Identification of a three-gene signature associated with immune infiltration to distinguish follicular thyroid cancer and adenoma: an integrated bioinformatic analysis

Yun Zhang (zhangyunendo@henu.edu.cn)
Henan Provincial People's Hospital https://orcid.org/0000-0001-7526-5714
WU Huan
Sun Yat-Sen Memorial Hospital
XIA Wei
Henan Provincial People's Hospital
WEI Wei
Henan Provincial People's Hospital
YUAN Huijuan
Henan Provincial People's Hospital

Primary research

Keywords: Follicular thyroid cancer, follicular thyroid adenoma, biomarker, immune infiltration, gene signature

Posted Date: February 23rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-239655/v1

Abstract

Background

Follicular thyroid cancer (FTC), accounting for about 15% of all thyroid cancers, was characterized by a more invasive nature and earlier hematogenous spread. The distinction between FTC and follicular thyroid adenoma (FTA) remained challenging for clinical scientists due to their highly similar cytological features at present, and frequently led to potentially unnecessary surgical procedures. The present study aimed to identify potential new biomarkers to improve diagnosis.

Methods

Three GEO datasets (GSE82208, GSE15045, and GSE29315) were downloaded to screen differentially expressed genes (DEGs) between FTC and FTA. GO and KEGG enrichment analysis were explored by ClusterProfiler package. A protein-protein interaction (PPI) network was established and hub genes were selected. Multivariate stepwise logistic regression analysis and receiver operating characteristic (ROC) analysis were used to construct and evaluate a gene signature model for differential diagnosis. Diagnostic efficacy of the model was further confirmed in an independent dataset GSE27155. CIBERSORT was used to characterize 22 infiltrating immune cell fractions of FTC and FTA samples.

Results

After gene integration, a total of 210 DEGs (55 upregulated and 155 downregulated) were identified, which were enriched in immune or cancer-related biological processes and pathways, such as cytokine-cytokine receptor interaction, fluid shear stress and atherosclerosis, and IL-17 signaling pathway. Seven hub genes were selected from PPI analysis. A three-gene signature (FOS, IL1B, and TOP2A) was constructed to distinguish FTC and FTA, with an area under the curve (AUC) of 0.839. Diagnostic efficacy of the signature was further conformed in an independent dataset GSE27155 with an AUC of 0.862. Moreover, the gene signature was associated with CD8 T-cell, which was significant different in FTC compared to FTA.

Conclusions

This study identified potential crucial genes and pathways that might be involved in the pathophysiology of FTC. Especially, a three-gene signature with diagnostic value was development. Furthermore, the landscape of immune infiltration in FTC was explored for the first time. The results would promote our understanding about the molecular mechanism of FTC development and provide potential target genes for differential diagnosis.

Background

Thyroid cancer was the most common endocrine malignancy with a rapidly increasing incidence in the past 30 years throughout the world [1]. Most of thyroid cancer was indolent and curable with low risk of recurrence or disease-specific mortality. An accurate preoperative diagnosis to identify malignant from lots of thyroid nodules was important, or it would lead to potentially unnecessary surgical procedures and potential complications. In this sense, follicular thyroid neoplasm might be the most controversial area in the thyroid pathology [2, 3].

Follicular thyroid cancer (FTC) and follicular thyroid adenoma (FTA) were the malignant and benign thyroid epithelial tumor showing follicular cell differentiation, respectively. FTC, accounting for about 15% of all thyroid cancers, was characterized by a more invasive nature and earlier hematogenous spread [4, 5]. The distinction between FTC and FTA was based on the presence of capsular and/or vascular invasion. However, neither of them could be easily identified by fine-needle aspiration biopsy (FNAB) cytology, the currently used methodology for preoperative diagnosis after clinical and ultrasound malignancy risk stratification [5, 6]. Several studies had investigated the gene expression profile and tried to find some markers to exclude malignancy preoperatively [7–16], however, reproducibility of results from different studies was rather low and no powerful molecular markers had been established. Although some genetic alterations had been found to play a fundamental role in FTC development in initial studies, such as RAS, H/K/NRAS, and PAX8-PPAR-γ [17–21], they were not found to be specific for FTC in subsequent studies, as these genetic alterations occurred in both FTCs and FTAs with similar frequencies [22–26]. So effective molecular markers to differentiate FTC from FTA were urgently needed.

Integrated bioinformatic analysis with multiple datasets had emerged recently as an efficacious novel approach to identify novel genes and comprehend the underlying molecular mechanisms of cancer [27, 28]. In the present study, we carried out a bioinformatic analyses based on four microarray datasets from Gene Expression Omnibus (GEO) database, to identify potential new markers to distinguish FTC and FTA. Three of them (GSE82208, GSE15045, and GSE29315) were used to screen differentially expressed genes (DEGs). Enrichment and protein–protein interaction (PPI) analysis were performed, and hub genes were selected. Subsequently, a gene signature was identified by logistic regression and further validated in another independent dataset GSE27155. Furthermore, association of the gene signature and tumor-infiltrating immune cells was analyzed. The flow chart of this study was shown in Fig. 1. This study would promote our understanding about the molecular mechanism of FTC development and provide potential target genes for differential diagnosis.

Methods

Gene Expression Profile Data

Four microarray gene expression profiles (GSE82208, GSE15045, GSE29315, and GSE27155) were obtained from GEO database (http://www.ncbi.nlm.nih.gov/geo). All included datasets met the following

inclusion criteria: 1) expression profiling by array; 2) tissue samples gathered from human FTC or FTA tissues; 3) included at least 10 samples. For GSE27155 and GSE29315, only FTC and FTA samples were selected for analysis in this study. Information about datasets was summarized in Table 1 and Additional file 1.

Table 1
The gene expression profile data characteristics.

Submission	Record	Platform	FTA	FTC	Original probes	Residue probes
Wojtas B, et al, 2016	GSE82208	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	25	27	54675	21655
Hinsch N, et al, 2009	GSE15045	GPL2986 ABI Human Genome Survey Microarray Version 2	4	8	32878	16752
Tomas G, et al, 2011	GSE29315	GPL8300 [HG_U95Av2] Affymetrix Human Genome U95 Version 2 Array	17	9	12625	8622
Giodano TJ, et al, 2011	GSE27155	GPL96 [HG-U133A] Affymetrix Human Genome U133A Array	10	13	22283	12548

Integrated Analysis of Microarray Datasets

Three GEO datasets (GSE82208, GSE15045 and GSE29315) were used for DEGs screening. The downloaded platform and matrix files were read and pre-processed utilizing Limma package in R/Bioconductor software [29]. The names of probes were converted to international standard gene names (gene symbols) and saved in TXT format. The probes without matching gene symbols were filtered out, and the values of different probes mapped to the same gene were averaged to obtain the final gene expression value. Limma package was also used to identify DEGs between FTC and FTA samples in each dataset. Subsequently, gene integration for the DEGs identified from the three test datasets was conducted employing RobustRankAggreg package [30]. $|\log2FC| > 0.5$ and P-value < 0.05 were defined as the threshold for identifying DEGs.

Functional and Pathway Enrichment Analysis

Gene Ontology (GO) enrichment analysis was one of the most important tools used to predict the potential functions of target genes based on biological processes (BP), cellular components (CC), and molecular function (MF) [31]. Kyoto Encyclopedia of Genes and Genomes (KEGG) was the major public database comprised of known genes and their biochemical functionalities [32]. So to investigate the functions and processes of the DEGs, ClusterProfiler package in R was utilized to annotate and visualize GO terms and KEGG pathway of upregulated and downregulated genes, respectively [33]. The gene annotation information was obtained from org.Hs.eg.Db. P-values were adjusted for multiple testing depending on the Benjamini–Hochberg False Discovery Rate (BH) method, and the adjusted P-values < 0.05 were regarded as the cut-off criteria for significant results.

PPI Network Construction and Hub Genes Identification

The Search Tool for the Retrieval of Interacting Genes (STRING) database provided information regarding the predicted and experimental interactions of proteins [34]. In this study, the identified DEGs were input into the online STRING database (https://string-db.org/) to detect the potential relationships and construct a PPI network. A combined score of ≥ 0.4 was used as the cut-off value. Thereafter, Cytoscape software (version 3.6.0) was used for the visual exploration of the PPI network. Moreover, the Cytoscape plug-in Molecular Complex Detection (MCODE) was used to screen notable modules in this PPI network. Degree cutoff = 2, Node Score Cutoff = 0.2, and K-Core = 2 were set as the advanced options. A MCODE score ≥ 5 was used as the selection criterion. Genes with a cut-off degree value of ≥ 24 were identified as hub genes. The enrichment analysis of each module and hub genes was performed by the Funrich software 3.1.3, including BP, CC, MF, transcription factor (TF) and biological pathway (BPA). A P-value < 0.05 was set as the cut-off criterion.

