



# Identification of key candidate genes and pathways in follicular variant papillary thyroid carcinoma by integrated bioinformatical analysis

Lanyu Jing<sup>#</sup>, Fada Xia<sup>#</sup>, Xin Du, Bo Jiang, Yong Chen, Xinying Li

Department of General Surgery, Xiangya Hospital, Central South University, Changsha 410008, China

*Contributions:* (I) Conception and design: L Jing, F Xia, X Li; (II) Administrative support: X Li; (III) Provision of study materials or patients: L Jing, F Xia, X Li; (IV) Collection and assembly of data: L Jing, X Li, Y Chen; (V) Data analysis and interpretation: L Jing, F Xia, X Du, B Jiang, X Li; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

<sup>#</sup>These authors contributed equally to this work.

*Correspondence to:* Xinying Li, MD, PhD, FACS. Department of General Surgery, Xiangya Hospital, Central South University, No.87 Xiangya Road, Changsha 410008, China. Email: [lixinyingcn@126.com](mailto:lixinyingcn@126.com).

**Background:** Follicular variant papillary thyroid carcinoma (FVPTC) is a heterogeneous group of tumors that differ morphologically, genetically, and clinically. This study aimed to investigate the gene mutation and gene expression profiles, especially the pathways in the interaction network and the diagnostic approaches of candidate markers of FVPTC.

**Methods:** The clinicopathological characteristics, gene mutation types, and mRNA expression profiles of patients with FVPTC were studied utilizing the data downloaded from The Cancer Genome Atlas (TCGA) database. Differentially expressed genes (DEGs) were identified, and functional enrichment analysis was applied. A protein-protein interaction (PPI) network was constructed to identify hub genes and receiver operating characteristic (ROC) analysis was used to evaluate candidate gene diagnostic values.

**Results:** RAS and BRAF mutations were the predominant mutation types in FVPTC. FVPTC was significantly correlated with the absence of extrathyroidal extension, lower N stage, and the low occurrence rate of BRAF mutation compared to classical PTC. Two thousand three hundred and forty-two FVPTC-related differentially expressed mRNAs (DEGs) and 420 FVPTC-specific DEGs were identified in this study. Function enrichment analysis revealed that these DEGs were involved in some pathways in cancer, including the PI3K-Akt signaling pathway and MAPK signaling pathways. The PPI network was constructed from 420 FVPTC-specific DEGs, and a sub-network, including 12 genes and 10 hub genes, was verified.

**Conclusions:** FVPTC was identified significantly relevant to remarkable alterations of gene mutation, DEGs, related pathways and the diagnostic performance of hub genes. Our study might provide further insights into the investigation of the tumorigenesis mechanism of FVPTC and assist in the discovery of new candidate diagnostic markers for FVPTC.

**Keywords:** Follicular variant papillary thyroid carcinoma (FVPTC); bioinformatics; protein-protein interaction network (PPI network); hub gene; receiver operating characteristic (ROC)

Submitted Jul 19, 2019. Accepted for publication Nov 01, 2019.

doi: [10.21037/tcr.2019.11.38](https://doi.org/10.21037/tcr.2019.11.38)

View this article at: <http://dx.doi.org/10.21037/tcr.2019.11.38>

## Introduction

Thyroid carcinoma is the most common endocrine malignancy, accounting for ~2.1% of all cancer diagnoses

worldwide, with ~77% of diagnoses occurrence in women, and the incidence has been increasing worldwide (1-4). The rate of new thyroid cancer cases has been raising by 3.8% on average each year for the last 10 years, while the

death rate gained an average of 0.7% each year between 2005–2015. The 5-year survival trend ranged from 92.3% to 98.6% (5). Differentiated thyroid carcinoma (DTC), including papillary thyroid carcinoma (PTC) and follicular thyroid carcinoma (FTC), account for 95% of thyroid cancer, and they have more favorable prognoses than poorly DTC (PDTC) and anaplastic thyroid cancer (ATC) (6). The most common variants of PTC include conventional PTC (classic/usual, cPTC), follicular variant PTC (FVPTC) and the tall cell variant PTC (TCPTC) (7).

FVPTC, one of the most frequent PTC variants (~30%), is a heterogeneous group of tumors that can be classified into two major groups, namely invasive/infiltrative FVPTC and noninvasive encapsulated FVPTC, which differ from each other in genes, clinic and morphology (8). The aggressive histopathologic features, such as thyroid capsule invasion, extrathyroidal extension and lymph node metastasis, are significantly less frequent in FVPTC than in cPTC, and the long-term outcome is either similar in both subtypes or more favorable in FVPTC compared with cPTC (9–11). Infiltrative FVPTC spreads to the lymph nodes, similar to cPTC, whereas encapsulated FVPTC behaves similar to the follicular adenoma (FA)/carcinoma group of tumors (12). Recently, the encapsulated/well demarcated non-invasive FVPTC was renamed “non-invasive follicular thyroid neoplasm with papillary-like nuclear feature” (NIFTP) by an international group of experts (12,13). FVPTC is less aggressive, so distinguishing between these two subtypes of PTC is necessary to avoid overtreatment and its complications. RNA sequencing provided a global view of the gene expression and enabled us to identify essential genes in the progression of thyroid carcinoma. Understanding the gene expression profiles of FVPTC as a whole will facilitate the research into the pathogenesis of this disease. However, the gene expression profiles, particularly the pathways in the interaction network and the diagnostic approaches of candidate markers of FVPTC, remain to be elucidated.

In this study, gene mutation and mRNA expression profiles of patients with FVPTC were obtained from The Cancer Genome Atlas (TCGA) database. Differentially expressed genes (DEGs) were identified, function enrichment and pathway analyses were performed, and then the protein-protein interaction (PPI) networks were constructed to identify the hub genes. We also performed receiver operating characteristic (ROC) analysis to evaluate the diagnostic values of the candidate genes. Our study may provide further insights into the investigation of the

tumorigenesis mechanism of FVPTC and assist in the discovery of new candidate diagnostic markers for FVPTC.

## Methods

### *TCGA data profiles*

The TCGA thyroid carcinoma (TCGA-THCA) dataset consisted of 507 cases of PTC and 59 normal tissue samples (14). Merged level-3 RNA sequencing data and clinical information, including the BRAF V600E mutation, were downloaded from the cBioPortal for Cancer Genomics (<http://www.cbioportal.org/index.do>) and Firebrowse (<http://firebrowse.org>). In total, 501 PTC cases and 59 normal samples with complete mRNA sequence and clinical data were identified. Among those 501 PTC cases, 357 cPTCs and 102 FVPTC ( $\geq 99\%$  follicular patterned) cases were selected for inclusion in this study (a total of 459 sample IDs are shown in *Tables S1,S2*). RNAseq by expectation-maximization (RSEM) values was used to quantify mRNA expression levels. This study is exempt from ethics, as all data is derived from multiple public databases and that personal information is not identifiable.

### *Screening of DEGs*

The differentially expressed mRNAs between the FVPTC and normal samples, between the cPTC and normal samples, and between the cPTC and FVPTC samples were analyzed with the Limma package (version 3.32.5) of R software (version 1.1.143). The function linear model fitting (lmFit) in the limma package was used to calculate fold changes (FC) and empirical Bayes statistics (eBayes) were used to estimate standard errors. Differentially expressed mRNAs with  $|\log_2FC| > 1.0$  and adjusted  $P < 0.05$  were considered to indicate a statistically significant difference. Gene expression values were converted to  $\log_{10}$  (RSEM) values using R. Hierarchical analysis of DEGs was achieved by Cluster 3.0 and then visualized via the Java TreeView 1.16r4 (<http://www.treeview.net>).

### *Functional enrichment analysis*

Candidate DEG functions and pathways enrichment were carried out using the database for annotation, visualization and integrated discovery (DAVID, <https://david.ncifcrf.gov/>) online system. Gene ontology analysis (GO, classification including biological process, cellular component and

molecular function) and Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment were performed. A P value <0.05 was set as the threshold.

### *PPI network analysis*

The PPI network was constructed using the search tool for the retrieval of interacting genes/proteins (STRING, <https://string-db.org>, minimum required interaction score was set at 0.4) database and visualized with Cytoscape software 3.5.1. Sub-network models were selected using the plugin molecular complex detection (MCODE) application in Cytoscape. The criteria for defining a sub-network were an MCODE score  $\geq 4$  and the number of nodes  $\geq 4$  as described (15). Function enrichment of genes in sub-networks was performed using STRING. Hub genes were explored using the CytoHubba plugin application (16) by topological analysis of maximal clique centrality (MCC, top 10 nodes ranked by MCC).

### *Statistical analysis*

Statistical analysis was carried out using SPSS 21.0 and R Studio (as described above). Chi-squared or Fisher's exact test was used to analyze the association between different histological diagnoses and clinicopathological parameters. Survival analysis was analyzed using Kaplan-Meier and log-rank tests. ROC curves and the area under the curve (AUC) were performed to test the specificity and sensitivity of the candidate genes in predicting diagnosis (17). Cut-off values were determined by calculating the maximized sum of specificity and sensitivity. A P value of <0.05 and AUC >0.7 was considered to be statistically significant.

