	1.1935	< 0.0001
TENM1	5.6507	0.0888	180	6.7176	0.1683	85	1.1888	< 0.0001	5.2242	0.0739	44	0.9245	0.0210	6.9922	0.1526	55	1.2374	< 0.0001

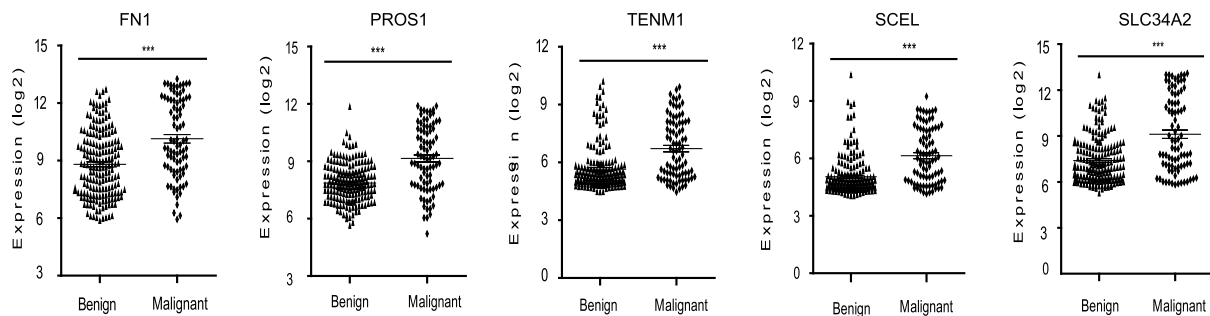
Notes:

FC: Fold Change

M/B : Malignant compared to Benign

B/B: Benign compared to Benign

B



C

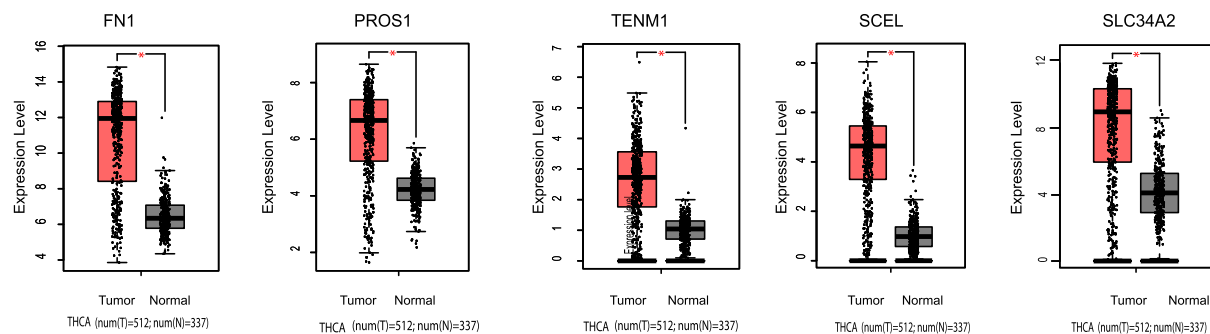
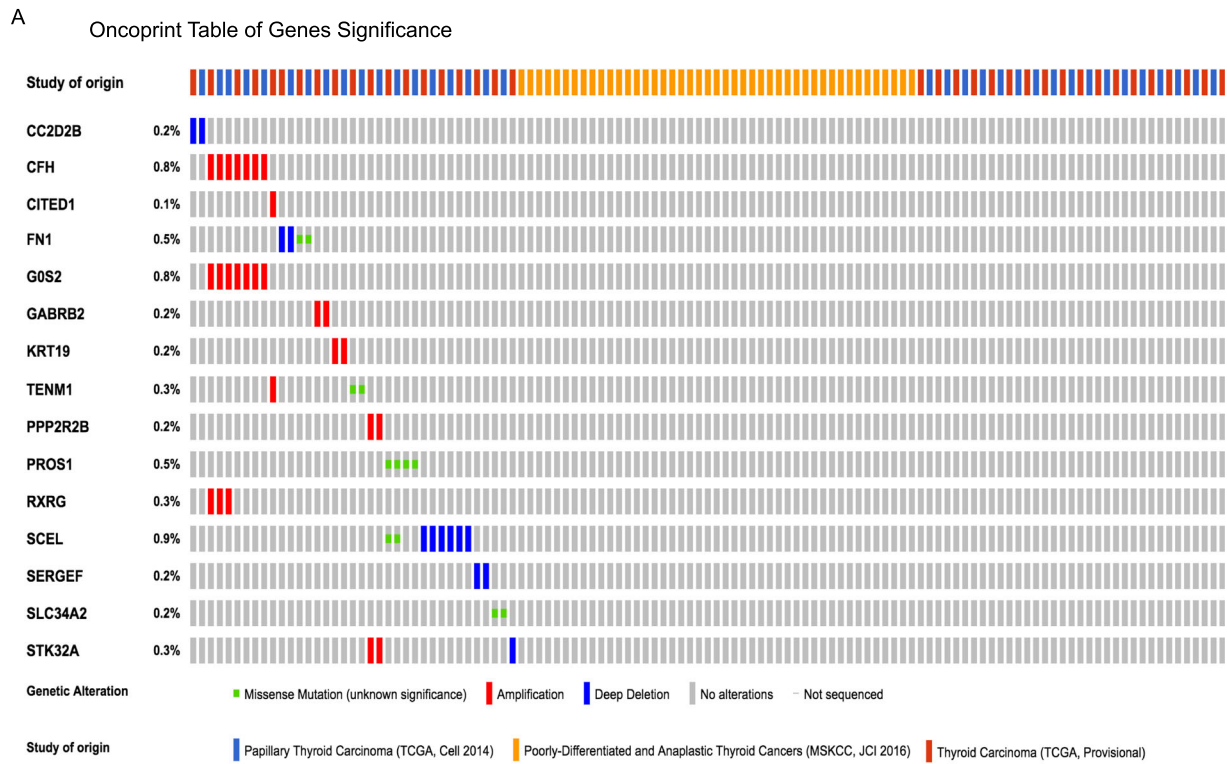