Expression Levels of Hub Genes in Malignant Thyroid Nodules

The expression of identified hub genes in FTC and FTA samples was explored in an independent dataset GSE27155, and the boxplot was performed to visualize the results.

The Cancer Genome Atlas (TCGA) database contained abundant genetic data of several other histological subtypes of thyroid cancer. UALCAN (http://ualcan.path.uab.edu/index.html) was a user-friendly, interactive web resource for analyzing cancer transcriptome data from TCGA [35]. In the current study, UALCAN was used to explore the expression levels of hub genes in TCGA thyroid cohort. A P-value < 0.05 was considered as statistically significant.

Gene Signature Selection for Differential Diagnosis of FTC and FTA

Logistic regression was a widely applied and useful statistical, non-linear model for predicting a binarized outcome based on a sequence of independent features. Receiver operating characteristic (ROC) curve was a comprehensive index used to reflect the sensitivity and specificity of continuous variables. In this study, logistic regression with stepwise backward selection was applied to construct a gene signature model for differential diagnosis of FTC and FTA. All the hub genes were used as independent variables. The diagnosis score was calculated as follows: $\text{Logic}(P) = \sum(C \times \text{EXPgene})$. EXP was the expression of the hub genes, and C was the regression coefficient for the corresponding gene in logistic regression analysis. FTA samples were treated as a reference group, whereas high-risk values indicated a highly probable occurrence of FTC. Then the diagnostic value of the model was evaluated with ROC curve by calculated the area under the curves (AUC). Furthermore, the selected gene signature was validated in independent dataset GSE27155. All analyses were conducted in R. All reported P-values were two-sided with $P < 0.05$ as significant.

Evaluation of Tumor-infiltrating Immune Cells and Its Association with the Gene Signature Selected

CIBERSORT was an analytical tool developed by Newman et al to characterize cell composition of complex tissues from normalized gene expression profiles [36]. In the present study, gene expression data of GSE82208 with a platform of Affymetrix U133 was uploaded to the CIBERSORT web portal (http://cibersort.stanford.edu/), to infer the relative proportions of 22 infiltrating immune cell types, with the algorithm run using the LM22 signature matrix at 1000 permutations. For each sample, all evaluated immune cell fractions combined were equal to 1. Besides, CIBERSORT calculated a global P-value for the immune cell fraction result based on the statistical probability of the presence of immune cells in the sample. All samples were included in this analysis regardless of P-value, for those samples with P-value > 0.05 might represent samples with low immune cell infiltrate. The relative proportions of 22 immune cell subpopulation were compared between FTC and FTA groups using Wilcoxon test. Spearman correlation was used to assess correlations between immune cell fractions and score of the gene signature selected above. A P-value < 0.05 was considered as statistically significant.

Results

Gene Expression Profile Data and DEGs Identified

In all, 210 genes (55 upregulated and 155 downregulated) were identified as DEGs in the FTC samples compared with the FTA ones, such as FOS, IL1B, and TOP2A (Figs. 2A-C and Additional file 2). According to the cut-off criteria, we screened 15 representative upregulated and downregulated genes (Fig. 2D).

GO and KEGG Enrichment Analyses

In order to gain insights into the biological functions of DEGs after gene integration, GO enrichment and KEGG pathway analysis were performed (Fig. 3 and Additional file 3). GO enrichment analysis revealed that 55 upregulated DEGs were mainly enriched in nuclear division and organelle fission in BP term, neuronal cell body, leading edge membrane, and condensed chromosome in CC term; however, downregulated DEGs were main enriched in ossification, response to metal ion, and epithelial cell proliferation in BP term, extracellular matrix in CC term, and DNA binding, carboxylic acid binding, and organic acid binding in MF term. In addition, 155 downregulated DEGs were analyzed by KEGG analysis and showed that they were most significant related to cytokine-cytokine receptor interaction, fluid shear stress and atherosclerosis, and IL-17 signaling pathway. No KEGG pathway was significant for upregulated DEGs.

PPI Network Construction and Hub Genes Identification

Based on the STRING database, we made a PPI network with 159 nodes and 513 edges to further explore the interaction among DEGs, as shown in Fig. 4A and Additional file 4. According to the degree of

importance, three modules (modules 1, 2 and 3) with score ≥ 5 were detected by MCODE, as shown in Figs. 4B-D.

Furthermore, enrichment analysis of module 1 and module 2 were performed and shown in Fig. 5 and Additional file 5. The chromosome segregation (P = 0.008) and regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolism (P < 0.001) were identified as the most significant BP term in module 1, while immune response (P = 0.046) was identified as the most significant BP term in module 2. Besides, for module 3, transport was identified as the most significant BP term (P = 0.003), and transporter activity was identified as the most significant MF term (P < 0.001). No pathway was significant.

Seven hub genes with high degree of connectivity were selected (Table 2). Enrichment analysis by FunRich of these hub genes were shown in Additional file 6 and 7. The cytokine activity (P < 0.001), DNA topoisomerase activity (P = 0.002), transcription factor activity (P = 0.003) and chemokine activity (P = 0.02) were identified as the most significant MF term, and immune response (P = 0.001) and regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolism (P = 0.014) were identified as the most important BP term.

Table 2
Hub genes with high degree of connectivity and their expression in TCGA.

Gene	Degree	MCODE Cluster	Type	P-value in TCGA Classical PTC vs Normal	P-value in TCGA FVPTC vs Normal
IL6	50	Cluster 1	Down-regulated	3.426300E-02	2.663100E-03
JUN	41	Cluster 2	Down-regulated	8.779300E-11	3.537800E-10
FOS	37	Cluster 2	Down-regulated	2.551300E-10	6.600900E-10
EGR1	32	Cluster 2	Down-regulated	7.098100E-10	3.974800E-09
IL1B	29	Cluster 2	Down-regulated	8.636200E-02	3.280900E-03
CCL2	25	Cluster 1	Down-regulated	7.001900E-03	1.420380E-04
TOP2A	24	Cluster 1	Up-regulated	4.364000E-02	9.968400E-01

Validation the Expression Levels of Hub Genes in TCGA and GSE27155

To validate the association of hub genes and malignant thyroid tumors, expression levels of hub genes in other histological subtypes of thyroid cancer in TCGA cohort was explored with UALCAN online tools. Results demonstrated that the expression of all seven hub genes in TCGA was also dysregulated and was consistent with that in GEO cohorts. As shown in Fig. 6A and Table 2, high expression of TOP2A and low expression of IL6, JUN, FOS, EGR1 and CCL2 were found in classical papillary thyroid cancer (PTC) tissue. Meanwhile, the expression levels of six downregulated hub genes, IL6, JUN, FOS, EGR1, IL1B, CCL2, were also downregulated in follicular variant of papillary thyroid cancer (FVPTC) compared to normal thyroid tissue. The expression level of TOP2A was upregulated in FVPTC but not significant.

The expression profiles of all 7 hub genes in GSE27155 were also explored. As shown in Fig. 6B, the expression level of CCL2 was significant downregulated in FTC samples compared to FTA samples, and the expression level of TOP2A was significant upregulated. It was consistent with the results above from the other three GEO datasets.

Gene Signature Selection and Validation for Differential Diagnosis of FTC and FTA

After a stepwise backward selection by logistic regression, three genes (FOS, IL1B, and TOP2A) were selected as signature genes to distinguish FTC patients from FTA patients (Fig. 7A). Using this combination of genes, we generated a final regression model with diagnostic score calculated as follows: Logic(P) = 2.216 + (-0.531 × EXPFOS) + (-0.239 × EXPIL1B) + (0.951 × EXPTOP2A). The model was significant in statistics, with a $\chi 2(3) = 36.959$, $P < 0.001$. The model had a sensitivity of 0.682 and a specificity of 0.783, together with positive predictive value 75% and negative predictive value 72%. Moreover, ROC analysis was performed for this diagnostic score system. As shown in Fig. 7C, the three-gene signature achieved an AUC as 0.839 (95% CI, 0.756–0.922, $P < 0.001$).

Then the score system of three-gene signature selected above was validated in another independent dataset GSE27155. As shown in Fig. 7D, the ROC curves showed that the score system was useful to distinguish FTC and FTA samples, with an AUC of 0.862 (95% CI, 0.714–1.000, $P = 0.004$). Taking all the four GEO dataset together, the AUC of the score system was 0.817 (95% CI, 0.740–0.894, $P < 0.001$, Fig. 8E).