## **Results**

### *Somatic mutation of FVPTCs and cPTCs*

Gene mutation frequencies in the patients with PTC were obtained using cBioPortal. Furthermore, subgroup gene mutation studies of FVPTC and cPTC were queried by selecting relevant case IDs through the cBioPortal website. In total, altered genes (a mutation in more than 2 cases) happened in 78 of 102 sequenced FVPTC cases (Figure S1). The top 10 gene mutation profiles are shown in Figure 1. Alteration of RAS was particularly prevalent in FVPTC with a mutation frequency of 31% (including NRAS, HRAS, and KRAS mutations, the frequencies were

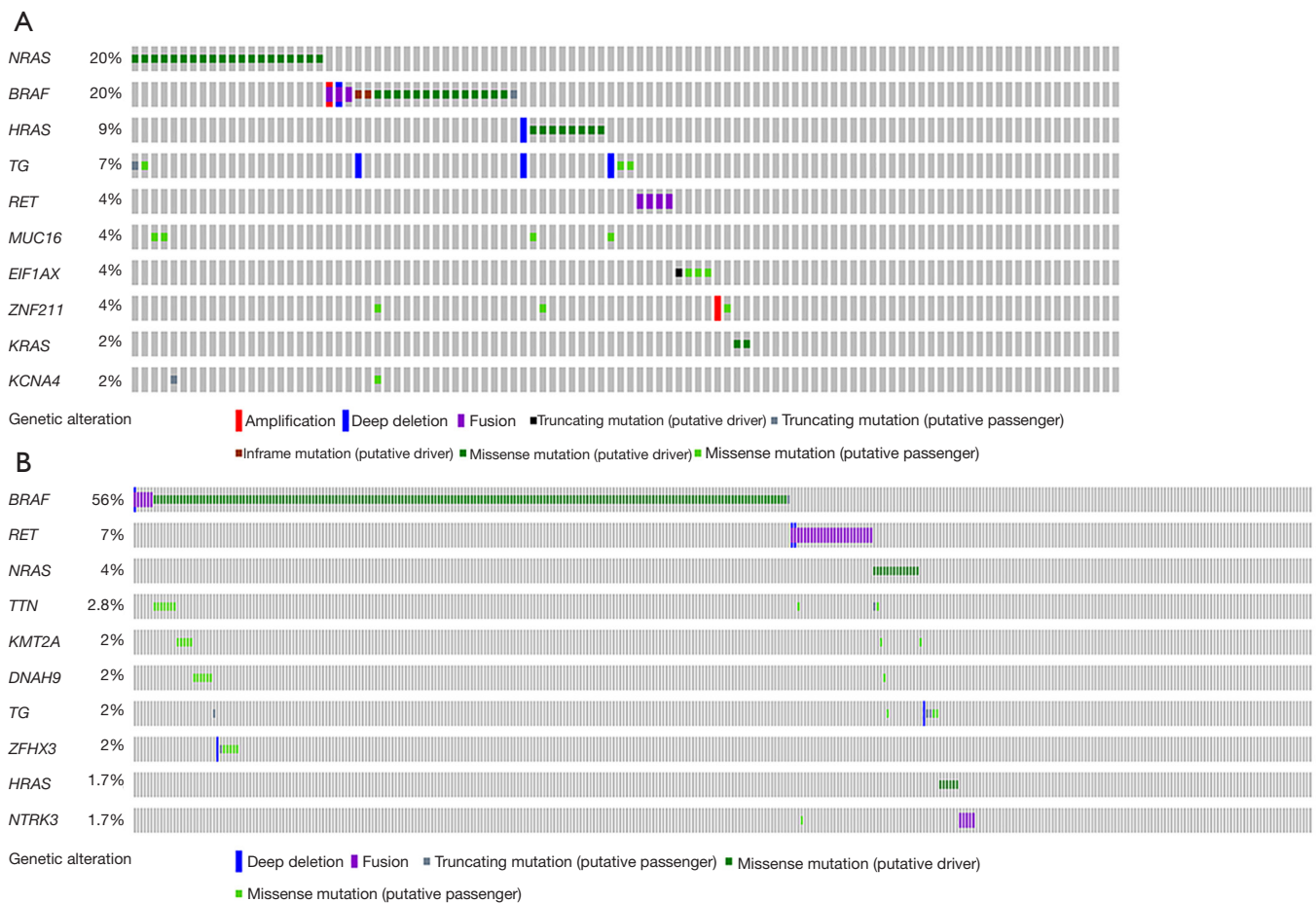
20%, 9%, and 2%, respectively). The BRAF mutation also took place at a high frequency (20%) in FVPTC. Distinguished with the FVPTC gene mutation pattern, the BRAF mutation was the predominant mutation type in cPTC (it occurred in 56% of cPTC patients), and attended by RET fusion (with a relatively low occurrence rate, at 9%). Notably, the incident rate of RAS mutation (including NRAS, HRAS, and KRAS mutations, the frequencies were 4%, 1.7%, and 0.71%, respectively) was much lower in the cPTC patients compared with the FVPTC patients.

### *FVPTC seems to associate with favorable clinicopathological parameters*

The correlation of two histological diagnoses with clinicopathological parameters was studied, to address the clinical differences of FVPTC and cPTC. The results showed that FVPTC was significantly correlated with the absence of extrathyroidal extension, lower N stage, and the low occurrence rate of BRAF mutation, while no significance was observed between patient gender, age, tumor focus type, T stage, and AJCC TNM stage (Table 1). Although a significant correlation ( $P=0.004$ ) was observed between FVPTC and M1 stage, all of the PTC patients with distant metastasis were at a very low rate (9/507). Additionally, Kaplan-Meier analysis revealed no disease-free or overall survival rate difference between the patients with FVPTC or cPTC. These results suggested that FVPTC patients seem to present favorable clinicopathological features compared with the cPTC patients. Moreover, the FVPTC patients with a BRAF mutation significantly correlated with older age, extrathyroidal extension, and advanced TNM stage (Table S3).

### *Screening FVPTC-specific differentially expressed mRNAs*

Differential analysis revealed 2,342 DEGs (1,695 downregulated and 647 upregulated genes) between the FVPTC and normal tissues (NT); 3,103 DEGs (1,723 downregulated and 1,380 upregulated genes) between the cPTC and NT with  $|\log_{2}FC| > 1.0$  and adjusted  $P < 0.05$ . DEGs between the FVPTC and NT were defined as FVPTC-related DEGs. Among these FVPTC-related DEGs, SLIT1, TMEM215, KLHDC8A, HAPLN1, and ZCCHC12 were the most significantly upregulated DEGs, while CCL21, DPT, MYOC, SCARA5, and RELN were the most significantly downregulated DEGs. Moreover, 1,625 DEGs were identified between FVPTC



**Figure 1** OncoPrints of mutated genes in FVPTCs and cPTCs. (A) Top 10 mutated genes in patients with FVPTC. RAS (including HRAS, NRAS, and KRAS) mutation and BRAF mutation occurred most frequently; (B) top 10 mutated genes in patients with cPTC; BRAF mutation was the predominant mutation type whereas RAS mutation occurred less often. OncoPrints were generated by the online cBioPortal system (updated to November 2017). FVPTC, follicular variant papillary thyroid carcinoma; cPTC, conventional papillary thyroid carcinoma.

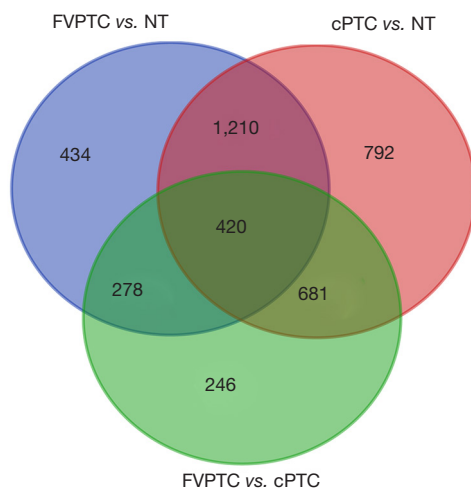
and cPTC with the same threshold, of which 1,191 were downregulated and 434 were upregulated. DEGs between FVPTC and cPTC that also belong to PTC-related DEGs (genes deregulated in both FVPTC and cPTC compared with NT) were defined as FVPTC-specific differentiated expressed genes (FVPTC-specific DEGs). According to this protocol, 420 FVPTC-specific DEGs, including 260 downregulated and 160 upregulated genes, were identified between the FVPTC and cPTC tissues (Figures 2,3). The average expression values in PTC tissues (including FVPTC and cPTC tissues); FC and P value details of these 420 deregulated genes are shown in Table S4. ZMAT4, SLC5A8, DIO1, MT1H, and CUX2 were the most significantly

upregulated DEGs, while TMPRSS6, SFTPB, SYT12, SLC27A6, and TMPRSS4 were the most significantly downregulated DEGs in the FVPTC tissues compared with the cPTC tissues. Interestingly, the expression level of most oncogenes (PTC-related upregulated DEGs identified in our study) in the FVPTC tissues were much lower than in the cPTC tissues, while the expression level of most tumor suppressors (PTC-related downregulated DEGs) were higher in the FVPTC tissues than in the cPTC tissues (Table S4). These results suggested that tumor-related genes are less dysregulated in FVPTC compared with cPTC, which may partly explain why FVPTC is less aggressive than cPTC.

**Table 1** Clinicopathological features of cPTC and FVPTC patients in the TCGA dataset

Variables	No. of patients	cPTC	FVPTC	P value
Gender				0.429
Male	122	98	24	
Female	337	259	78	
Age				0.396
≤45	224	178	46	
>45	235	179	56	
Multifocality				0.198
Unifocal	253	202	51	
Multifocal	198	148	50	
Extrathyroidal extension				<0.001*
None	319	230	89	
Minimal (T3) + moderate/advanced (T4a)	123	111	12	
T stage				0.163
T1 + T2	296	224	72	
T3 + T4	161	131	30	
N stage				<0.001*
N0	213	148	65	
N1	197	184	13	
M stage				0.004**
M0	253	221	32	
M1	9	4	5	
AJCC TNM stage				0.059
Stage I + II	318	240	78	
Stage III + IV	139	116	23	
BRAF mutation				<0.001*
Without mutation	239	157	82	
BRAF V600E	206	192	14	
Disease free status				0.363 <sup>a</sup>
Disease free	406	311	95	
Recurred/progressed	39	33	6	
Overall survival status				0.199 <sup>a</sup>
Living	443	342	101	
Deceased	16	15	1	

Patients with censored data were excluded. \*, P<0.05; \*\*, P value by Fisher's exact test; <sup>a</sup>, P value by log-rank test. FVPTC, follicular variant papillary thyroid carcinoma; cPTC, conventional papillary thyroid carcinoma; TCGA, The Cancer Genome Atlas.



**Figure 2** Identification of 420 FVPTC-specific DEGs. There were 420 overlapping DEGs from the three group comparisons (FVPTV *vs.* NT, cPTV *vs.* NT and FVPTV *vs.* cPTC tissues). FVPTC, follicular variant papillary thyroid carcinoma; DEG, differentially expressed gene; cPTC, conventional papillary thyroid carcinoma; NT, normal tissues.