Fig. 3 Validation of the gene expression levels of novel biomarkers between histopathological benign and malignant. **a** Expression comparison of 24 signature genes among benign and malignant cases groups; **b** Validation the expression level of potential biomarkers of FN1, TENM1, SCEL, SCL34A2, PROS1. **c** Validation based on TCGA data via GEPIA, including FN1, TENM1, SCEL, SCL34A2, PROS1 (***, represent p value < 0.0001; *, represents p value < 0.01)



B
The Mutations of Genes Significance in Thyroid Cancer cases

Gene	Study	Sample ID	Cancer Type	Protein Change	Mutation Type	Copy #	COSMIC	Allele Freq (T)	# Mut in Sample
FN1	Papillary Thyroid Carcinoma	TCGA-J8-A300-01	Papillary Thyroid Cancer	R534P	Missense	Diploid	1	0.33	9
	Thyroid Carcinoma (TCGA, Provisional)	TCGA-J8-A300-01	Thyroid Carcinoma	R534P	Missense	Diploid	1	0.33	9
PROS1	Papillary Thyroid Carcinoma	TCGA-EM-A22L-01	Papillary Thyroid Cancer	K200I	Missense	Diploid	1	0.36	10
	Thyroid Carcinoma (TCGA, Provisional)	TCGA-EM-A22L-01	Follicular Thyroid Cancer	K200I	Missense	Diploid	1	0.36	10
	Papillary Thyroid Carcinoma	TCGA-EM-A4FV-01	Papillary Thyroid Cancer	Q571K	Missense	Diploid	1	0.21	20
	Thyroid Carcinoma (TCGA, Provisional)	TCGA-EM-A4FV-01	Thyroid Carcinoma	Q571K	Missense	Diploid	1	0.21	21
SCEL	Papillary Thyroid Carcinoma	TCGA-EM-A22L-01	Papillary Thyroid Cancer	T320S	Missense	Diploid	1	0.19	10
	Thyroid Carcinoma (TCGA, Provisional)	TCGA-EM-A22L-01	Thyroid Carcinoma	T320S	Missense	Diploid	1	0.19	10
SLC34A2	Papillary Thyroid Carcinoma	TCGA-BJ-A2NA-01	Papillary Thyroid Cancer	T688M	Missense	Diploid	1	0.38	25
	Thyroid Carcinoma (TCGA, Provisional)	TCGA-BJ-A2NA-01	Thyroid Carcinoma	T688M	Missense	Diploid	1	0.38	27
TENM1	Papillary Thyroid Carcinoma	TCGA-ET-A3DO-01	Papillary Thyroid Cancer	S1131F	Missense	Diploid	1	0.05	17
	Thyroid Carcinoma (TCGA, Provisional)	TCGA-ET-A3DO-01	Thyroid Carcinoma	S1131F	Missense	Diploid	1	0.05	17

Fig. 4 Validate Signature genes by Blast with independent cases via the cBioportal database. **a** Oncoprint table of significant signature genes. Through the cBioPortal Cancer Genomics database, the genes significance (GS) in blue module were explored multidimensional cancer genomics datasets in the context of clinical data and biologic pathways. The Oncoprint table summarizes genomic alterations in all queried genes across samples. Each row represents a gene, and each column represents a tumor sample. Red bars indicate gene amplifications, blue bars are homozygous deletions, and green squares are nonsynonymous mutations. **b** The summary Mutations table of query genes. The tabular view provides additional information about all mutations in each query genes