The Landscape of Immune Infiltration and Its Association with the Gene Signature Selected

Using CIBERSORT algorithm, we investigated the different composition of tumor-infiltrating immune cells between FTC and FTA tissue. As shown in Fig. 8A-C, M0 and M2 macrophages accounted for the largest proportion among the 22 immune cell subpopulations, whereas the fractions of T cells CD4 memory activated, dendritic cells activated, and neutrophils were rare. The proportions of immune cells varied significantly between both intra- and intergroup. Compared with FTA tissue, FTC tissue generally contained a lower proportion for plasma cell and CD8 T-cells (Fig. 8B, $P < 0.05$). The proportion of CD8 T-

cells was moderately correlated with Macrophages M0 and T cells CD4 memory resting (Fig. 8D). Moreover, the proportion of CD8 T-cells showed significant negative Spearman correlation (P = 0.039) and Pearson correlation (P = 0.017) with score of three-gene signature selected above.

Discussion

Distinguishing between FTC and FTA remained challenging for clinical scientists resulting in equivocal histopathological diagnoses. Therefore, additional molecular markers to rule out malignancy and provide accurate diagnoses preoperative were urgently needed. In this study, we utilized a bioinformatics method to identify the potential biomarker and underlying molecular mechanisms of FTC.

Totally, 210 genes were identified from three profile datasets, including 55 upregulated and 155 downregulated genes, which were differently expressed in FTC samples compared with FTA ones. Chi J et al had performed a miRNA-mRNA regulatory network analysis and identified 86 DEGs and 32 differentially expressed miRNAs between FTC and FTA [37]. Similar work was done by Hu S et al [38]. In comparison to their study focused on individual dataset, our study integrated microarray data with relatively large sample size from multiple GEO datasets. Moreover, in-depth functional enrichment analysis was carried out and a PPI network was built.

Especially, a three-gene signature with differential diagnostic value to distinguish FTC from FTA was selected in our study and further validated in independent dataset. The AUC in all four GEO datasets reached 0.817. It would be very useful and better than any of single gene as biomarker.

Among the various mechanisms involved in cancer development, mounting evidence had found that the dysfunction of immune system might exert an important role. In a malignant environment, immune system homeostasis and control of self-tolerance were significantly altered [39]. In our study, cytokine-cytokine receptor interaction and IL-17 signaling pathway were the most significant pathways in KEGG analysis, suggest that dysfunction of immune system was also important in FTC compared to FTA. Cytokine-cytokine interaction pathway had also been reported to be associated with anaplastic thyroid cancer (ATC), the most aggressive thyroid cancer [40]. IL-17 was important for host defense and contributed to the pathogenesis of autoimmune diseases and cancer. Expression of IL-17 had been found to be upregulated in thyroid cancer tissues when compared with benign lesions including FTA, and was associated with recurrence and mortality [41]. IL-17RB could enhance thyroid cancer cell invasion and metastasis via ERK1/2 pathway-mediated MMP-9 expression [42]. Furthermore, 7 hub genes identified in our study were also significant enriched in immune response.

To further explore immune response involved in development of FTC, tumor-infiltrating immune cells were assessed. Association of immune cell infiltration and prognosis of patients with differentiated thyroid cancers (DTC) had been reported repeatedly [43–45]. However, few studies referred FTC. Here, we hypothesized that FTC would have different compositions of immune cell infiltration from FTA. And we used CIBERSORT deconvolution software to infer the relative proportions of 22 distinct leukocyte cell types in samples of GSE82208, for microarray gene expression data from Affymetrix U133 platform was

most suitable for this software. The results showed that, FTC tissue contained a lower proportion for CD8 T-cells than FTA tissue. Moreover, the proportion of CD8 T-cells showed significant negative correlation with score of three-gene signature selected above. Cytotoxic CD8 T-cells had been demonstrated to be related to protective anti-tumor immune responses that could eliminate tumor cells [46]. CD8 T-cell infiltration with COX2 expression might predict relapse in DTC patients [47]. However, Cunha et al had also reported DTC patients with chronic lymphocytic thyroiditis background and increased CD8 T-cell tumor infiltration showed increased disease-free survival [43]. The conflicting results might suggest a complex role in thyroid cancer which should be further explored.

Six downregulated genes, IL6, JUN, FOS, EGR1, IL1B, CCL2, and one upregulated gene, TOP2A, were identified as hub genes in our PPI analysis. The dysregulated expression of all seven hub genes was validated in other histological type of thyroid cancer in TCGA, supporting their role in thyroid cancers. Most of them were involved in carcinogenesis, for example, JUN was considered as key genes of PTC [48, 49], and FOS had been reported to contribute to the progression of ATC [40]. TOP2A gene encoded a DNA topoisomerase that was involved in DNA replication and DNA metabolic processes. High expression of TOP2A was closely related to malignant biological behaviors such as proliferation and invasion, and more importantly TOP2A had been served as a cancer target in clinical application [50]. Immunohistochemical analysis showed that TOP2A correlated with thyroid tumor histology and it was more frequently expressed in tumors with aggressive clinical behavior [51]. Also, TOP2A had been identified as a hub gene for ATC [52, 53].

Several limitations of our study should be noted. First, although our study recruited 4 datasets, the sample size was still relatively small, and clinical factors, such as sex, age, tumor staging, were not considered. So, we could not construct a personalized prediction model for FTC and FTA to utilize in the clinic. Second, the data used in our study was accessed from a public database while the quality of the data could not be appraised. Third, due to lack of clinical samples, the results obtained solely by means of bioinformatic analysis and were not be confirmed by an independent clinical and experimental studies, such as q-PCR. So further studies with larger sample sizes and experimental verification would be necessary to confirm these findings in the future.

Conclusions

In conclusion, we performed an integrated bioinformatic analysis from four GEO datasets and identified several core genes and pathways in FTC. Furthermore, a three-gene signature with diagnostic value was selected and validated by independent dataset. The landscape of immune infiltration in FTC was explored for the first time. These results would not only be helpful for differential diagnosis, but also provided a novel perspective on the molecular mechanisms underlying development of FTC. However, further experimental studies with larger cohort and more clinical data were required to confirm the findings.

Abbreviations

ATC
anaplastic thyroid cancer
AUC
area under the curves
BH
Benjamini–Hochberg False Discovery Rate
BP
biological processes
BPA
biological pathway
CC
cellular components
DEGs
differentially expressed genes
DTC
differentiated thyroid cancers
FNAB
fine-needle aspiration biopsy
FTA
follicular thyroid adenoma
FTC
follicular thyroid cancer
FVPTC
follicular variant of papillary thyroid cancer
GEO
Gene Expression Omnibus
GO
Gene Ontology
KEGG
Kyoto Encyclopedia of Genes and Genomes
MF
molecular function
MCODE
Molecular Complex Detection
PPI
protein–protein interaction
PTC
papillary thyroid cancer
ROC
Receiver operating characteristic

STRING
Search Tool for the Retrieval of Interacting Genes database
TCGA
The Cancer Genome Atlas database
TF
transcription factor

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and material: Publicly available datasets were analyzed in this study, which can

be found in GEO database (https://www.ncbi.nlm.nih.gov/geo/).

Competing interests: The authors declare that they have no competing interests.

Funding: This study was supported by the National Natural Science Foundation of China (81900721).

Authors' contributions: Conceptualization, ZY and YH; Data curation and formal analysis, ZY and WH; Validation, XW and WW; Writing original draft, ZY and WH; Writing-review & editing, XW and Y H.

Acknowledgements: Not applicable.

References

[1] Kitahara CM, Sosa JA. The changing incidence of thyroid cancer. Nat Rev Endocrinol. 2016;12:646-53.

[2] Sherman SI. Thyroid carcinoma. Lancet. 2003;361:501-11.

[3] LiVolsi VA, Baloch ZW. Follicular-patterned tumors of the thyroid: the battle of benign vs. malignant vs. so-called uncertain. Endocr Pathol. 2011;22:184-9.

[4] Grani G, Lamartina L, Durante C, Filetti S, Cooper DS. Follicular thyroid cancer and Hürthle cell carcinoma: challenges in diagnosis, treatment, and clinical management. Lancet Diabetes Endocrinol. 2018;6:500-14.

[5] Daniels GH. Follicular Thyroid carcinoma: a perspective. Thyroid. 2018;28:1229-42.

[6] Yamashina M. Follicular neoplasms of the thyroid. Total circumferential evaluation of the fibrous capsule. Am J Surg Pathol. 1992;16:392-400.