### GO and KEGG pathway enrichment analysis of FVPTC-related and -specific DEGs

In order to gain a better understanding of the gene function and signaling pathways of FVPTC-related and -specific DEGs, online GO and KEGG pathway enrichment analysis were carried out using the DAVID system. For FVPTC-related DEGs, the top 5 enriched GO terms for biological process were cell adhesion, inflammatory response, immune response, cell-cell signaling, and signal transduction. The top 5 enriched GO terms for cellular component were integral component of plasma membrane, extracellular space, extracellular region, plasma membrane, and proteinaceous extracellular matrix. For molecular function heparin binding, calcium ion binding, chemokine activity, receptor activity, and Ras guanyl-nucleotide exchange factor activity were considerably enriched (Figure 4A). Furthermore, the enriched KEGG pathways of FVPTC-related DEGs included pathways in cancer, cytokine-cytokine receptor interactions, neuroactive ligand-receptor interactions, the PI3K-Akt signaling pathway, the MAPK signaling pathway, and the cell adhesion pathway (ranked by enriched gene number, Figure 4B).

For FVPTC-specific DEGs, we found that the top 5 enriched GO terms for biological process were response to proteolysis, positive regulation of gene expression, keratan

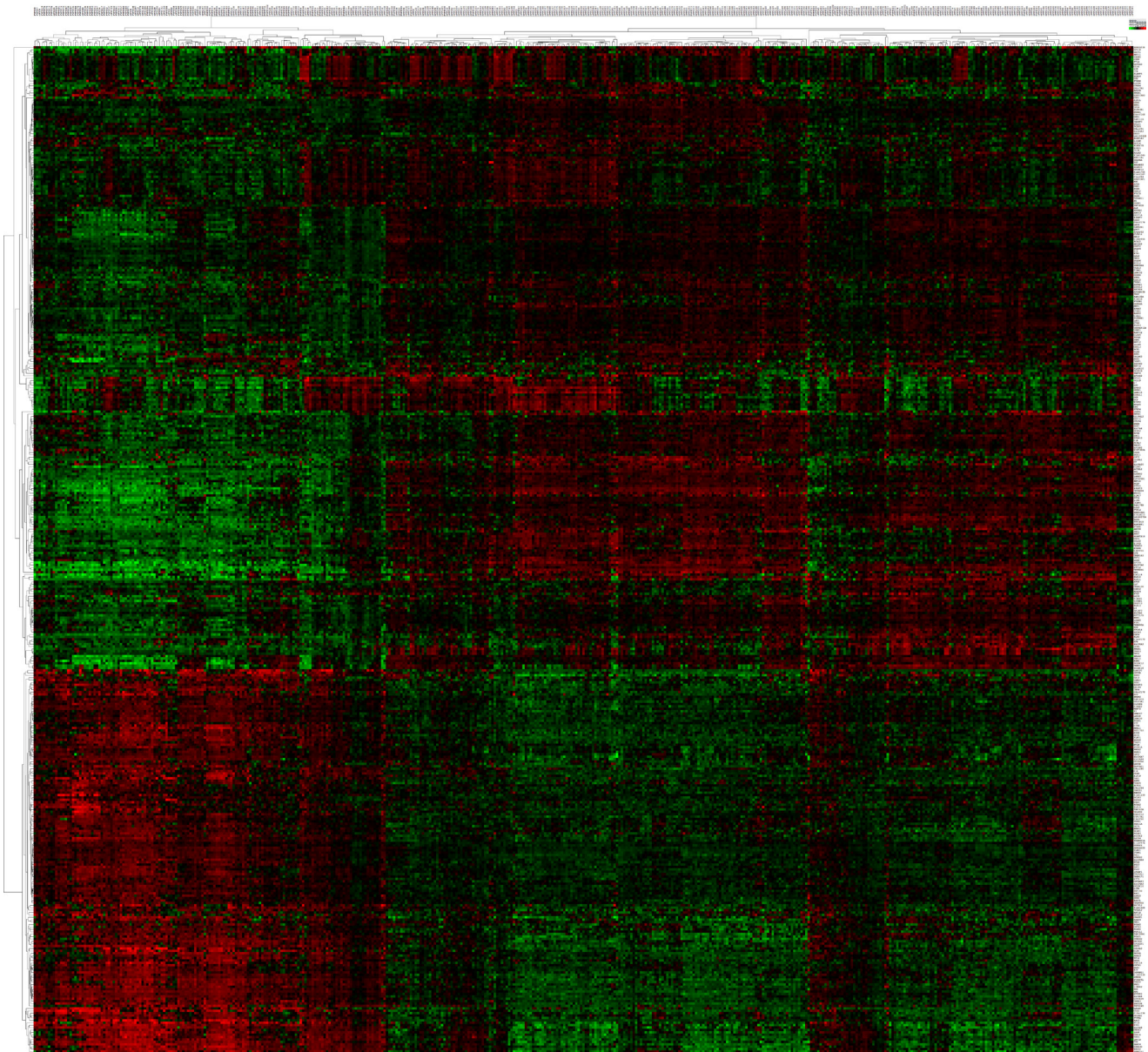
sulfate biosynthetic process, signal transduction, and cell adhesion. The top 5 enriched GO terms for cellular components were extracellular region, extracellular space, plasma membrane, proteinaceous extracellular matrix, and integral component of the plasma membrane. For molecular function, serine-type endopeptidase activity, serine-type peptidase activity, calcium ion binding, protein kinase inhibitor activity, and protease binding were significantly enriched (Figure 4C). Furthermore, those FVPTC-specific DEGs were significantly enriched in pathways in complement and coagulation cascades, tyrosine metabolism, cytokine-cytokine receptor interaction, cell adhesion molecules, and the chemokine signaling pathway (Figure 4D).

### PPI network construction and Hub gene selection

In order to find the FVPTC-specific biomarkers, the PPI network was constructed only on 420 FVPTC-specific DEGs. By uploading the PPI network into the Cytoscape software, we found 228 nodules and 534 edges in the DEG network (Figure 5A). A total of 5 sub-networks were found using the definition criteria. The top sub-network, which included 12 nodules and 66 edges, was selected, and GO and KEGG pathway analyses were performed. Genes annotated in this sub-network included PYY, ADCY8, LPAR5, ADORA1, GRM4, CCL19, SSTR3, NPW, CCL21, CXCR5, C3, and NMU. The most significant enrichment terms were the G-protein coupled receptor signaling pathway, CCR7 chemokine receptor binding in GO, and the chemokine signaling pathway in KEGG (Figure 5B). The top 10 hub genes were annotated as LPAR5, NMU, CCL19, CCL21, C3, CXCR5, ADCY8, PYY, GRM4, and ADORA1 (ranked by MCC scores). All of these hub genes were included in the top sub-network. All of these results suggested that these hub genes might play an important role in the development of FVPTC.

### Diagnostic performance of hub genes

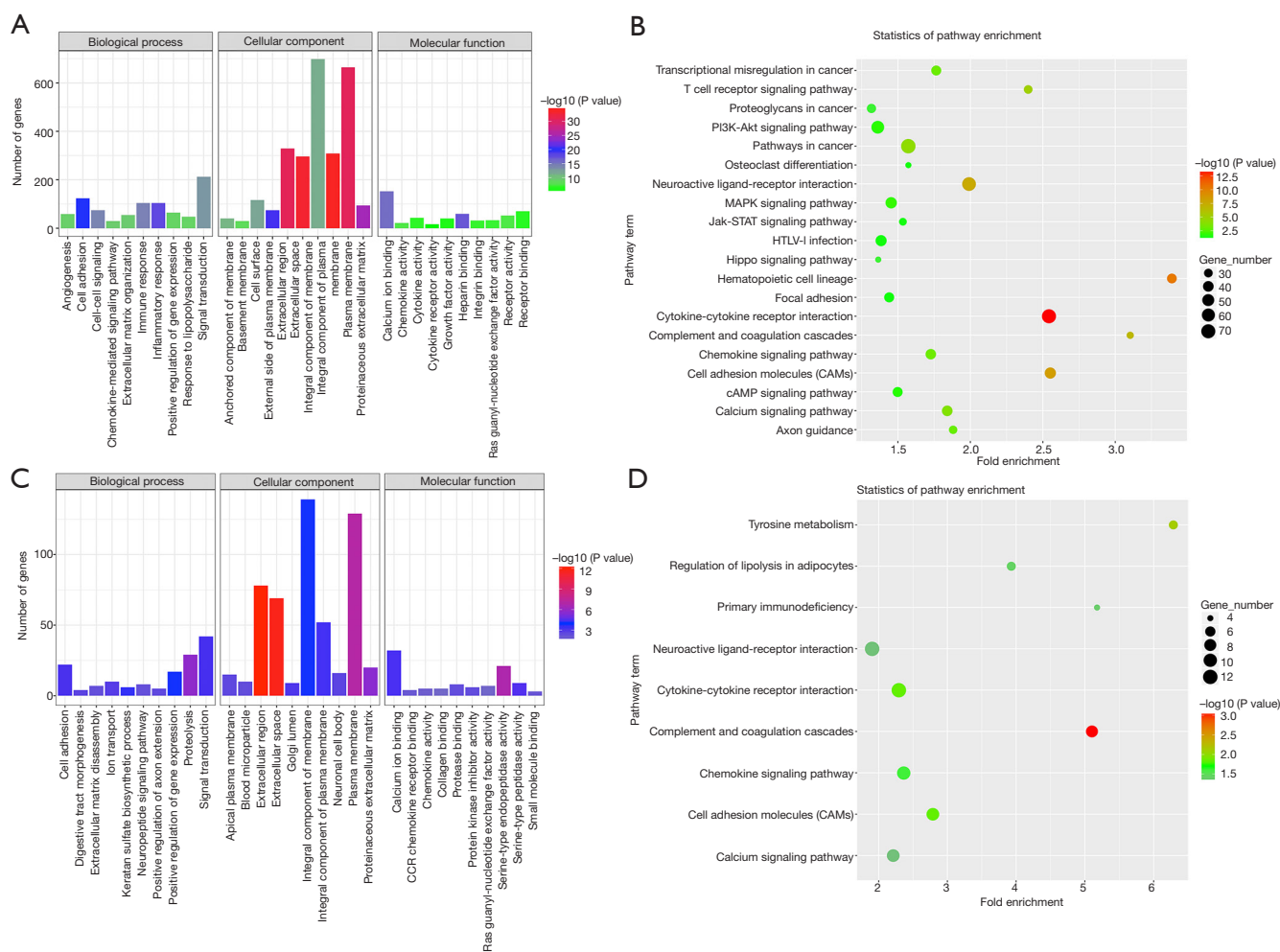
The diagnostic performance of the hub genes in differentiating FVPTC and cPTC from NT, and in differentiating FVPTC from cPTC (three independent groups) was evaluated by ROC analysis. To identify the discriminatory ability of these candidate genes in each group, AUC >0.7 was fixed as the threshold. First, we accessed the discriminatory ability of the hub genes between the FVPTC and NT. The upregulated genes ADCY8,



**Figure 3** Heatmap of the top 420 FVPTC-specific DEGs. Hierarchical analysis of the 420 FVPTC-specific DEGs based on the expression values in FVPTC and cPTC tissues [all of the values are presented with  $\log_{10}$  (RSEM), P value <0.05]. FVPTC, follicular variant papillary thyroid carcinoma; DEG, differentially expressed gene; cPTC, conventional papillary thyroid carcinoma; RSEM, RNAseq by expectation-maximization.