Table 1 GO and KEGG pathway enrichment analysis of genes significance

A KEGG Pathway Analysis				
Pathway analysis	Hit	Total	P-Value	Q-Value
cluster1 Enrichment Score : 2.61				
KEGG THYROID CANCER	<u>2</u>	29	0.0013	0.0121
KEGG NON SMALL CELL LUNG CANCER	<u>2</u>	54	0.0046	0.0165
cluster2 Enrichment Score : 2.49				
KEGG SMALL CELL LUNG CANCER	<u>3</u>	84	0.0005	0.0087
KEGG PATHWAYS IN CANCER	<u>3</u>	325	0.0218	0.0328
cluster3 Enrichment Score : 2.18				
REACTOME TIGHT JUNCTION INTERACTIONS	<u>2</u>	30	0.0014	0.0086
KEGG LEUKOCYTE TRANSENDOTHELIAL MIGRATION	<u>2</u>	116	0.02	0.0328
KEGG TIGHT JUNCTION	<u>3</u>	132	0.0018	0.0081
REACTOME CELL CELL ADHESION SYSTEMS	<u>2</u>	59	0.0055	0.0164
REACTOME CELL JUNCTION ORGANIZATION	<u>2</u>	84	0.0108	0.0244
KEGG CELL ADHESION MOLECULES CAMS	<u>2</u>	133	0.0259	0.0358
KEGG COMPLEMENT AND COAGULATION CASCADES	<u>2</u>	69	0.0074	0.0191
cluster4 Enrichment Score : 1.64				
REACTOME PLATELET DEGRANULATION	<u>2</u>	86	0.0113	0.0227
REACTOME PLATELET ACTIVATION	<u>2</u>	167	0.0394	0.0507
REACTOME CELL SURFACE INTERACTIONS AT THE VASCULAR WALL	<u>2</u>	94	0.0134	0.0242
REACTOME FORMATION OF PLATELET PLUG	<u>2</u>	186	0.0479	0.0575
B GO Analysis				
GO Term	Hit	Total	P-Value	Q-Value
cluster1 Enrichment Score : 6.57				
response to lipid	<u>8</u>	945	6.95E-09	3.50E-06
response to hormone	<u>7</u>	1134	2.06E-05	0.0015
response to organic cyclic compound	<u>8</u>	959	9.73E-09	2.45E-06

with cytology indeterminate, we designed the whole project workflow, selected specific dataset (GSE34289) and applied the WGCNA approach to analyze the gene expression profiling of thyroid nodule that generated from FNA clinical samples (Fig. 1). The gene expression profile contained 172 specific genes designed for promise diagnostic assessment [8], which are mainly involving in variously biological and cellular processes that related to energy metabolism, cell differentiation, and cellular development. and aimed to discover some novel biomarkers to accelerate the precise clinical diagnostics for thyroid cancer. It is representative to discover the signature genes significantly associated with thyroid nodules malignancy through gene expression profiling analysis of these cases. Furthermore, this dataset was generated from 49 national widely clinical sites, collected from 3789 patients and evaluated 4812 thyroid nodules samples (size > 1 cm) in United States and well characterized with higher standard. It obtained 577 cytological indeterminate aspirates and finally selected 265 indeterminate

nodules for further analysis through blinded histopathological review [25]. In addition, this dataset also contained two groups of cases labelled as cytology-benign/histopathology-benign (44 cases) and cytology-malignant/histopathology-malignant (55) with validated cytopathology and histopathological features (Additional file Table S1). Based on this dataset characters, we utilized totally 364 samples and split into 4 groups in this study. The expression level of identified signature genes will be explored in cytology-benign/histopathology-benign and cytology-malignant/histopathology-malignant groups as negative and positive control. Moreover, as designed in the workflow, the discovered signature genes would be validated through another independently investigations of thyroid carcinomas patient's cases in TCGA via GEIPA (Fig. 3c) and indicated these genes sensitiveness with statistical analysis.

Compared with other computational methodology, the WGCNA have unique merits, which could be robust and sensitive detection of the subset of genes co-

expression as functional modules from the entire transcriptome and without pre-filtering to cause selective bias or losing useful information [20]. It was designed to discover the networks and genes associated with phenotypes of target by using unsupervised clustering and constructing gene module. The constructed gene co-expression module consists of a group of genes that maintain a consistent expression relationship and share a common biological regulation function that independent of a priori defined gene sets or pathways [31]. Previously, WGCNA has been successfully applied to biomarker discovery for cancer diagnostics, such as discovered microRNA expression network in prostate cancer [32] and identified ASPM as a potential biomarker in glioblastoma [33].

In our study, our results firstly demonstrated that it is reasonable to build on the co-expression networks with clinical traits (histopathology, age, gender) using Pearson correlation analysis (Additional file Figure S1). To discover the related modules to histopathological phenotype, we calculated the modules statistic significantly with the linear mixed effects model for testing the association of the node to the histological phenotype. We analyzed the gene expression profile data and identified three module eigengenes (ME), blue, Green and Turquoise module, are significantly associated with histological feature of malignant thyroid cancer (Fig. 2, Additional file Figure S4). Through the Eigengene dendrogram analysis, we discovered the most significantly hub, the blue module that contained 24 genes, related to histological feature (Fig. 2d&e).

To validate the WGCNA analysis results, we took two approaches to test signature genes positive expression and correlation. Firstly, we compared the signature genes expression level between sample group 2/3/4 and group one separately, which represent as test, negative and positive control. Gone through the results, we setup the fold change > 1.05 or fold change < 0.98 (plus P value < 0.001) as the cutoff for statistics significant standard. As results indicated, all 24 signature genes do not significant expression difference although some genes expression with lower or higher level in the double cytology/histopathology benign case group, which defined as negative control. While 22 signature genes were significantly higher expression ($FC > 1.0$, P value < 0.001), except *PPP2R2B* with lower expression ($FC = 0.9227$, P value = 0.0042) and *CC2D2B* with higher expression ($FC = 1.0725$, P value = 0.002), in the double cytology/histopathology malignant cases group, which works as positive control. These results indicated that 24 signature genes could significantly differentiate the malignancy and benign cases (positive rate = 91.67%, 22/24). For the cytology-Indeterminate/histopathology-malignant group, the 23 signature genes show significant