[7] Borup R, Rossing M, Henao R, Yamamoto Y, Krogdahl A, Godballe C, et al. Molecular signatures of thyroid follicular neoplasia. Endocr Relat Cancer. 2010; 7:691-708.

[8] Stolf BS, Santos MM, Simao DF, Diaz JP, Cristo EB, Hirata R Jr, et

1. Class distinction between follicular adenomas and follicular carcinomas of the thyroid gland on the basis of their signature expression. Cancer. 2006;106:1891-900.

[9] Hinsch N, Frank M, Döring C, Vorländer C, Hansmann ML. QPRT: a potential marker for follicular thyroid carcinoma including minimal invasive variant; a gene expression, RNA and immunohistochemical study. BMC Cancer. 2009;9:93.

[10] Williams MD, Zhang L, Elliott DD, Perrier ND, Lozano G, Clayman GL, et al. Differential gene expression profiling of aggressive and nonaggressive follicular carcinomas. Hum Pathol. 2011;42:1213-20.

[11] Chudova D, Wilde JI, Wang ET, Wang H, Rabbee N, Egidio CM, et al. Molecular classification of thyroid nodules using high-dimensionality genomic data. J Clin Endocrinol Metab. 2010;95:5296-304.

[12] Keutgen XM, Filicori F, Crowley MJ, Wang Y, Scognamiglio T, Hoda R, et al. A panel of four miRNAs accurately differentiates malignant from benign indeterminate thyroid lesions on fine needle aspiration. Clin Cancer Res. 2012;18:2032-8.

[13] Yoo SK, Lee S, Kim SJ, Jee HG, Kim BA, Cho H, et al. Comprehensive analysis of the transcriptional and mutational landscape of follicular and papillary thyroid cancers. PLoS Genet. 2016;12:e1006239.

[14] Jung SH, Kim MS, Jung CK, Park HC, Kim SY, Liu J, et al. Mutational burdens and evolutionary ages of thyroid follicular adenoma are comparable to those of follicular carcinoma. Oncotarget. 2016;7:69638-48.

[15] Swierniak M, Pfeifer A, Stokowy T, Rusinek D, Chekan M, Lange D, et al. Somatic mutation profiling of follicular thyroid cancer by next generation sequencing. Mol Cell Endocrinol. 2016;433:130-7.

[16] Nicolson NG, Murtha TD, Dong W, Paulsson JO, Choi J, Barbieri AL, et al. Comprehensive genetic analysis of follicular thyroid carcinoma predicts prognosis independent of histology. J Clin Endocrinol Metab. 2018;103:2640-50.

[17] Xing M. Clinical utility of RAS mutations in thyroid cancer: a blurred picture now emerging clearer. BMC Med. 2016;14:12.

[18] Xing M. Molecular pathogenesis and mechanisms of thyroid cancer. Nat Rev Cancer. 2013;13:184-99.

[19] Xing M, Haugen BR, Schlumberger M. Progress in molecular-based management of differentiated thyroid cancer. Lancet. 2013;381:1058-69.

[20] Liu D, Yang C, Bojdani E, Murugan AK, Xing M. Identification of RASAL1 as a major tumor suppressor gene in thyroid cancer. J Natl Cancer Inst. 2013;105:1617-27.

[21] Vasko V, Ferrand M, Di Cristofaro J, Carayon P, Henry JF, de Micco C. Specific pattern of RAS oncogene mutations in follicular thyroid tumors. J Clin Endocrinol Metab. 2003;88:2745-52.

[22] Cheung L, Messina M, Gill A, Clarkson A, Learoyd D, Delbridge L, et al. Detection of the PAX8-PPAR gamma fusion oncogene in both follicular thyroid carcinomas and adenomas. J Clin Endocrinol Metab. 2003;88:354-7.

[23] Sahin M, Allard BL, Yates M, Powell JG, Wang XL, Hay ID, et al. PPARgamma staining as a surrogate for PAX8/PPARgamma fusion oncogene expression in follicular neoplasms: clinicopathological correlation and histopathological diagnostic value. J Clin Endocrinol Metab. 2005;90:463-8.

[24] Kloos RT, Reynolds JD, Walsh PS, Wilde JI, Tom EY, Pagan M, et al. Does addition of BRAF V600E mutation testing modify sensitivity or specificity of the Afirma Gene Expression Classifier in cytologically indeterminate thyroid nodules. J Clin Endocrinol Metab. 2013;98:E761-8.

[25] Fukahori M, Yoshida A, Hayashi H, Yoshihara M, Matsukuma S, Sakuma Y, et al. The associations between RAS mutations and clinical characteristics in follicular thyroid tumors: new insights from a single center and a large patient cohort. Thyroid. 2012;22:683-9.

[26] Zhu Z, Gandhi M, Nikiforova MN, Fischer AH, Nikiforov YE. Molecular profile and clinical-pathologic features of the follicular variant of papillary thyroid carcinoma. An unusually high prevalence of ras mutations. Am J Clin Pathol. 2003;120:71-7.

[27] Zhao J, Lv T, Quan J, Zhao W, Song J, Li Z, et al. Identification of target genes in cardiomyopathy with fibrosis and cardiac remodeling. J Biomed Sci. 2018;25:63.

[28] Zhao L, Fong AHW, Liu N, Cho WCS. Molecular subtyping of nasopharyngeal carcinoma (NPC) and a microRNA-based prognostic model for distant metastasis. J Biomed Sci. 2018;25:16.

[29] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.

[30] Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics. 2012;28:573-80.

[31] Tweedie S, Ashburner M, Falls K, Leyland P, McQuilton P, Marygold S, et al. FlyBase: enhancing Drosophila Gene Ontology annotations. Nucleic Acids Res. 2009;37:D555-9.

[32] Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017;45:D353-61.

[33] Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16:284-7.

[34] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43:D447-52.

[35] Chandrashekar DS, Bashel B, Balasubramanya S, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, et al. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. Neoplasia. 2017;19:649-58.

[36] Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12:453-7.

[37] Chi J, Zheng X, Gao M, Zhao J, Li D, Li J, et al.Integrated microRNA-mRNA analyses of distinct expression profiles in follicular thyroid tumors. Oncol Lett. 2017;14:7153-60.

[38] Hu S, Liao Y, Zheng J, Gou L, Regmi A, Zafar MI, et al. In silico integration approach reveals key microRNAs and their target genes in follicular thyroid carcinoma. Biomed Res Int. 2019; 2019:2725192.

[39] Mascitelli L, Goldstein MR. Statin immunomodulation and thyroid cancer. Clin Endocrinol (Oxf). 2015;82:620.

[40] Huang Y, Tao Y, Li X, Chang S, Jiang B, Li F, et al. Bioinformatics analysis of key genes and latent pathway interactions based on the anaplastic thyroid carcinoma gene expression profile. Oncol Lett. 2017;13:167-76.

[41] Carvalho D, Zanetti BR, Miranda L, Hassumi-Fukasawa MK, Miranda-Camargo F, Crispim J, et al. High IL-17 expression is associated with an unfavorable prognosis in thyroid cancer. Oncol Lett. 2017;13:1925-31.

[42] Ren L, Xu Y, Liu C, Wang S, Qin G. IL-17RB enhances thyroid cancer cell invasion and metastasis via ERK1/2 pathway-mediated MMP-9 expression. Mol Immunol. 2017;90:126-35.

[43] Cunha LL, Morari EC, Guihen AC, Razolli D, Gerhard R, Nonogaki S, et al. Infiltration of a mixture of immune cells may be related to good prognosis in patients with differentiated thyroid carcinoma. Clin Endocrinol (Oxf). 2012;77:918-25.

[44] French JD, Weber ZJ, Fretwell DL, Said S, Klopper JP, Haugen BR. Tumor-associated lymphocytes and increased FoxP3+ regulatory T cell frequency correlate with more aggressive papillary thyroid cancer. J Clin Endocrinol Metab. 2010;95:2325-33.

[45] Cunha LL, Marcello MA, Ward LS. The role of the inflammatory microenvironment in thyroid carcinogenesis. Endocr Relat Cancer. 2014;21:R85-103.

[46] Hadrup S, Donia M, Thor Straten P. Effector CD4 and CD8 T cells and their role in the tumor microenvironment. Cancer Microenviron. 2013;6:123-33.

[47] Cunha LL, Marcello MA, Nonogaki S, Morari EC, Soares FA, Vassallo J, et al. CD8+ tumour-infiltrating lymphocytes and COX2 expression may predict relapse in differentiated thyroid cancer. Clin Endocrinol (Oxf). 2015;83:246-53.

[48] Yu J, Mai W, Cui Y, Kong L. Key genes and pathways predicted in papillary thyroid carcinoma based on bioinformatics analysis. J Endocrinol Invest. 2016;39:1285-93.