ADORA1, and LPAR5 (*Figure 6A*) and the downregulated genes C3, CCL19, CCL21, CXCR5, and PYY (*Figure 6B*) were selected as the candidate genes with an AUC >0.7. Then, the discriminatory ability of the hub genes between the cPTC and NT was assessed. There were 8 hub genes, excluding CCL19 and CXCR5, possessed discriminatory

abilities (*Figure 6C,D*). Additionally, C3 and ADORA1 exhibited the best prediction power with the highest AUCs (1 and 0.964, respectively). Finally, ROC analysis showed that only the downregulated genes (the genes downregulated in the FVPTCs compared with the cPTCs, which included ADCY8, ADORA1, C3, GRM4, LPAR5,



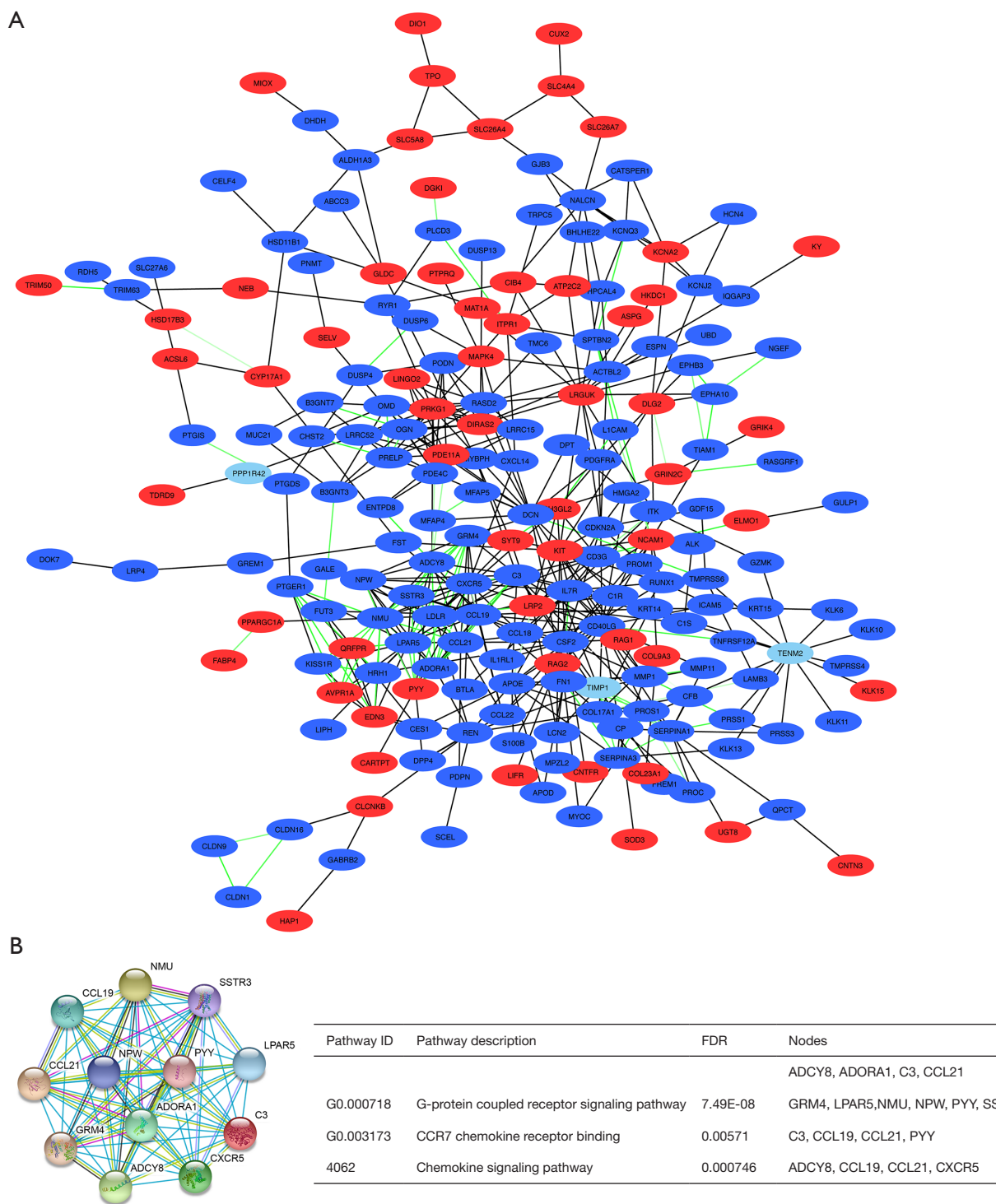
**Figure 4** GO and KEGG pathway enrichment analysis of FVPTC DEGs. (A) GO annotation for FVPTC-related DEGs. GO biological process, cellular component and molecular function terms; the number of enriched genes and  $-\log_{10}(P \text{ value})$  are presented in the diagram; (B) KEGG pathway enrichment analysis of FVPTC-related DEGs. Pathway terms,  $-\log_{10}(P \text{ value})$ , enriched gene numbers and fold enrichment score are presented in the diagram; (C) GO annotation for FVPTC-specific DEGs; (D) KEGG pathway enrichment analysis of FVPTC-specific DEGs. GO, gene ontology; KEGG, Kyoto encyclopedia of genes and genomes; FVPTC, follicular variant papillary thyroid carcinoma; DEG, differentially expressed gene.

and NUM) could robustly distinguish FVPTC from cPTC (Figure 6E). Furthermore, ADCY8, ADORA1, C3, and LPAR5 were identified in all of the three independent groups. The cut-off values and AUCs of these four genes for distinguishing the three independent groups are presented in Table 2. Collectively, our results suggest that combinatorial ADCY8, ADORA1, C3, and LPAR5 could not only be used for the differential diagnosis of PTC, including FVPTC and cPTC, from normal samples, but also for the differential diagnosis of FVPTC from cPTC samples.

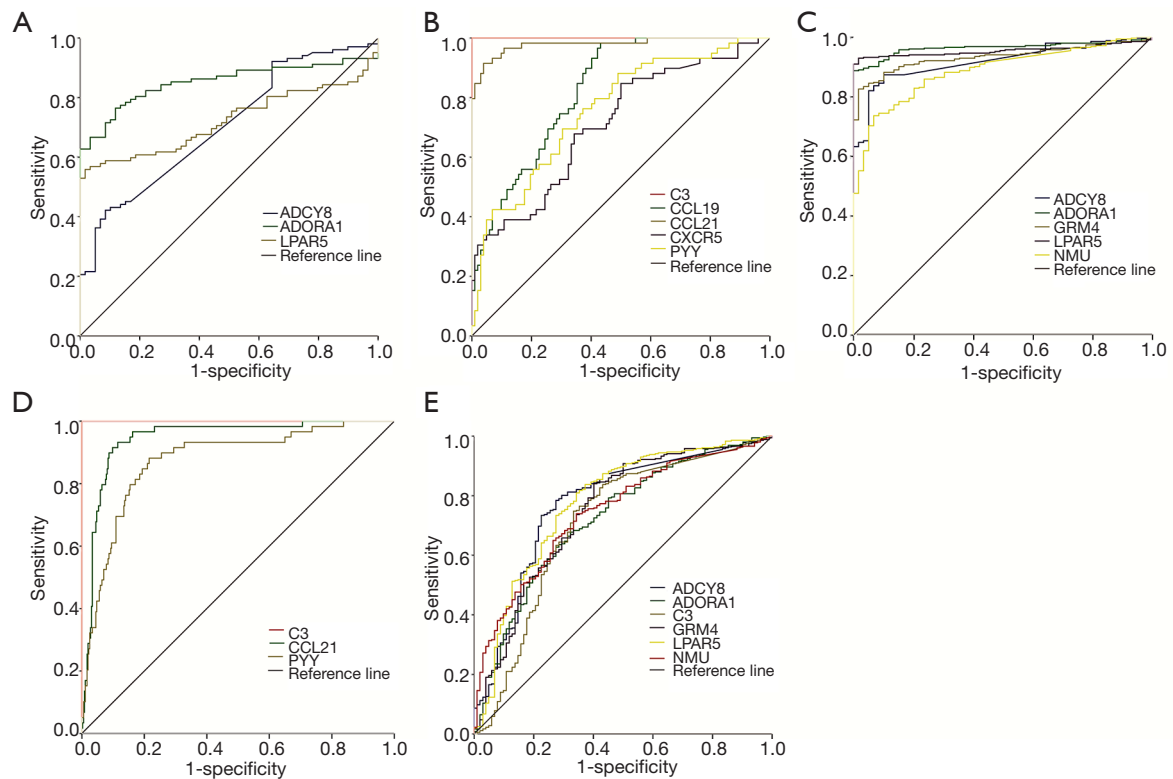
## Discussion

The gene mutation pattern and clinical characteristics of FVPTC were studied in the present study. Multi-group analyses were performed to identify aberrantly expressed genes in FVPTC. The FVPTC-related DEGs and FVPTC-specific DEGs were obtained from the TCGA-THCA dataset, and we proceeded with functional and pathway enrichment analysis. We highlighted the diagnostic values of several hub genes essentially, not only in distinguishing FVPTC and cPTC from NT, but also in





**Figure 5** PPI network of FVPTC-specific DEGs. (A) Network of 420 FVPTC-specific DEGs included 228 nodules and 534 edges (red nodules represent the upregulated genes and blue nodules represent the downregulated genes). Green edges indicate a combined score of interactive genes >0.9; (B) the top sub-network included 12 nodules and 66 edges. The top GO and KEGG terms of the enrichment results of these genes are shown in the right form. PPI, protein-protein interaction; FVPTC, follicular variant papillary thyroid carcinoma; DEG, differentially expressed gene; GO, gene ontology; KEGG, Kyoto encyclopedia of genes and genomes; FDR, false discovery rate.