higher expression (Fold change > 1.0 , P value < 0.001) than benchmark (positive rate = 95.83%, 23/24), while only *PPP2R2B* with lower expression (P value = 0.0039) (Table 1). Furthermore, identified 5 potential biomarker genes were all significantly higher expression ($FC > 1.15$, P value < 0.0001) in both cytology-indeterminate/histopathology-malignant group and double cytology/histopathology malignant group (Fig. 3a&b). These results suggest these 5 genes have potential to be the biomarker candidate for differentiation the malignancy among the indeterminate cases. In the secondary approach, we explored these 24 genes through the cBioPortal Cancer Genomic database, which is containing many published cancer studies datasets from CCLE and TCGA [28], and verified through 3 independent thyroid cancer investigations that contained 915 patient's datasets. The 16 genes were matched in 37 (4.0%) of 915 patient's cases with genetic alternation of missense mutation, amplification, deep deletion and copy-number alterations, and listed as *CC2D2B*, *CFH*, *CITED1*, *FNI*, *GOS2*, *GABRB2*, *KRT19*, *TENM1*, *PPP2R2B*, *PROS1*, *RXRG*, *SCEL*, *SERGEF*, *SLC34A2* and *STK32A*. Some of these generic alternations were associated with papillary thyroid carcinoma metastasis to brain [34] and could be useful as histopathological biomarkers for papillary thyroid carcinoma [25, 35]. In addition, the queried results also discovered 5 genes with missense mutation significantly (P value < 0.01) associated thyroid cancer cases, listed as *FNI* (R534P), *PROS1* ((K200I), (Q571K)), *SCEL* (T320S), *SLC34A2*(T688M) and *TENM1* (S1131F) (Fig. 4b). We compared these genes expression level between groups, and our results indicated these genes could significantly differentiate the cases of benign and malignant among cytological indeterminate cases (Fig. 3a). In addition, 5 potential biomarkers significantly higher expression in malignant cases (P value < 0.0001 , Fig. 3b). Furthermore, to validate these signature genes, we blast the 5 potential biomarkers through TGCA database with other independent clinical cases (about 849 cases, Fig. 3c). These 5 genes are also significantly higher expression in thyroid cancer cases (P value < 0.01) explored in GEPIA with thyroid cancer clinical cases data. These two validation approach made results more convincible.

Through the Genclip analysis, we found that these genes are mainly concentrate on the GO pathways that involving in physiological response of hormone and steroid hormone, and regulation of cell migration and adhesion, cell junction interaction, etc. (Table 1, Additional file Figure S6A). The involving pathways are significantly concentrated in subgroups of thyroid cancer, non-small cell lung cancer and cancer, process, signaling, extracellular region, and transporter activity (Additional file Figure S6B, and Table 1). It indicates that these functions may be associated with metabolism and accelerated growth

and development of obesity individuals. Notably, the results of GO enrichment analysis also provide more significant pathways with biological annotations (Table 1).

Checked with published literature, these 5 genes were reported associated with the transform progress of multiple carcinomas. Fibronectin 1 (*FNI*) is a glycoprotein existing with soluble dimer or multimeric form in different conditions. *FNI* is involved in multiple cell adhesion and migration processes, including embryogenesis, wound healing, blood coagulation, host defense, and found with higher expression in metastasis [36]. It was reported that *FNI* is over expression in the Papillary Thyroid Carcinoma [37] and listed as potential biomarker for diagnostics. *PROS1* (Protein S 1) is a vitamin K-dependent plasma protein that works as a cofactor of the anticoagulant protease. It could activate protein C (APC) and inhibit blood coagulation [38]. The genetic mutation of this gene will result in autosomal dominant hereditary thrombophilia [39] and malignant glioma [40]. The *PROS1* and *FNI* others 12 genes alternations were identified as important diagnostic biomarkers for thyroid cancer through the meta-analysis the gene expression profiling of clinical thyroid nodules [41]. *SCEL* (Sciellin) is the precursor to the cornified envelope of terminally differentiated keratinocytes. *SCEL* is overexpressed in the papillary thyroid carcinoma and worked as key regulator in mesenchymal-to-epithelial transition and dynamically regulated through the metastasis process [36]. *SCEL* was high expression in thyroid tumor tissue and significantly associated with I-131 [42]. *TENM1* (Teneurin transmembrane protein 1) is involving in pattern formation and morphogenesis [43]. *TENM1* was overexpression in thyroid cancer and associated with thyroidal invasion [44] and identified as potential marker of papillary thyroid carcinoma progress [36, 45]. *SLC34A2* (solute carrier family 34 member 2) is a member of the *SLC34* solute carrier protein family and coded for pH-sensitive sodium-dependent phosphate transporter (NaPi2b), which is a multi-transmembrane [46] and The physiological function of *SLC34A2* is transcellular inorganic phosphate absorption and maintenance of phosphate homeostasis [47] and cell differentiation. *SLC34A2* is overexpressed in multiple cancer types, including lung, ovarian, and thyroid cancers [48] and identified as potential therapeutic target for non-small cell lung and Ovarian cancer [48]. Combined with these independent research results, it is strong suggest that *FNI* (R534P), *PROS1* ((K200I), (Q571K)), *SCEL* (T320S), *SLC34A2* (T688M) and *TENM1* (S1131F) are potential novel biomarker candidates significantly associated with thyroid carcinomas and could differentiate the malignant thyroid nodule among cytological indeterminate cases. As mentioned previously, the 7-gene MT biomarkers panel was broadly recommended to evaluate the residual cytological indeterminate thyroid nodules and