[49] Chen W, Liu Q, Lv Y, Xu D, Chen W, Yu J. Special role of JUN in papillary thyroid carcinoma based on bioinformatics analysis. World J Surg Oncol. 2017;15:119.

[50] Węsierska-Gądek J, Składanowski A. Therapeutic intervention by the simultaneous inhibition of DNA repair and type I or type II DNA topoisomerases: one strategy, many outcomes. Future Med Chem. 2012;4:51-72.

[51] Ludvíková M, Holubec L Jr, Ryska A, Topolcan O. Proliferative markers in diagnosis of thyroid tumors: a comparative study of MIB-1 and topoisomerase II-a immunostaining. Anticancer Res. 2005;25:1835-40.

[52] Gao X, Wang J, Zhang S. Integrated bioinformatics analysis of hub genes and pathways in anaplastic thyroid carcinomas. Int J Endocrinol. 2019;2019:9651380.

[53] Hu S, Liao Y, Chen L. Identification of key pathways and genes in anaplastic thyroid carcinoma via integrated bioinformatics analysis. Med Sci Monit. 2018;24:6438-48.

Figures

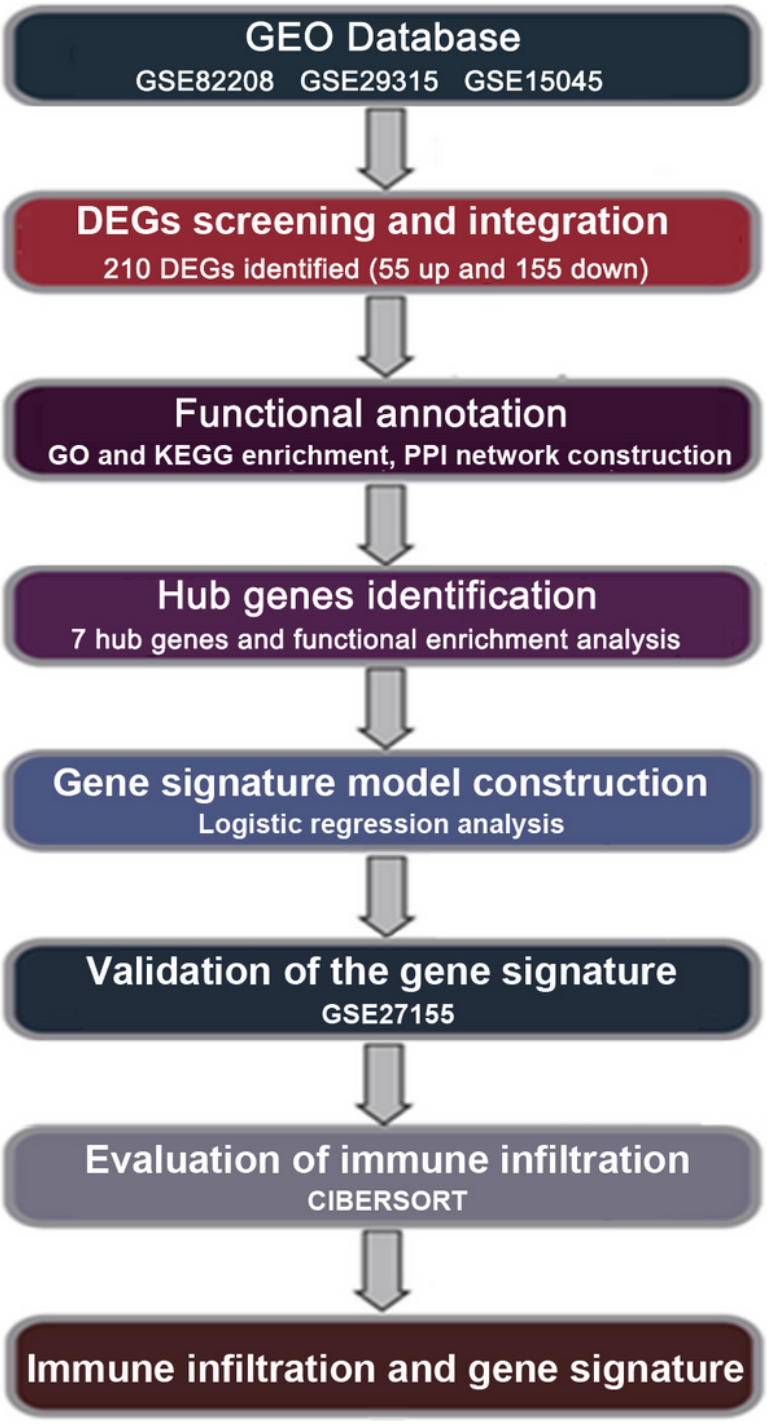

Figure 1

Workflow of the present study.

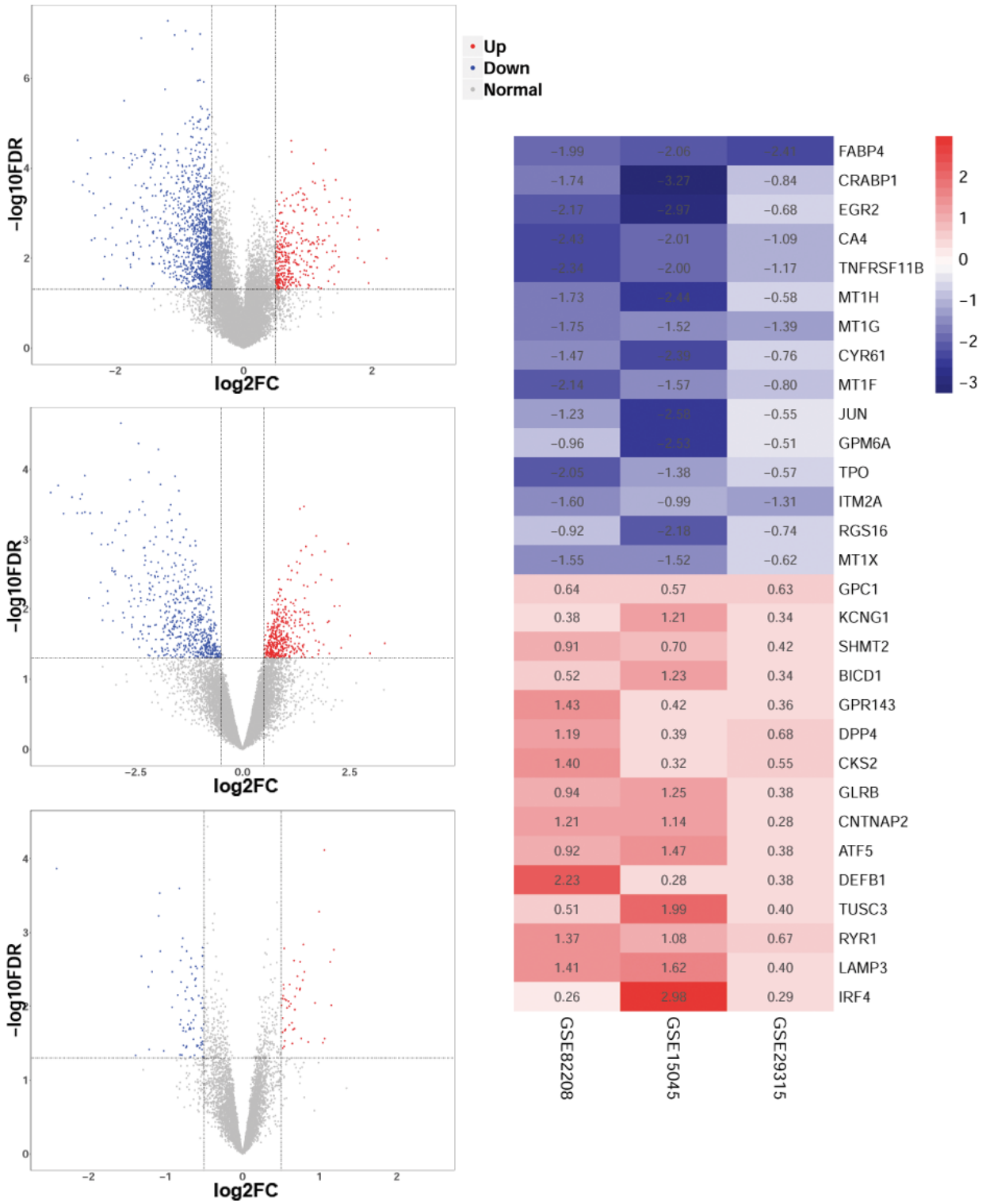

Figure 2

Identification of DEGs in FTC samples compared with the FTA ones. (A) Volcano plot of GSE82208; (B) volcano plot of GSE15045; (C) volcano plot of GSE29315; and (D) heat map of representative DEGs. Each column represented one dataset and each row represented one gene. The number in each rectangle represented the log2FC value. Blue represented downregulated in FTC samples, red represented

upregulated and white represents that there was no different expression. The gradual color ranged from blue to red represented the changing process from downregulation to upregulation.