**Figure 6** ROC curves for diagnostic performance of hub genes. Diagnostic performance of upregulated (A) and downregulated (B) hub genes in differentiating FVPTC from NT; diagnostic performance of upregulated (C) and downregulated (D) hub genes in differentiating cPTC from NT; (E) ROC analysis shows that only six downregulated genes could distinguish FVPTC from cPTC. ROC, receiver operating characteristic; FVPTC, follicular variant papillary thyroid carcinoma; cPTC, conventional papillary thyroid carcinoma; NT, normal tissues.

**Table 2** Cut-off values and AUCs of four candidate genes for distinguishing the three independent groups

Candidate genes	Cut-off values (log <sub>2</sub> RSEM)			AUC (sensitivity & specificity)		
	FVPTC vs. NT	cPTC vs. NT	FVPTC vs. cPTC	FVPTC vs. NT	cPTC vs. NT	FVPTC vs. cPTC
<i>ADCY8</i>	>0.631	>0.317	<2.421	0.704 (0.422, 0.915)	0.920 (0.874, 0.898)	0.774 (0.784, 0.725)
<i>ADORA1</i>	>7.740	>8.404	<9.246	0.852 (0.765, 0.881)	0.964 (0.888, 1.000)	0.724 (0.678, 0.676)
<i>C3</i>	<7.767	<7.230	<2.286	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	0.710 (0.748, 0.667)
<i>LPAR5</i>	>8.325	>8.263	<9.579	0.722 (0.559, 0.983)	0.955 (0.930, 0.983)	0.778 (0.810, 0.657)

NT, normal tissues; AUC, area under the curve; FVPTC, follicular variant papillary thyroid carcinoma; cPTC, conventional papillary thyroid carcinoma; RSEM, RNAseq by expectation-maximization.

discriminating FVPTCs from cPTCs. Through the TCGA-THCA data analysis, Stokowy *et al.* revealed molecular differences, including microRNA and mRNA, between FVPTC and cPTC. Putative miR-152-3p with the TGFA regulatory pair was discovered in FVPTC (18). Saiselet *et al.* illustrated that several downregulated miRNAs could distinguish histological classical variants and FVPTC in

the TCGA dataset (19). To our knowledge, this is the first comprehensive study on gene mutation, screening FVPTC-specific DEGs, and the diagnostic approach for candidate biomarkers of FVPTC by utilizing TCGA-THCA data.

We found that more than half of the FVPTCs harbored oncogenic mutations, either the RAS (31%) or BRAF (20%) mutation. According to other reports,

BRAF and RAS were the predominant mutations and mutually exclusive in all of the FVPTC cases (20). RAS mutations was demonstrated to get absolute predominance in encapsulated FVPTCs compared to infiltrative FVPTV (21). The incidence of the BRAF V600E mutation in encapsulated FVPTCs was relatively low, whereas an equal incidence of BRAF V600E or RAS mutation was observed in infiltrative FVPTCs (21,22). Yoo *et al.* revealed that the mutational profile of EFVPTC was similar to FA, while that of infiltrative FVPTC was similar to cPTC (23). We revealed that FVPTC seems to associate with favorable clinicopathological parameters compared with cPTCs, while FVPTC patients with the BRAF mutation significantly correlated with older age, extrathyroidal extension, and advanced TNM stage. Although the TCGA-THCA dataset did not provide the subgroup details of FVPTC, the FVPTC patients with the BRAF mutation appear to belong to a group with more aggressive behavior. Despite differences in the cytological classification and molecular profiles between noninvasive and infiltrative FVPTC, preoperative cytologic examination was indistinguishable between these two subgroups (21,22). Hence, diagnostic facilitation using molecular profiles combined with fine-needle aspirates (FNA) cytology should be evaluated by further studies with larger patient cohorts.

Slit guidance ligand 1 (SLIT1) belongs to the SLIT family of secreted proteins that mediate positional interactions between cells and their environment during development by signaling from ROBO receptors. SLITs and ROBOs are considered candidate tumor suppressor genes because their promoters are frequently hypermethylated in epithelial cancer (24). Hypermethylation of promotor region triggered SLIT1 downregulation, which has been reported in murine mammary glands or human breast carcinoma cells (24), glioma tumors (25), and early and advanced gastric cancers (26). PML knockdown led to impaired cell migration in glioblastoma, and the cell migration reduction was demonstrated to be associated with a marked upregulation of SLIT1 mRNA and protein levels, and of ROBO2. Furthermore, the PML/SLIT1 axis controls the sensitivity of glioblastoma cells to arsenic trioxide (27). There are only a few literature reviews the SLIT1 expression in thyroid normal and cancer tissues. Other than the downregulated expression pattern in other types of cancers, we found that SLIT1 was the top upregulated gene in FVPTC compared with the NT. Based on the above, the mechanism and function of upregulated SLIT1 in FVPTC need to be elucidated urgently. The C-C

motif Chemokine ligand 21 (CCL21)/CCR7 axis promoted metastasis and survival of cancer cells by modulating EMT and cancer related pathways, and CCL21 correlated with tumor metastasis and prognosis (28,29). Lu *et al.* disclosed that CCL21 could facilitate chemoresistance and the stem cell property of colorectal cancer cells through AKT/GSK-3 $\beta$ /Snail signals (30). Zhang *et al.* disclosed that exogenous 100 ng/mL CCL21 markedly promoted cell proliferation, and CCL21/CCR7 interaction augmented the proportion of cells in G2/M via enhancing the p-ERK expression level (31). CCL21 was significantly lower in Ewing sarcoma patients with metastases than in patients without metastases. A higher RNA expression level of CCL21 was apparently corrected with good chemotherapeutic response and improved outcome (32). CCR7 was highly expressed in PTC (31), whereas the CCL21 expression level in PTC and the function of dysregulation of CCL21 should be elucidated. Hyaluronan- and proteoglycan-linked protein 1 (HAPLN1) showed essential effect on tumor proliferation, invasion, and metastasis. HAPLN1 is 23-fold overexpressed in stage 1 mesothelioma (33) and highly expressed in hepatocellular carcinomas (34) and oral cancer tissues (35). Zinc finger CCHC-type containing 12 (ZCCHC12) has important biological functions and acts as a metastasis-related oncogene in PTC (36). In our study, these cancer-related genes were also found to be dysregulated in FVPTCs; hence their function and mechanism in FVPTC need to be further investigated. Furthermore, the enriched KEGG pathways of FVPTC-related DEGs with the most genes were the pathways in cancer, the PI3K-Akt signaling pathway, the MAPK signaling pathway, and the cell adhesion pathway. Consistent with the results of other studies, the activation of the MAPK and PI3K pathways was observed in FVPTC (37). These data support that FVPTC shares the same pathway abnormalities with cPTC, although it has a different gene mutation pattern. Furthermore, the mild abnormality of molecular function, biological process, and related pathways involved in 420 FVPTC-specific DEGs may be one of the reasons for the lower malignant degree of FVPTC.

FNA cytology has the inherent limitation that indeterminate cytology results cannot distinguish between FA, FTC, or FVPTC. Molecular FNA diagnostics (combined FNA with molecular diagnostic approaches) is an essential complementary addition to FNA cytology (38). Previous studies have described the identification of biomarkers to improve the differential diagnosis of FVPTC from cPTC and benign lesions.

Liu *et al.* identified that HBME-1 was profitable in the differentiation between FA and FVPTC (39). KAP-1, which was presented in 72% of FVPTC cases, could differentiate nodular hyperplastic lesions from both cPTC and FVPTC, while it was unable to distinguish FVPTC from FA and FTC (40). Hoftijzer *et al.* used a tissue array containing benign thyroid and thyroid carcinoma (including FVPTC) tissues to evaluate the diagnostic accuracy of RAR and RXR subtype protein expression for the differential diagnosis of thyroid neoplasms. In comparison between FA and FVPTC, nRAR, nRARA, cRXRB and mRXRB had sensitivities and specificities above 70%, and the highest specificity was for nRAR (91%) (41). Our study is the first to reveal that several biomarkers showed excellent predictive performances both in the differential diagnosis of PTC (including FVPTC and cPTC) from normal samples and in the differential diagnosis of FVPTC from cPTC samples. Future clinical studies with more cost-effective approaches, such as RT-PCR and immunohistochemical staining, instead of RNA sequencing should be conducted on a larger cohort of FVPTC samples to validate the clinical utility of the individual or combined hub genes identified in this study.

## Conclusions

Our study identified the remarkable alterations of gene mutation, DEGs and related pathways in FVPTC. We demonstrated that the dysregulation of cancer-related genes might lead to diverse behavior in FVPTCs, even though the FVPTC shared the same pathway abnormalities with cPTC. In this study, we demonstrated 420 FVPTC-specific DEGs may be one of the reasons for the lower malignant degree of FVPTC. The diagnostic performance of hub genes may provide a novel mechanism in FVPTC and additional diagnostic targets of FVPTC. Differential diagnostic facilitation of FVPTC from cPTC and of FVPTC subgroups using molecular profiles should be evaluated by further studied with larger patients' cohorts. More work is needed to elucidate the functional mechanism of these genes and to validate the clinical utility of these biomarkers with more cost-effective approaches.

## Acknowledgments

*Funding:* This work was funded by the National Natural Science Foundation of China (NSFC) (81672885).