estimate with high specificity (~ about 90%) [14–16]. However, the sensitivity of 7-gene MT biomarkers panel testing showed huge variation (from 44 to 100%) in clinical practice [6, 17]. It suggests that there are some unknown biomarkers existing in these indeterminate cases. It is possible that our identified 5 novel biomarkers genes could contribute to enhance specificity of previously 7-gene MT biomarkers panel. The combined application of these two panel biomarkers would get more promise and precise clinical diagnostic results for nodules malignancy in some cases population.

There are some limitations and several novelties in our study. The FNA yield cytology indeterminate cases are including subtype of follicular lesion, follicular neoplasm and suspicious or malignancy. For the first limitation from this dataset, lacking of these cases histopathological information about thyroid cancer subtype sorting, such as follicular adenoma (FA), follicular carcinoma (FC) and papillary thyroid carcinoma (PTC), we could not track these potential biomarkers back to original patient's pathological status and dig deeper insight. Secondly, the gene expression profile was limited to 172 genes for promise diagnostic assessment [8]. It will cause to miss other genes significantly associated with malignant thyroid carcinomas by this pre-filter selection. Thirdly, due to the bioinformatics analysis nature, the discovered specific GO pathways and KEGG pathways were referred from previously literatures and did not be further investigated. Although we explored these significant genes associated with histological feature through TCGA database via cBioPortal and compared with the other two groups case in the same dataset, these potential biomarkers will be required to verify with storing patient's cases according to subtype of thyroid cancer by immunohistochemistry (IHC) or other genetic detection method, like qPCR or sequencing in coming research work. Therefore, more number and sorting subtype patient's cases are mandatory to verify these potential biomarkers for thyroid cancer precise diagnostics in the future cohort study. On the other side, our study has several novelties. Firstly, we applied reverse strategy by using WGCNA approach to discover the genes significantly associated with malignant histopathological feature in clinical FNA samples with cytological indeterminate feature. In parallels, compared these signature genes expression level among histopathological benign and malignant groups, the results indicated that signature genes have significant positive overexpression in Malignant groups and negative overexpression in benign group (Table 1). Secondly, we inquired the key functional & GO pathways that genes significance in module involving in the progress of thyroid carcinomas by Genclip enrichment analysis (Table 1). The results will be a clause for the next step research. Thirdly, exploring through TCGA database, we discovered 5 novel potential biomarkers to differentiate

the malignant and benign thyroid nodules, which were identified as potential biomarkers of malignant thyroid cancer in previously independent researches. Furthermore, these 5 genes were validated with significant higher expression level in the TCGA thyroid cancer cases (Fig. 3c). These results are partially as evidences to support our results and research strategy.

Conclusions

Our study identified five novel signature genes with missense mutation, *FN1* (R534P), *PROS1*(K200I), (Q571K)), *SCEL* (T320S), *SLC34A2*(T688M) and *TENM1* (S1131F) that highlighted as potential biomarkers to rule out nodules malignancy. These novel results provide new insight and strategy to identify these potential biomarkers and differentiate malignant histopathological thyroid nodules with cytological Indeterminate. The clinical validation and application of these prognostic biomarkers will facilitate the precise diagnostics and help to enhance the healthcare efficiency for thyroid cancer patients.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12885-020-6676-z>.

Additional file 1: Figure S1. Sample dendrogram and Clinic Feature traits heatmap. Clustering dendrogram of samples based on their Euclidean distance. The clinical feature traits were histopathology, gender and age. The white color means a low value, red means a high value, and grey represents a missing entry.

Additional file 2: Figure S2. Analysis of network topology for various soft-thresholding powers. In panel (A), the scale-free topology model fit index (signed R^2 , y-axis) shows as a function of the soft-thresholding power (x-axis). In panel (B), the mean connectivity (k_i , y-axis) displays as a function of the soft-thresholding power (x-axis) under different weighting coefficients. The connectivity k_i of node i equals the number of its direct connections to other nodes. $P(k)$ indicates the frequency distribution of the connectivity. The higher the coefficient, the closer the network is to the distribution of the scale free network.

Additional file 3: Figure S3. Heatmap plot of genes network. The heatmap represents the Topological Overlap Matrix (TOM) among all Genes used for analysis. Light color represents low overlap and progressively darker red color represents higher overlap. Blocks of darker colors along the diagonal are the modules. The gene dendrogram and module assignment are also shown along the left side and the top.

Additional file 4: Figure S4. Clustering dendrogram of Genes, with dissimilarity based on topological overlap. Different colors index different modules. Six modules are identified. Grey bars represent Genes that do not belong to any other modules and are not co-expressed.

Additional file 5: Figure S5. The scatterplots of Gene Significance (GS) for histology vs. Module Membership (MM) in the all modules (A~E). There is a highly significant correlation between GS and MM in this module, implying that the most important (central) elements of blue module also tend to be highly correlated with thyroid nodule histology trait.