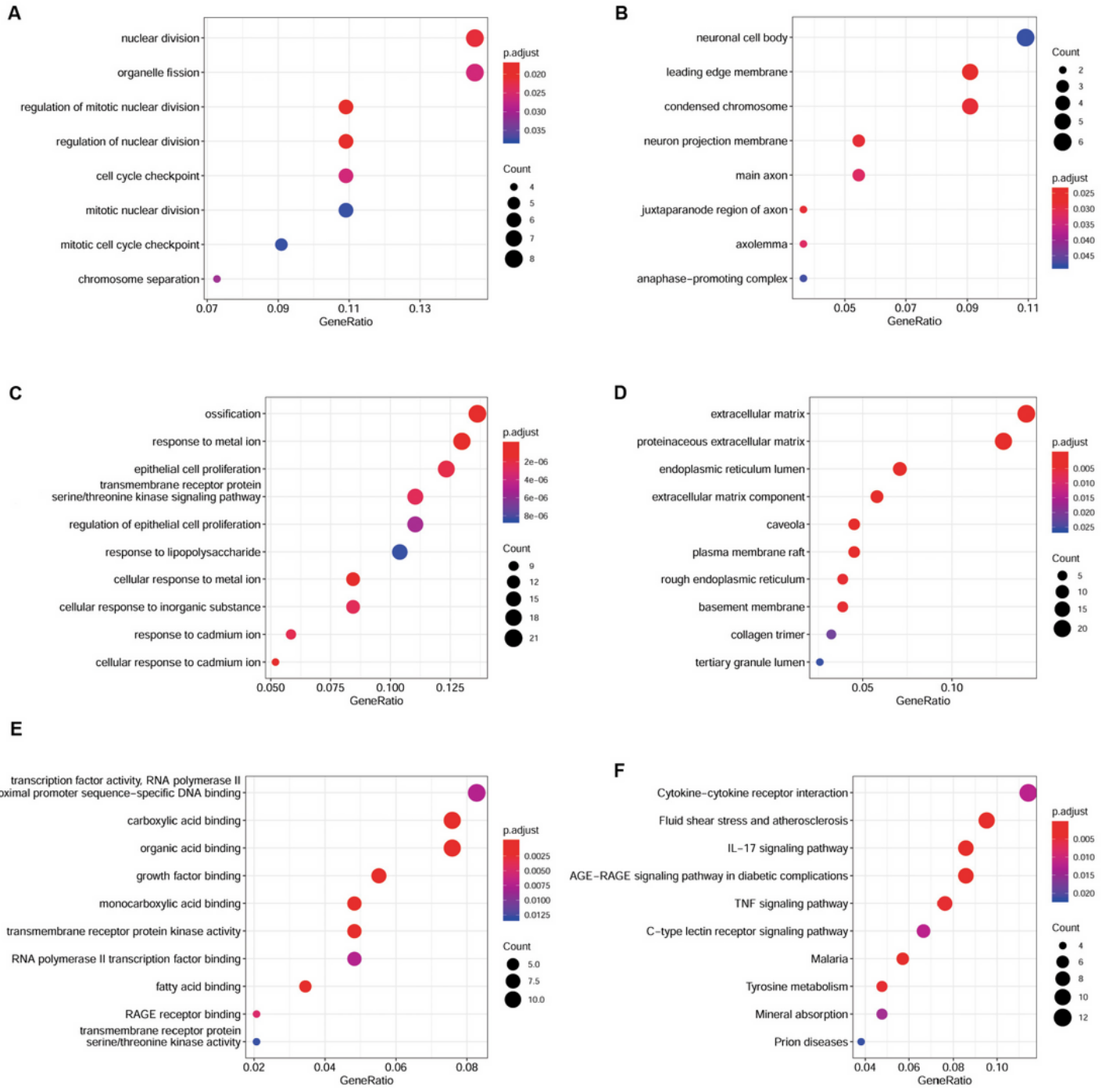

Figure 3

GO functional and KEGG pathway enrichment analysis of DEGs. The y-axis showed significantly enriched GO terms and KEGG pathways of DEGs, and the x-axis showed the gene ratio, FDR < 0.05. The size of the dot indicated the number of target genes in the pathway, and the color of the dot reflected the different P-

value range. (A-B) upregulated GO terms in BP and CC; (C-E) downregulated GO terms in BP, CC and MF; (F) downregulated KEGG pathway.

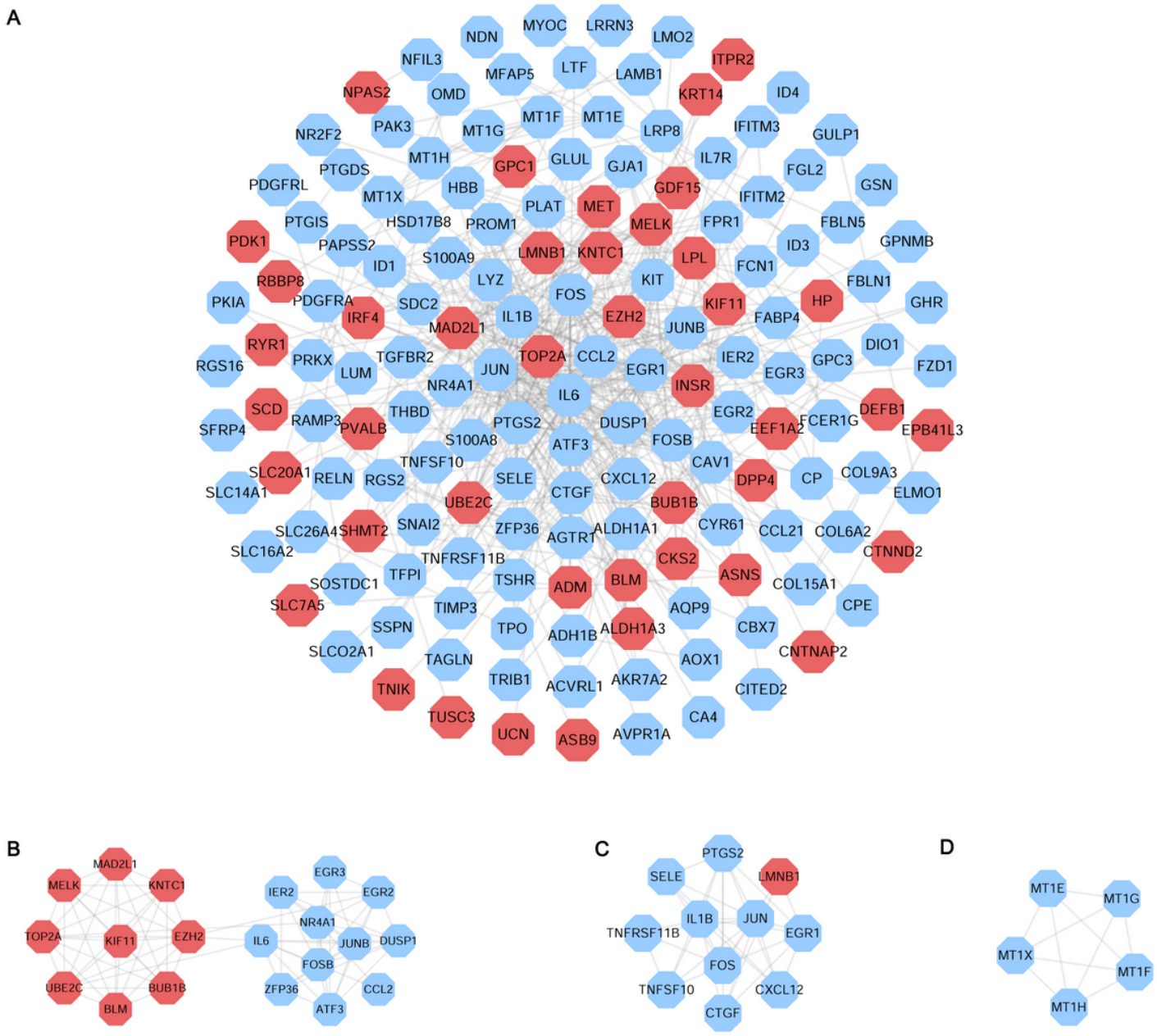

Figure 4

PPI network of DEGs and significant modules with a score of ≥5.0. Red nodes, upregulated genes; Blue nodes, downregulated genes. (A) PPI network of DEGs in FTC samples compared with the FTA ones; (B) Module 1, MCODE score = 8.105; (C) Module 2, MCODE score = 6.8; (B) Module 3, MCODE score = 5.