## Footnote

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/tcr.2019.11.38>). The authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

1. Davies L, Welch HG. Current thyroid cancer trends in the United States. *JAMA Otolaryngol Head Neck Surg* 2014;140:317-22.
2. Shaha AR, Tuttle RM. Thyroid cancer staging and genomics. *Ann Transl Med* 2019;7:S49.
3. Li M, Zhao B, Qu W, et al. Uncovering the potential miRNAs and mRNAs in follicular variant of papillary thyroid carcinoma in the Cancer Genome Atlas database. *Transl Cancer Res* 2019;8:1158-69.
4. Ferlay J, Soerjomataram I, Ervik M, et al. GLOBOCAN 2012 v1.0, cancer incidence and mortality worldwide: IARC CancerBase No. 11. Lyon: International Agency for Research on Cancer, 2013.
5. SEER. Cancer stat facts: thyroid cancer. Available online: <http://seer.cancer.gov/statfacts/html/thyro.html>
6. Vuong HG, Kondo T, Duong UNP, et al. Prognostic impact of vascular invasion in differentiated thyroid carcinoma: a systematic review and meta-analysis. *Eur J Endocrinol* 2017;177:207-16.
7. Lloyd RV, Buehler D, Khanafshar E. Papillary thyroid carcinoma variants. *Head Neck Pathol* 2011;5:51-6.
8. Xu B, Ghossein R. Evolution of the histologic classification of thyroid neoplasms and its impact on

- clinical management. *Eur J Surg Oncol* 2018;44:338-47.
9. Tunca F, Sormaz IC, Iscan Y, et al. Comparison of histopathological features and prognosis of classical and follicular variant papillary thyroid carcinoma. *J Endocrinol Invest* 2015;38:1327-34.
  10. Shi X, Liu R, Basolo F, et al. Differential clinicopathological risk and prognosis of major papillary thyroid cancer variants. *J Clin Endocrinol Metab* 2016;101:264-74.
  11. Kim SK, Park I, Woo JW, et al. Follicular and diffuse sclerosing variant papillary thyroid carcinomas as independent predictive factors of loco-regional recurrence: a comparison study using propensity score matching. *Thyroid* 2016;26:1077-84.
  12. Tallini G, Tuttle RM, Ghossein RA. The history of the follicular variant of papillary thyroid carcinoma. *J Clin Endocrinol Metab* 2017;102:15-22.
  13. Nikiforov YE, Seethala RR, Tallini G, et al. Nomenclature revision for encapsulated follicular variant of papillary thyroid carcinoma: a paradigm shift to reduce overtreatment of indolent tumors. *JAMA Oncol* 2016;2:1023-9.
  14. Cancer Genome Atlas Research Network. Integrated genomic characterization of papillary thyroid carcinoma. *Cell* 2014;159:676-90.
  15. Li H, Liu JW, Liu S, et al. Bioinformatics-based identification of methylated-differentially expressed genes and related pathways in gastric cancer. *Dig Dis Sci* 2017;62:3029-39.
  16. Chin CH, Chen SH, Wu HH, et al. CytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* 2014;8 Suppl 4:S11.
  17. Li J, Fine JP. ROC analysis with multiple classes and multiple tests: methodology and its application in microarray studies. *Biostatistics* 2008;9:566-76.
  18. Stokowy T, Gawel D, Wojtas B. Differences in miRNA and mRNA profile of papillary thyroid cancer variants. *Int J Endocrinol* 2016;2016:1427042.
  19. Saiselet M, Gacquer D, Spinette A, et al. New global analysis of the microRNA transcriptome of primary tumors and lymph node metastases of papillary thyroid cancer. *BMC Genomics* 2015;16:828.
  20. McFadden DG, Dias-Santagata D, Sadow PM, et al. Identification of oncogenic mutations and gene fusions in the follicular variant of papillary thyroid carcinoma. *J Clin Endocrinol Metab* 2014;99:E2457-62.
  21. Kim TH, Lee M, Kwon AY, et al. Molecular genotyping of the non-invasive encapsulated follicular variant of papillary thyroid carcinoma. *Histopathology* 2018;72:648-61.
  22. Zhao L, Dias-Santagata D, Sadow PM, et al. Cytological, molecular, and clinical features of noninvasive follicular thyroid neoplasm with papillary-like nuclear features versus invasive forms of follicular variant of papillary thyroid carcinoma. *Cancer Cytopathol* 2017;125:323-31.
  23. Yoo SK, Lee S, Kim SJ, et al. Comprehensive analysis of the transcriptional and mutational landscape of follicular and papillary thyroid cancers. *PLoS Genet* 2016;12:e1006239.
  24. Marlow R, Strickland P, Lee JS, et al. SLITs suppress tumor growth in vivo by silencing Sdf1/Cxcr4 within breast epithelium. *Cancer Res* 2008;68:7819-27.
  25. Dickinson RE, Dallol A, Bieche I, et al. Epigenetic inactivation of SLIT3 and SLIT1 genes in human cancers. *Br J Cancer* 2004;91:2071-8.
  26. Kim M, Kim JH, Baek SJ, et al. Specific expression and methylation of SLIT1, SLIT2, SLIT3, and miR-218 in gastric cancer subtypes. *Int J Oncol* 2016;48:2497-507.
  27. Amodeo V, Deli A, Betts J, et al. A PML/slit axis controls physiological cell migration and cancer invasion in the CNS. *Cell Rep* 2017;20:411-26.
  28. Zhang L, Wang D, Li Y, et al. CCL21/CCR7 axis contributed to CD133+ pancreatic cancer stem-like cell metastasis via EMT and Erk/NF- $\kappa$ B pathway. *PLoS One* 2016;11:e0158529.
  29. Xiong Y, Huang F, Li X, et al. CCL21/CCR7 interaction promotes cellular migration and invasion via modulation of the MEK/ERK1/2 signaling pathway and correlates with lymphatic metastatic spread and poor prognosis in urinary bladder cancer. *Int J Oncol* 2017;51:75-90.
  30. Lu LL, Chen XH, Zhang G, et al. CCL21 facilitates chemoresistance and cancer stem cell-like properties of colorectal cancer cells through AKT/GSK-3 $\beta$ /snail signals. *Oxid Med Cell Longev* 2016;2016:5874127.
  31. Zhang YY, Liu ZB, Ye XG, et al. Iodine regulates G2/M progression induced by CCL21/CCR7 interaction in primary cultures of papillary thyroid cancer cells with RET/PTC expression. *Mol Med Rep* 2016;14:3941-6.
  32. Sand LG, Berghuis D, Szuhai K, et al. Expression of CCL21 in Ewing sarcoma shows an inverse correlation with metastases and is a candidate target for immunotherapy. *Cancer Immunol Immunother* 2016;65:995-1002.
  33. Ivanova AV, Goparaju CM, Ivanov SV, et al. Protumorigenic role of HAPLN1 and its IgV domain in malignant pleural mesothelioma. *Clin Cancer Res* 2009;15:2602-11.

34. Mebarki S, Désert R, Sulpice L, et al. De novo HAPLN1 expression hallmarks Wnt-induced stem cell and fibrogenic networks leading to aggressive human hepatocellular carcinomas. *Oncotarget* 2016;7:39026-43.
35. Zhuang Z, Jian P, Longjiang L, et al. Oral cancer cells with different potential of lymphatic metastasis displayed distinct biologic behaviors and gene expression profiles. *J Oral Pathol Med* 2010;39:168-75.
36. Wang O, Zheng Z, Wang Q, et al. ZCCHC12, a novel oncogene in papillary thyroid cancer. *J Cancer Res Clin Oncol* 2017;143:1679-86.
37. Santarpia L, Myers JN, Sherman SI, et al. Genetic alterations in the RAS/RAF/mitogen-activated protein kinase and phosphatidylinositol 3-kinase/Akt signaling pathways in the follicular variant of papillary thyroid carcinoma. *Cancer* 2010;116:2974-83.
38. Eszlinger M, Lau L, Ghaznavi S, et al. Molecular profiling of thyroid nodule fine-needle aspiration cytology. *Nat Rev Endocrinol* 2017;13:415-24.
39. Liu YY, Morreau H, Kievit J, et al. Combined immunostaining with galectin-3, fibronectin-1, CITED-1, Hector Battifora mesothelial-1, cytokeratin-19, peroxisome proliferator-activated receptor- $\gamma$ , and sodium/iodide symporter antibodies for the differential diagnosis of non-medullary thyroid carcinoma. *Eur J Endocrinol* 2008;158:375-84.
40. Martins MB, Marcello MA, Morari EC, et al. Clinical utility of KAP-1 expression in thyroid lesions. *Endocr Pathol* 2013;24:77-82.
41. Hoftijzer HC, Liu YY, Morreau H, et al. Retinoic acid receptor and retinoid X receptor subtype expression for the differential diagnosis of thyroid neoplasms. *Eur J Endocrinol* 2009;160:631-8.