Additional file 6: Figure S6. GO and pathway analysis of Genes Significant (GS). Clustering analysis of the biological functions of 22 genes in previous studies for GO (A) and Pathway (B) generated by the GenClip software. In the heatmap, the black color represents that the biological function of the corresponding gene-term association has not been reported yet. While light green color means that the corresponding gene-

term association positively has been reported. The color scale bar for proportion of genes associated were labelled.

Additional file 7: Table S1. The samples information and case group sorting.

Additional file 8: Table S2. The pick soft threshold for Module.

Additional file 9: Table S3. List of the 29 Genes in the grey module.

Additional file 10: Table S4. List of the 24 significant genes in the blue module.

Additional file 11: Table S5. The Copy-number Alterations of significant genes.

Additional file 12: Table S6. The Mutual Exclusivity tab of significant gene pairs. The genes pairs alternated in Thyroid cancer are mutual exclusivity. The tab provides summary statistics significant on mutual exclusivity and co-occurrence of genomic alterations in each pair of query genes. The mutual exclusivity is significant for the other two gene pairs ($P < 0.05$). The P values are determined by a Fisher's exact test with the null hypothesis that the frequency of occurrence of a pair of alterations in two genes is proportional to their uncorrelated occurrence in each gene.

Additional file 13: Table S7. The Genetic Alterations type of query genes.

Abbreviations

7-gene MT: 7-gene molecular panel; CITED1: Cbp/p300 Interacting transactivator with Glu/Asp rich Carboxy-terminal domain 1; FN1: Fibronectin 1; FNA: Fine-needle aspiration; GO: Gene ontology; GS: Gene significance; ME: Module eigengene; MM: Module membership; MS: Module significance; PCA: Principal component analysis; PROS1: Protein S; SCEL : Scellin; SLC34A2: Solute carrier family 34 member 2; TCGA: The Cancer Genome Atlas; TENM1: Teneurin transmembrane protein 1; TOM: Topological Overlap Matrix; WGCNA: Weighted gene co-expression network analysis

Acknowledgements

We highly value the support of the medical doctors of Department of Radiology, Affiliated Renmin Hospital of Jiangsu University as well as from the colleagues of the Life Science Institute of Jiangsu University.

Authors' contributions

SL, YYH and DW designed the study; DW, SL, YYH collected data, performed data analysis, and drafted the manuscript; DW, SL, YYH, SH and YH interpreted and summarized the results; DW, SL, YYH wrote and revised the manuscript; all authors have read and approved the final version of manuscript.

Funding

In this study, all activities of study design and collection, analysis, and interpretation data and in writing manuscript, plus the cost of manuscript publication, were supported by the Funding from Jiangsu University oversea outstanding talent (17JDG001). It was awarded to Dr. Shubai Liu and Dr. Yingying He.

Availability of data and materials

The original data used in this study was mentioned in the section "Gene Expression Data Source" [GSE34289]. The secondary datasets used or generated by analysis in this study are available from online supplementary.

Ethics approval and consent to participate

The original written consent from donor was described by original study operators Erik et al. (2012). We also followed related ethics regulatory rules according to previously statement, this study was discussed and got approved our institute committee of ethic.

The full name of the ethics committee of Jiangsu university, life science college:

Shudong Hu, MD, associate professor
Yongzhong Hou, Ph.D., Professor
Yingying He, Ph.D., professor
Shubai Liu, Ph.D., Professor
Haifeng Shi, Ph.D., Professor

Consent for publication

Not Applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Institute of Life Sciences, Jiangsu University, 301 Xuefu Road, JinKou District, Zhenjiang 212013, PR China. ²Department of Radiology, Affiliated Renmin Hospital of Jiangsu University, Zhenjiang, Jiangsu, PR China.