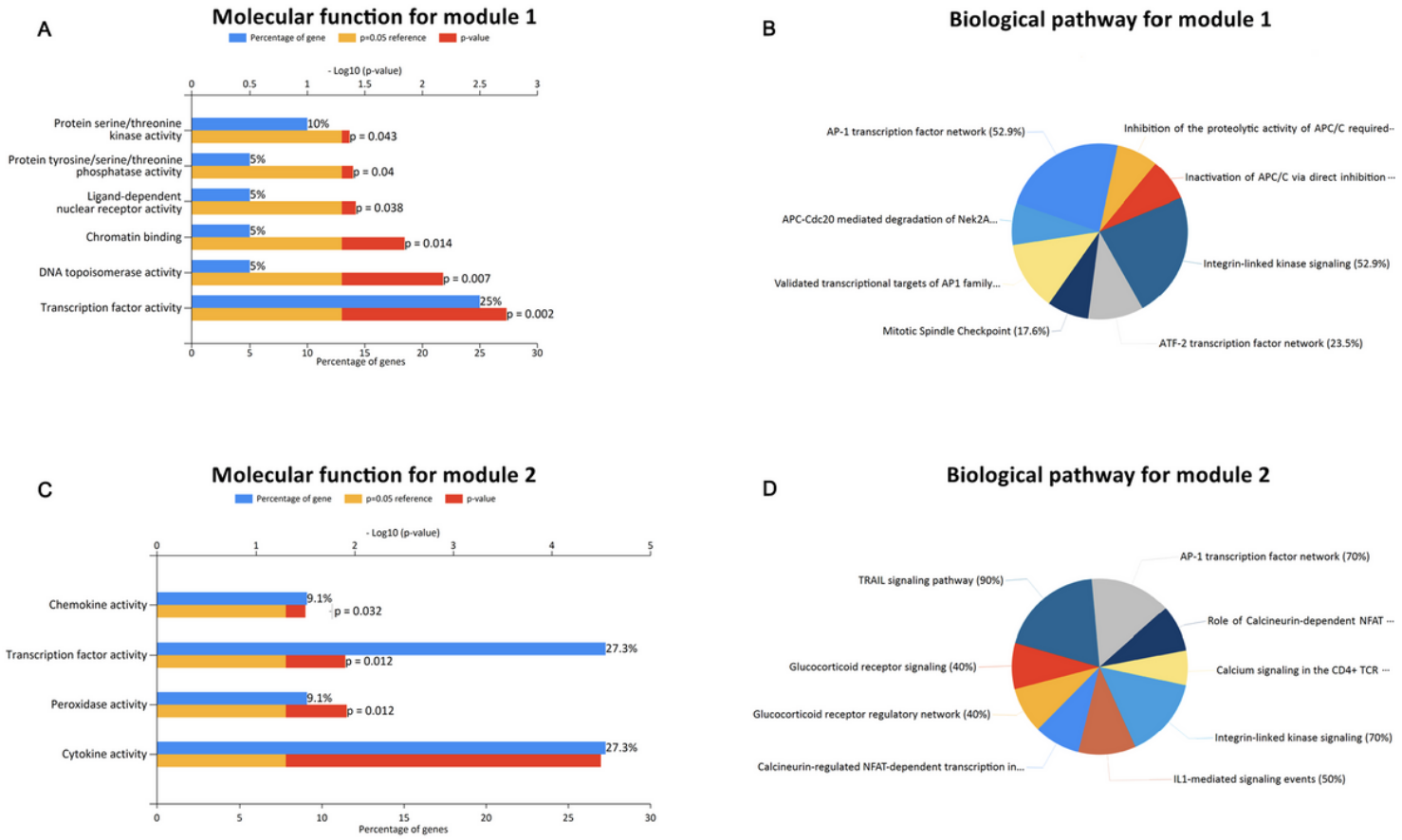

Figure 5

Enrichment analysis of module 1 and module 2. (A, C) MF for module 1 and module 2; X axis represented percentage of genes or –log10 (P-value); Y axis represented GO terms. (B, D) Biological pathway for module 1 and module 2.

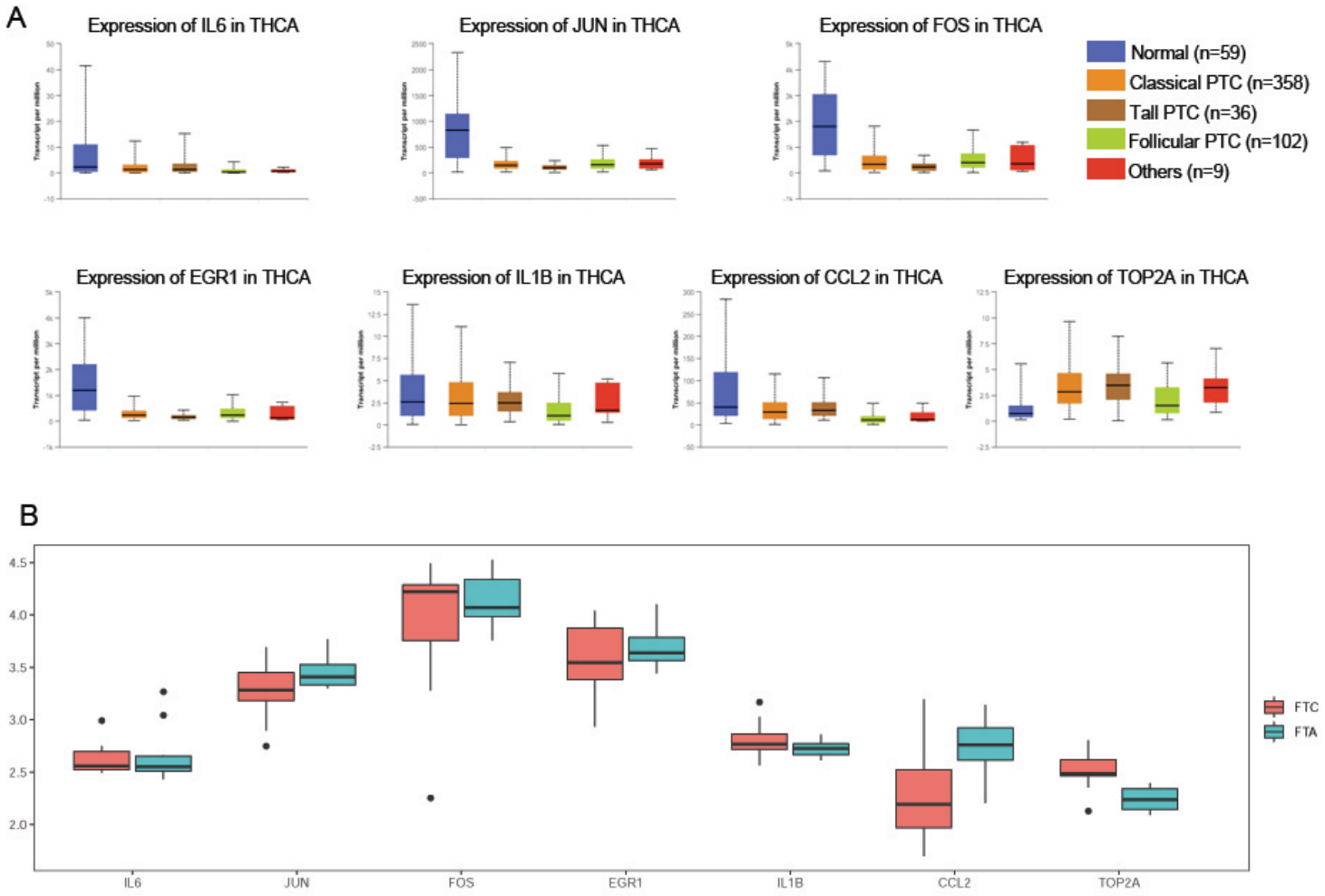

Figure 6

Expression levels of hub genes. (A) The expression levels of hub genes in each histological subtype of thyroid cancer in TCGA. (B) Expression level of hub genes in FTC and FTA samples of GSE27155 dataset.