**Cite this article as:** Jing L, Xia F, Du X, Jiang B, Chen Y, Li X. Identification of key candidate genes and pathways in follicular variant papillary thyroid carcinoma by integrated bioinformatical analysis. *Transl Cancer Res* 2020;9(2):477-490. doi: 10.21037/tcr.2019.11.38

Table S1 TCGA patient IDs of 357 cPTC patients with mRNA sequence and clinical data

Sample IDs	Sample IDs	Sample IDs	Sample IDs
TCGA-4C-A93U	TCGA-DJ-A3US	TCGA-EL-A3T7	TCGA-ET-A3BU
TCGA-BJ-A0YZ	TCGA-DJ-A3UU	TCGA-EL-A3T8	TCGA-ET-A3BV
TCGA-BJ-A0Z0	TCGA-DJ-A3UW	TCGA-EL-A3T9	TCGA-ET-A3BW
TCGA-BJ-A0Z3	TCGA-DJ-A3UX	TCGA-EL-A3TA	TCGA-ET-A3BX
TCGA-BJ-A0Z5	TCGA-DJ-A3UY	TCGA-EL-A3TB	TCGA-ET-A3DO
TCGA-BJ-A0ZA	TCGA-DJ-A3UZ	TCGA-EL-A3ZG	TCGA-ET-A3DP
TCGA-BJ-A0ZB	TCGA-DJ-A3V2	TCGA-EL-A3ZH	TCGA-ET-A3DR
TCGA-BJ-A0ZC	TCGA-DJ-A3V3	TCGA-EL-A3ZK	TCGA-ET-A3DS
TCGA-BJ-A0ZE	TCGA-DJ-A3V9	TCGA-EL-A3ZL	TCGA-ET-A3DT
TCGA-BJ-A0ZJ	TCGA-DJ-A3VA	TCGA-EL-A3ZM	TCGA-ET-A3DU
TCGA-BJ-A18Y	TCGA-DJ-A3VD	TCGA-EL-A3ZN	TCGA-ET-A40S
TCGA-BJ-A18Z	TCGA-DJ-A3VE	TCGA-EL-A3ZO	TCGA-ET-A40T
TCGA-BJ-A190	TCGA-DJ-A3VI	TCGA-EL-A3ZP	TCGA-FE-A230
TCGA-BJ-A192	TCGA-DJ-A4UL	TCGA-EL-A3ZQ	TCGA-FE-A231
TCGA-BJ-A28R	TCGA-DJ-A4UP	TCGA-EL-A3ZR	TCGA-FE-A232
TCGA-BJ-A28S	TCGA-DJ-A4UT	TCGA-EL-A3ZS	TCGA-FE-A233
TCGA-BJ-A28T	TCGA-DJ-A4V0	TCGA-EL-A3ZT	TCGA-FE-A234
TCGA-BJ-A28V	TCGA-DJ-A4V2	TCGA-EL-A4JV	TCGA-FE-A235
TCGA-BJ-A28W	TCGA-DJ-A4V5	TCGA-EL-A4JW	TCGA-FE-A236
TCGA-BJ-A28Z	TCGA-DO-A1K0	TCGA-EL-A4JX	TCGA-FE-A237
TCGA-BJ-A291	TCGA-DO-A2HM	TCGA-EL-A4JZ	TCGA-FE-A238
TCGA-BJ-A2N7	TCGA-E3-A3E0	TCGA-EL-A4K0	TCGA-FE-A239
TCGA-BJ-A2N8	TCGA-E3-A3E1	TCGA-EL-A4K1	TCGA-FE-A23A
TCGA-BJ-A2N9	TCGA-E3-A3E3	TCGA-EL-A4K2	TCGA-FE-A3PB
TCGA-BJ-A2NA	TCGA-E3-A3E5	TCGA-EL-A4K4	TCGA-FE-A3PC
TCGA-BJ-A2P4	TCGA-E8-A242	TCGA-EL-A4K6	TCGA-FK-A3SB
TCGA-BJ-A3EZ	TCGA-E8-A2EA	TCGA-EL-A4K7	TCGA-FK-A3SD
TCGA-BJ-A3F0	TCGA-E8-A2JQ	TCGA-EL-A4K9	TCGA-FK-A3SE
TCGA-BJ-A3PR	TCGA-E8-A3X7	TCGA-EL-A4KD	TCGA-FK-A3SG
TCGA-BJ-A3PT	TCGA-E8-A413	TCGA-EL-A4KG	TCGA-FK-A3SH
TCGA-BJ-A3PU	TCGA-E8-A414	TCGA-EL-A4KH	TCGA-FK-A4UB
TCGA-BJ-A45C	TCGA-E8-A415	TCGA-EL-A4KI	TCGA-FY-A2QD
TCGA-BJ-A45E	TCGA-E8-A416	TCGA-EM-A1CS	TCGA-FY-A3BL
TCGA-BJ-A45F	TCGA-E8-A417	TCGA-EM-A1CT	TCGA-FY-A3I4
TCGA-BJ-A45H	TCGA-E8-A418	TCGA-EM-A1CU	TCGA-FY-A3NM
TCGA-BJ-A45K	TCGA-E8-A419	TCGA-EM-A1CV	TCGA-FY-A3NN
TCGA-BJ-A408	TCGA-E8-A432	TCGA-EM-A22I	TCGA-FY-A3ON
TCGA-CE-A13K	TCGA-E8-A433	TCGA-EM-A22K	TCGA-FY-A3R6
TCGA-CE-A27D	TCGA-E8-A434	TCGA-EM-A22M	TCGA-FY-A3R7
TCGA-CE-A3MD	TCGA-E8-A436	TCGA-EM-A22O	TCGA-FY-A3R8
TCGA-CE-A3ME	TCGA-E8-A437	TCGA-EM-A22P	TCGA-FY-A3R9
TCGA-CE-A481	TCGA-E8-A438	TCGA-EM-A2CS	TCGA-FY-A3RA
TCGA-CE-A482	TCGA-E8-A44K	TCGA-EM-A2OX	TCGA-FY-A3TY
TCGA-CE-A483	TCGA-E8-A44M	TCGA-EM-A2OZ	TCGA-FY-A3YR
TCGA-CE-A484	TCGA-EL-A3CL	TCGA-EM-A2P0	TCGA-FY-A40K
TCGA-CE-A485	TCGA-EL-A3CM	TCGA-EM-A2P1	TCGA-FY-A40L
TCGA-DE-A0XZ	TCGA-EL-A3CN	TCGA-EM-A2P3	TCGA-FY-A40M
TCGA-DE-A0Y2	TCGA-EL-A3CO	TCGA-EM-A3AK	TCGA-FY-A4B0
TCGA-DE-A0Y3	TCGA-EL-A3CR	TCGA-EM-A3AN	TCGA-FY-A4B3
TCGA-DE-A3KN	TCGA-EL-A3CS	TCGA-EM-A3AO	TCGA-FY-A4B4
TCGA-DE-A4M8	TCGA-EL-A3CT	TCGA-EM-A3AQ	TCGA-FY-A76V
TCGA-DE-A4M9	TCGA-EL-A3CU	TCGA-EM-A3AR	TCGA-GE-A2C6
TCGA-DE-A4MA	TCGA-EL-A3CV	TCGA-EM-A3FJ	TCGA-H2-A26U
TCGA-DE-A4MB	TCGA-EL-A3CW	TCGA-EM-A3FK	TCGA-H2-A2K9
TCGA-DE-A4MC	TCGA-EL-A3CX	TCGA-EM-A3FM	TCGA-H2-A3RI
TCGA-DE-A4MD	TCGA-EL-A3CY	TCGA-EM-A3FO	TCGA-H2-A421
TCGA-DE-A69J	TCGA-EL-A3CZ	TCGA-EM-A3FQ	TCGA-H2-A422
TCGA-DE-A69K	TCGA-EL-A3D0	TCGA-EM-A3FR	TCGA-IM-A3EB
TCGA-DE-A7U5	TCGA-EL-A3D1	TCGA-EM-A3O3	TCGA-IM-A3ED
TCGA-DJ-A130	TCGA-EL-A3D4	TCGA-EM-A3O7	TCGA-IM-A420
TCGA-DJ-A13P	TCGA-EL-A3D5	TCGA-EM-A3SU	TCGA-J8-A3NZ
TCGA-DJ-A13T	TCGA-EL-A3D6	TCGA-EM-A3SX	TCGA-J8-A3O0
TCGA-DJ-A13U	TCGA-EL-A3GO	TCGA-EM-A3SZ	TCGA-J8-A3O1
TCGA-DJ-A13V	TCGA-EL-A3GP	TCGA-EM-A4FF	TCGA-J8-A3O2
TCGA-DJ-A1QD	TCGA-EL-A3GR	TCGA-EM-A4FM	TCGA-J8-A3YD
TCGA-DJ-A1QE	TCGA-EL-A3GS	TCGA-EM-A4FN	TCGA-J8-A3YE
TCGA-DJ-A1QF	TCGA-EL-A3GU	TCGA-EM-A4FO	TCGA-J8-A3YF
TCGA-DJ-A1QI	TCGA-EL-A3GV	TCGA-EM-A4FV	TCGA-J8-A3YG
TCGA-DJ-A1QN	TCGA-EL-A3GW	TCGA-ET-A25J	TCGA-J8-A3YH
TCGA-DJ-A1QQ	TCGA-EL-A3GX	TCGA-ET-A25K	TCGA-J8-A42S
TCGA-DJ-A2PN	TCGA-EL-A3GY	TCGA-ET-A25M	TCGA-J8-A4HW
TCGA-DJ-A2PO	TCGA-EL-A3GZ	TCGA-ET-A25N	TCGA-J8-A4HY
TCGA-DJ-A2PQ	TCGA-EL-A3H1	TCGA-ET-A25O	TCGA-KS-A41F
TCGA-DJ-A2PR	TCGA-EL-A3H2	TCGA-ET-A25P	TCGA-KS-A41J
TCGA-DJ-A2PS	TCGA-EL-A3H3	TCGA-ET-A2MX	TCGA-KS-A41I
TCGA-DJ-A2PU	TCGA-EL-A3H4	TCGA-ET-A2MY	TCGA-KS-A413
TCGA-DJ-A2PV	TCGA-EL-A3H5	TCGA-ET-A2MZ	TCGA-KS-A415
TCGA-DJ-A2PW	TCGA-EL-A3H7	TCGA-ET-A2N0	TCGA-KS-A417
TCGA-DJ-A2PZ	TCGA-EL-A3H8	TCGA-ET-A39J	TCGA-KS-A419
TCGA-DJ-A2Q0	TCGA-EL-A3MW	TCGA-ET-A39K	TCGA-KS-A41B
TCGA-DJ-A2Q1	TCGA-EL-A3MX	TCGA-ET-A39L	TCGA-KS-A41C
TCGA-DJ-A2Q4	TCGA-EL-A3MY	TCGA-ET-A39M	TCGA-L6-A4EP
TCGA-DJ-A2Q5	TCGA-EL-A3MZ	TCGA-ET-A39N	TCGA-L6-A4EQ
TCGA-DJ-A2Q6	TCGA-EL-A3N2	TCGA-ET-A39P	TCGA-MK-A4N6
TCGA-DJ-A2Q7	TCGA-EL-A3N3	TCGA-ET-A39R	TCGA-MK-A4N7
TCGA-DJ-A2QC	TCGA-EL-A3T0	TCGA-ET-A39S	TCGA-MK-A4N9
TCGA-DJ-A3UK	TCGA-EL-A3T1	TCGA-ET-A3BN	TCGA-QD-A8IV
TCGA-DJ-A3UM	TCGA-EL-A3T2	TCGA-ET-A3BP	-
TCGA-DJ-A3UN	TCGA-EL-A3T3	TCGA-ET-A3BQ	-
TCGA-DJ-A3UO	TCGA-EL-A3T6	TCGA-ET-A3BS	-