Received: 22 February 2019 Accepted: 24 February 2020

Published online: 12 March 2020

References

- Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J. Cancer statistics in China, 2015. *CA Cancer J Clin*. 2016;66(2):115–32.
- Chen AY, Jemal A, Ward EM. Increasing incidence of differentiated thyroid cancer in the United States, 1988–2005. *Cancer*. 2009;115(16):3801–7.
- Treglia G, Aktolun C, Chiti A, Frangos S, Giovannella L, Hoffmann M, Iakovou I, Mihailovic J, Krause BJ, Langsteger W, et al. The 2015 revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma: the “evidence-based” refusal to endorse them by EANM due to the “not evidence-based” marginalization of the role of nuclear medicine. *Eur J Nucl Med Mol Imaging*. 2016;43(8):1486–90.
- Bak M, Peter I, Nyari T, Simon P, Ujlaky M, Boer A, Kasler M. On-site fine-needle aspiration cytology of thyroid nodules. Quality assurance of the Bethesda system for reporting thyroid cytopathology (2008). *Orv Hetil*. 2015;156(41):1661–6.
- Mallick UK, American Thyroid A. The revised American Thyroid Association management guidelines 2009 for patients with differentiated thyroid cancer: an evidence-based risk-adapted approach. *Clin Oncol (R Coll Radiol)*. 2010;22(6):472–4.
- Nikiforov YE, Ohori NP, Hodak SP, Carty SE, LeBeau SO, Ferris RL, Yip L, Seethala RR, Tublin ME, Stang MT, et al. Impact of mutational testing on the diagnosis and management of patients with cytologically indeterminate thyroid nodules: a prospective analysis of 1056 FNA samples. *J Clin Endocrinol Metab*. 2011;96(11):3390–7.
- Wang CC, Friedman L, Kennedy GC, Wang H, Kebebew E, Steward DL, Zeiger MA, Westra WH, Wang Y, Khanafshar E, et al. A large multicenter correlation study of thyroid nodule cytopathology and histopathology. *Thyroid*. 2011;21(3):243–51.
- Chudova D, Wilde JI, Wang ET, Wang H, Rabbee N, Egidio CM, Reynolds J, Tom E, Pagan M, Rigi CT, et al. Molecular classification of thyroid nodules using high-dimensionality genomic data. *J Clin Endocrinol Metab*. 2010;95(12):5296–304.
- Pagedar NA, Chen DH, Wasman JK, Savvides P, Schluchter MD, Wilhelm SM, Lavertu P. Molecular classification of thyroid nodules by cytology. *Laryngoscope*. 2008;118(4):692–6.
- Nam SY, Han BK, Ko EY, Kang SS, Hahn SY, Hwang JY, Nam MY, Kim JW, Chung JH, Oh YL, et al. BRAF V600E mutation analysis of thyroid nodules needle aspirates in relation to their ultrasonographic classification: a potential guide for selection of samples for molecular analysis. *Thyroid*. 2010;20(3):273–9.
- Cancer Genome Atlas Research N. Integrated genomic characterization of papillary thyroid carcinoma. *Cell*. 2014;159(3):676–90.
- Grogan RH, Mitmaker EJ, Clark OH. The evolution of biomarkers in thyroid cancer—from mass screening to a personalized biosignature. *Cancers (Basel)*. 2010;2(2):885–912.
- Yip L, Ferris RL. Clinical application of molecular testing of fine-needle aspiration specimens in thyroid nodules. *Otolaryngol Clin N Am*. 2014;47(4):557–71.
- Giordano TJ, Beaudenon-Huibregtse S, Shinde R, Langfield L, Vinco M, Laosinchai-Wolf W, Labourier E. Molecular testing for oncogenic gene mutations in thyroid lesions: a case-control validation study in 413 postsurgical specimens. *Hum Pathol*. 2014;45(7):1339–47.
- Beaudenon-Huibregtse S, Alexander EK, Guttler RB, Hershman JM, Babu V, Blevins TC, Moore P, Andruss B, Labourier E. Centralized molecular testing for oncogenic gene mutations complements the local cytopathologic diagnosis of thyroid nodules. *Thyroid*. 2014;24(10):1479–87.
- Nayar R, Ivanovic M. The indeterminate thyroid fine-needle aspiration: experience from an academic center using terminology similar to that proposed in the 2007 national cancer institute thyroid fine needle aspiration state of the science conference. *Cancer*. 2009;117(3):195–202.
- Nikiforov YE, Steward DL, Robinson-Smith TM, Haugen BR, Klopper JP, Zhu Z, Fagin JA, Falciglia M, Weber K, Nikiforova MN. Molecular testing for mutations in improving the fine-needle aspiration diagnosis of thyroid nodules. *J Clin Endocrinol Metab*. 2009;94(6):2092–8.
- Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, Pacini F, Randolph GW, Sawka AM, Schlumberger M, et al. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2016;26(1):1–133.
- Haugen BR. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: what is new and what has changed? *Cancer*. 2017;123(3):372–81.
- Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005;4:Article17.
- Zhang J, Baddoo M, Han C, Strong MJ, Cvitanovic J, Moroz K, Dash S, Flemington EK, Wu T. Gene network analysis reveals a novel 22-gene signature of carbon metabolism in hepatocellular carcinoma. *Oncotarget*. 2016;7(31):49232–45.
- Guo Y, Xing Y. Weighted gene co-expression network analysis of pneumocytes under exposure to a carcinogenic dose of chloroprene. *Life Sci*. 2016;151:339–47.
- Zhu XL, Ai ZH, Wang J, Xu YL, Teng YC. Weighted gene co-expression network analysis in identification of endometrial cancer prognosis markers. *Asian Pac J Cancer Prev*. 2012;13(9):4607–11.
- Shi K, Bing ZT, Cao GQ, Guo L, Cao YN, Jiang HO, Zhang MX. Identify the signature genes for diagnose of uveal melanoma by weight gene co-expression network analysis. *Int J Ophthalmol*. 2015;8(2):269–74.
- Alexander EK, Kennedy GC, Baloch ZW, Cibas ES, Chudova D, Diggans J, Friedman L, Kloos RT, LiVolsi VA, Mandel SJ, et al. Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. *N Engl J Med*. 2012;367(8):705–15.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
- Li A, Horvath S. Network neighborhood analysis with the multi-node topological overlap measure. *Bioinformatics*. 2007;23(2):222–31.
- Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401–4.
- Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):pl1.
- Huang ZX, Tian HY, Hu ZF, Zhou YB, Zhao J, Yao KT. GenCLiP: a software program for clustering gene lists by literature profiling and constructing gene co-occurrence networks related to custom keywords. *BMC Bioinformatics*. 2008;9:308.
- Stuart JM, Segal E, Koller D, Kim SK. A gene-coexpression network for global discovery of conserved genetic modules. *Science*. 2003;302(5643):249–55.
- Wang L, Tang H, Thayanithy V, Subramanian S, Oberg AL, Cunningham JM, Cerhan JR, Steer CJ, Thibodeau SN. Gene networks and microRNAs implicated in aggressive prostate cancer. *Cancer Res*. 2009;69(24):9490–7.
- Horvath S, Zhang B, Carlson M, Lu KV, Zhu S, Felciano RM, Laurance MF, Zhao W, Qi S, Chen Z, et al. Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proc Natl Acad Sci U S A*. 2006;103(46):17402–7.
- Schulten HJ, Hussein D, Al-Adwani F, Karim S, Al-Maghrabi J, Al-Sharif M, Jamal A, Bakhshab S, Weaver J, Al-Ghamdi F, et al. Microarray expression profiling identifies genes, including cytokines, and biofunctions, as diapedesis, associated with a brain metastasis from a papillary thyroid carcinoma. *Am J Cancer Res*. 2016;6(10):2140–61.
- Nakamura N, Erickson LA, Jin L, Kajita S, Zhang H, Qian X, Rumilla K, Lloyd RV. Immunohistochemical separation of follicular variant of papillary thyroid carcinoma from follicular adenoma. *Endocr Pathol*. 2006;17(3):213–23.
- Huang Y, Prasad M, Lemon WJ, Hampel H, Wright FA, Kornacker K, LiVolsi V, Frankel W, Kloos RT, Eng C, et al. Gene expression in papillary thyroid