A

	B	S.E	Wals	P	OR	95% CI
TOP2A	.951	.209	20.706	<0.001	2.588	1.718~3.897
FOS	-.531	.146	13.136	<0.001	0.588	0.442~0.784
IL1B	-.239	.140	2.913	0.088	0.788	0.599~1.036
Constant	2.216	1.020	4.714	0.030	9.167	

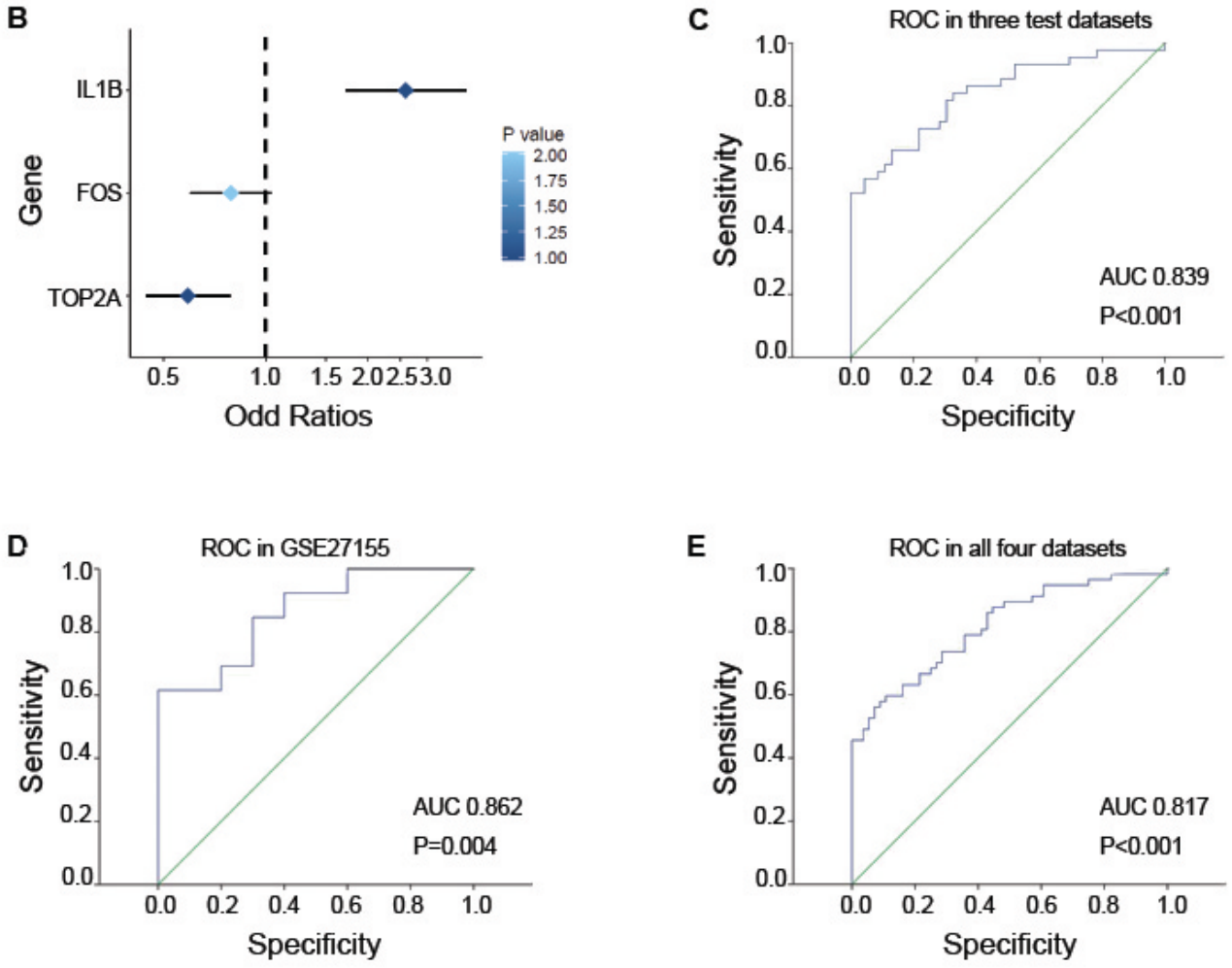

Figure 7

Identification of a three-gene signature by logistic regression. (A) Score model of the gene signature selected by logistic regression. (B) Forest map of the score model. (C) ROC curve of the score model in three test datasets (GSE82208, GSE15045, and GSE29035). (D) Validation of the score model by ROC curve in independent dataset GSE27155. (E) ROC curve of the score model in all four datasets.

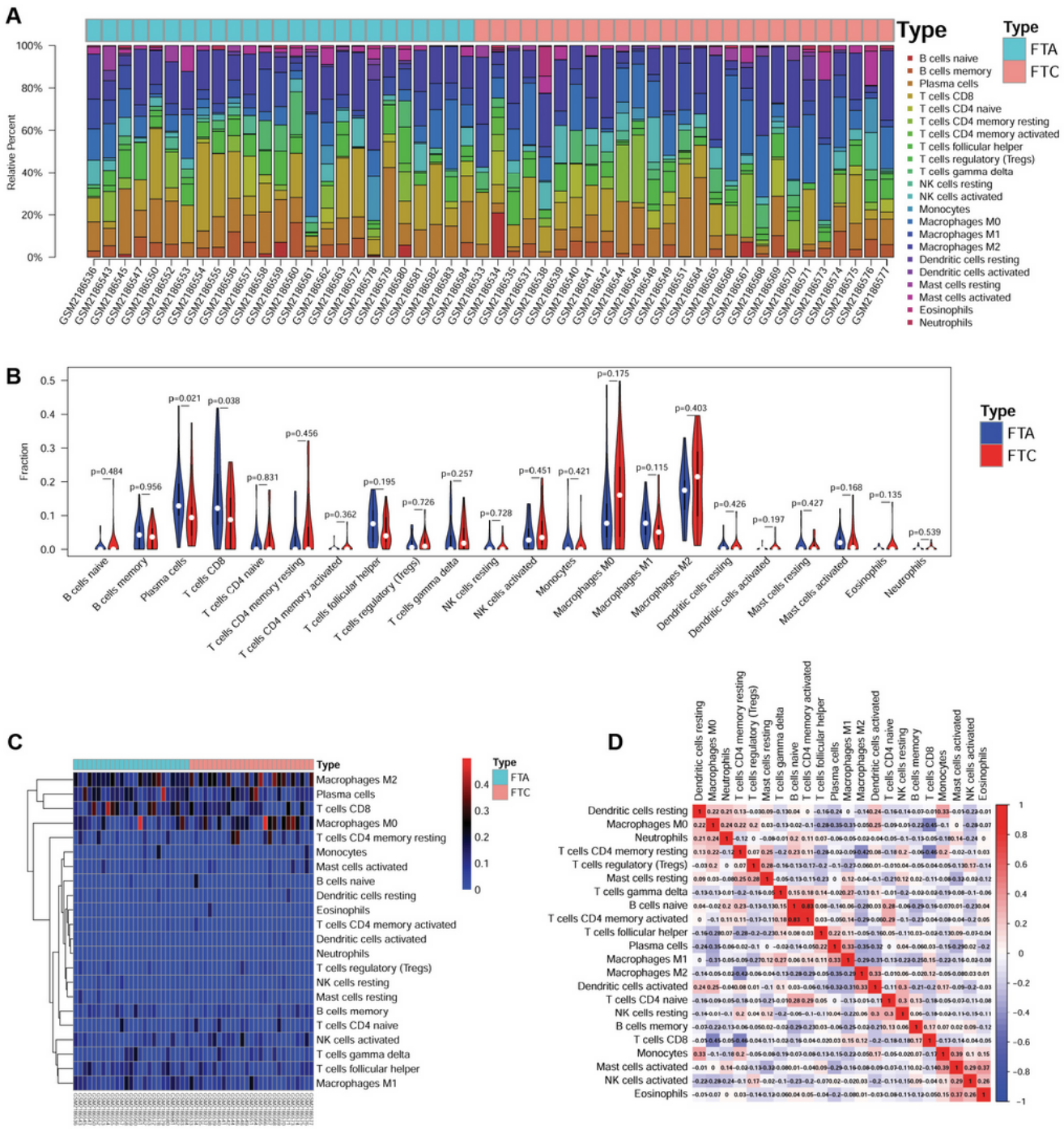

Figure 8

The landscape of immune infiltration in samples of GSE82208. (A) Distribution of immune cell fractions in each sample. (B) Difference of immune infiltration between FTC and FTA tissue. (C) Heatmap of immune cell fractions. (D) Heatmap of the correlation among immune cells.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile7.pdf
- SupplementaryMaterialsTables.xlsx