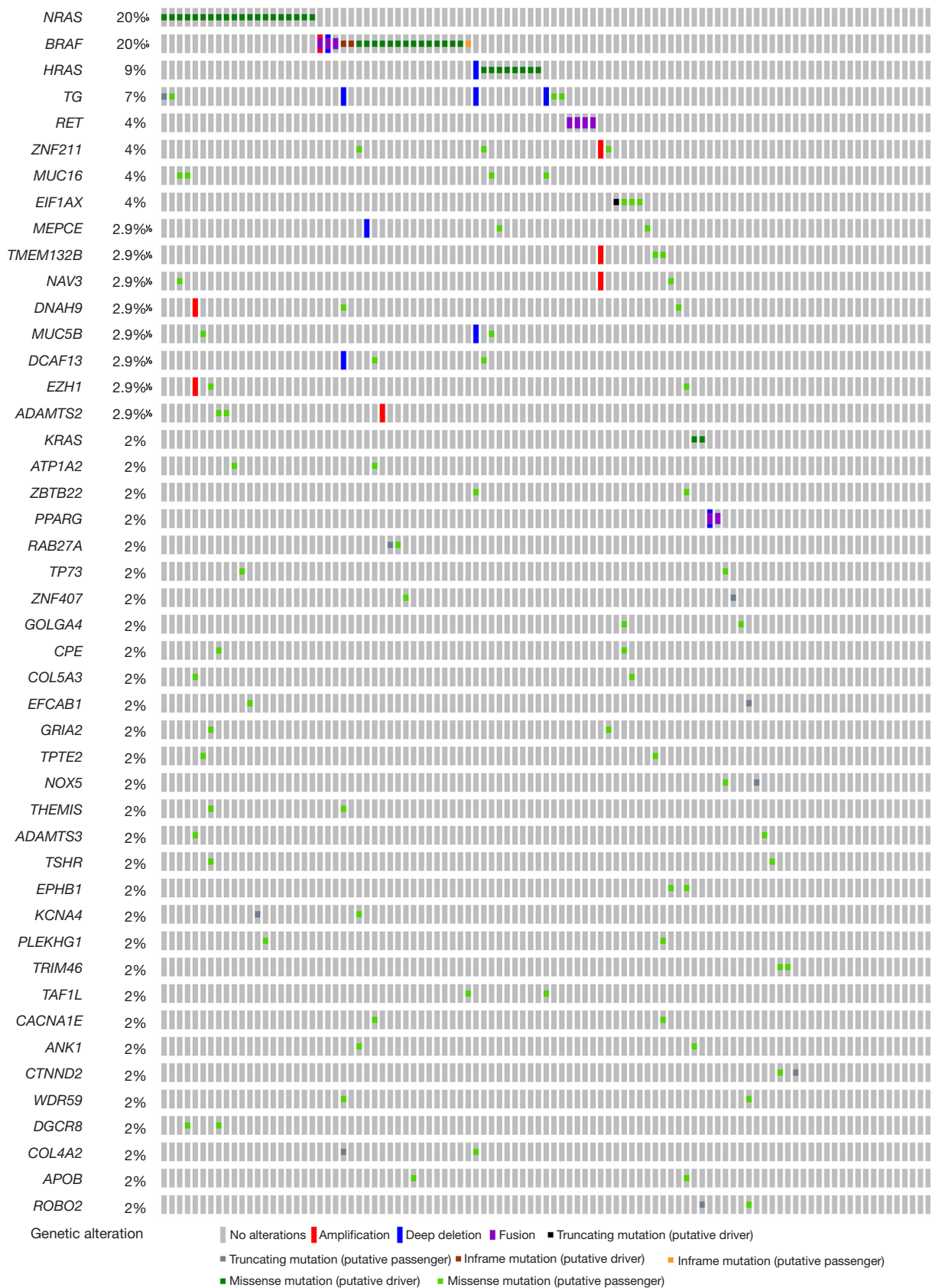
cPTC, conventional papillary thyroid carcinoma; TCGA, The Cancer Genome Atlas.

**Table S2** TCGA patient IDs of 102 FVPTC patients with mRNA sequence and clinical data

Sample IDs	Sample IDs	Sample IDs	Sample IDs
TCGA-BJ-A0Z2	TCGA-DJ-A3VM	TCGA-EM-A2CT	TCGA-ET-A25I
TCGA-BJ-A0ZG	TCGA-DJ-A4UR	TCGA-EM-A2CU	TCGA-ET-A2N3
TCGA-BJ-A45D	TCGA-DO-A1JZ	TCGA-EM-A2OV	TCGA-ET-A2N4
TCGA-BJ-A45G	TCGA-E3-A3DZ	TCGA-EM-A2OW	TCGA-ET-A2N5
TCGA-BJ-A4O9	TCGA-E3-A3E2	TCGA-EM-A2OY	TCGA-ET-A39I
TCGA-DE-A2OL	TCGA-EL-A3CP	TCGA-EM-A2P2	TCGA-ET-A39T
TCGA-DJ-A13M	TCGA-EL-A3GQ	TCGA-EM-A3AI	TCGA-ET-A3DQ
TCGA-DJ-A13R	TCGA-EM-A1CW	TCGA-EM-A3AJ	TCGA-ET-A3DV
TCGA-DJ-A13S	TCGA-EM-A1YA	TCGA-EM-A3AL	TCGA-ET-A40P
TCGA-DJ-A13W	TCGA-EM-A1YB	TCGA-EM-A3AP	TCGA-ET-A40R
TCGA-DJ-A13X	TCGA-EM-A1YC	TCGA-EM-A3FL	TCGA-ET-A4KQ
TCGA-DJ-A1QG	TCGA-EM-A1YD	TCGA-EM-A3FN	TCGA-FE-A22Z
TCGA-DJ-A1QL	TCGA-EM-A1YE	TCGA-EM-A3FP	TCGA-FE-A3PA
TCGA-DJ-A1QM	TCGA-EM-A22J	TCGA-EM-A3O6	TCGA-FE-A3PD
TCGA-DJ-A2PP	TCGA-EM-A22L	TCGA-EM-A3O8	TCGA-FK-A3S3
TCGA-DJ-A2PX	TCGA-EM-A22N	TCGA-EM-A3O9	TCGA-FY-A3I5
TCGA-DJ-A2Q2	TCGA-EM-A22Q	TCGA-EM-A3OA	TCGA-FY-A3NP
TCGA-DJ-A2QA	TCGA-EM-A2CJ	TCGA-EM-A3OB	TCGA-FY-A3W9
TCGA-DJ-A2QB	TCGA-EM-A2CK	TCGA-EM-A3ST	TCGA-FY-A3WA
TCGA-DJ-A3UP	TCGA-EM-A2CL	TCGA-EM-A3SY	TCGA-FY-A40N
TCGA-DJ-A3UR	TCGA-EM-A2CN	TCGA-EM-A4FH	TCGA-H2-A3RH
TCGA-DJ-A3UT	TCGA-EM-A2CO	TCGA-EM-A4FK	TCGA-IM-A41Z
TCGA-DJ-A3VG	TCGA-EM-A2CP	TCGA-EM-A4FQ	TCGA-KS-A41I
TCGA-DJ-A3VK	TCGA-EM-A2CQ	TCGA-EM-A4FU	TCGA-KS-A41L
TCGA-DJ-A3VL	TCGA-EM-A2CR	TCGA-EM-A4G1	TCGA-KS-A4ID
TCGA-L6-A4ET	-	-	-
TCGA-MK-A84Z	-	-	-

FVPTC, follicular variant papillary thyroid carcinoma; TCGA, The Cancer Genome Atlas.





**Figure S1** OncoPrints of all mutated genes in FVPTCs. In total, altered genes (a mutation in more than 2 cases) were detected in 78 of 102 sequenced FVPTC cases. FVPTC, follicular variant papillary thyroid carcinoma.

**Table S3** Clinicopathological features of FVPTC patients with or without BRAF mutation in the TCGA-THCA dataset

Variables	No. of patients	FVPTC with BRAF mutation	FVPTC without BRAF mutation	P value
Gender				0.777
Male	24	4	20	
Female	78	16	62	
Age				0.049*
≤45	46	5	41	
>45	56	15	41	
Multifocality				0.625
Unifocal	51	9	42	
Multifocal	50	11	39	
Extrathyroidal extension				0.012*
None	89	14	75	
Minimal (T3) + moderate/advanced (T4a)	12	6	6	
T stage				0.105
T1 + T2	72	11	61	
T3 + T4	30	9	21	
N stage				0.483
N0	65	14	51	
N1	13	4	9	
M stage				1.000
M0	32	11	21	
M1	5	1	4	
AJCC TNM stage				0.012*
Stage I + II	78	10	68	
Stage III + IV	23	9	14	
Disease free status				0.442 <sup>a</sup>
Disease free	95	18	77	
Recurred/progressed	6	2	4	
Overall survival status				1.000 <sup>a</sup>
Living	101	20	81	
Deceased	1	0	1	

Patients with censored data were excluded. P values were calculated by *t*-test or Fisher's exact test. \*, P<0.05; <sup>a</sup>, P value by log-rank test. FVPTC, follicular variant papillary thyroid carcinoma; TCGA-THCA, The Cancer Genome Atlas thyroid carcinoma.

Table S4 Average expression values, FC, and P values of 402 FVPTC-specific DEGs from the three comparison groups (FVPTC vs. NT; cPTC vs. NT and FVPTC vs. cPTC classes)

Table with columns: Gene name, Average expression, log2(RSEM), FVPTC vs. cPTC LogFC, adj.Padj, FVPTC vs. normal LogFC, adj.Padj, cPTC vs. normal LogFC, adj.Padj. The table lists 402 genes and their corresponding expression and significance values across the three comparison groups.

FC, fold changes; FVPTC, follicular variant papillary thyroid carcinoma; DEG, differentially expressed gene; NT, normal tissues; cPTC, conventional papillary thyroid carcinoma; RSEM, RNASeq by expectation-maximization.