- carcinoma reveals highly consistent profiles. *Proc Natl Acad Sci U S A*. 2001; 98(26):15044–9.
37. da Silveira Mitteldorf CA, de Sousa-Canavez JM, Leite KR, Massumoto C, Camara-Lopes LH. FN1, GALE, MET, and QPCT overexpression in papillary thyroid carcinoma: molecular analysis using frozen tissue and routine fine-needle aspiration biopsy samples. *Diagn Cytopathol*. 2011;39(8):556–61.
 38. Dahlback B, Stenflo J. High molecular weight complex in human plasma between vitamin K-dependent protein S and complement component C4b-binding protein. *Proc Natl Acad Sci U S A*. 1981;78(4):2512–6.
 39. Taniguchi F, Morishita E, Sekiya A, Nomoto H, Katsu S, Kaneko S, Asakura H, Ohtake S. Gene analysis of six cases of congenital protein S deficiency and functional analysis of protein S mutations (A139V, C449F, R451Q, C475F, A525V and D599TfsTer13). *Thromb Res*. 2017;151:8–16.
 40. Milinkovic V, Bankovic J, Rakic M, Stankovic T, Skender-Gazibara M, Ruzdijic S, Tanic N. Identification of novel genetic alterations in samples of malignant glioma patients. *PLoS One*. 2013;8(12):e82108.
 41. Griffith OL, Melck A, Jones SJ, Wiseman SM. Meta-analysis and meta-review of thyroid cancer gene expression profiling studies identifies important diagnostic biomarkers. *J Clin Oncol*. 2006;24(31):5043–51.
 42. Abend M, Pfeiffer RM, Ruf C, Hatch M, Bogdanova TI, Tronko MD, Hartmann J, Meineke V, Mabuchi K, Brenner AV. Iodine-131 dose-dependent gene expression: alterations in both normal and tumour thyroid tissues of post-Chernobyl thyroid cancers. *Br J Cancer*. 2013;109(8):2286–94.
 43. Tucker RP, Chiquet-Ehrismann R. Teneurins: a conserved family of transmembrane proteins involved in intercellular signaling during development. *Dev Biol*. 2006;290(2):237–45.
 44. Nikolova DN, Zembutsu H, Sechanov T, Vidinov K, Kee LS, Ivanova R, Becheva E, Kocova M, Toncheva D, Nakamura Y. Genome-wide gene expression profiles of thyroid carcinoma: identification of molecular targets for treatment of thyroid carcinoma. *Oncol Rep*. 2008;20(1):105–21.
 45. Cheng SP, Chen MJ, Chien MN, Lin CH, Lee JJ, Liu CL. Overexpression of teneurin transmembrane protein 1 is a potential marker of disease progression in papillary thyroid carcinoma. *Clin Exp Med*. 2017;17(4):555–64.
 46. Xu H, Bai L, Collins JF, Ghishan FK. Molecular cloning, functional characterization, tissue distribution, and chromosomal localization of a human, small intestinal sodium-phosphate (Na⁺-Pi) transporter (SLC34A2). *Genomics*. 1999;62(2):281–4.
 47. Virkki LV, Biber J, Murer H, Forster IC. Phosphate transporters: a tale of two solute carrier families. *Am J Physiol Renal Physiol*. 2007;293(3):F643–54.
 48. Lin K, Rubinfeld B, Zhang C, Firestein R, Harstad E, Roth L, Tsai SP, Schutten M, Xu K, Hristopoulos M, et al. Preclinical development of an anti-NaPi2b (SLC34A2) antibody-drug conjugate as a therapeutic for non-small cell lung and ovarian cancers. *Clin Cancer Res*. 2015;21(22):5139–50.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

