Identification of Candidate Biomarkers Related to Diagnosis and Prognosis of Thyroid Cancer via Transcriptomic Profile Analysis

Qiang Li
Public Health College, Shanghai Jiao Tong University School https://orcid.org/0000-0001-8711-2643

Sheng Jiang
The Second Affiliated Hospital of Chengdu Medical College, China National Nuclear Corporation 416 Hospital

Tienan Feng
Hongqiao International Institute of Medicine, Shanghai Tongren Hospital/Clinical Research Institute, Shanghai Jiao Tong University School of Medicine

Tengteng Zhu (ztt@shsmu.edu.cn)
Shanghai Jiao Tong University School of Medicine https://orcid.org/0000-0001-7589-7228

Biyun Qian
Hongqiao International Institute of Medicine, Shanghai Tongren Hospital/Clinical Research Institute, Shanghai Jiao Tong University School of Medicine

Research article

Keywords: Thyroid cancer, Diagnosis, Prognosis, Bioinformatic analysis, Immunohistochemistry

Posted Date: October 19th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-92586/v1

Abstract

Background

The detection rate of thyroid cancer (TC) has been continuously improved due to the development of detection technology. Compared with other cancers, the gene profile plays a more prominent role in the diagnosis and treatment of TC.

Methods

Four datasets from Gene Expression Omnibus (GEO) was used to perform transcriptomic profile analysis. The overlapping differentially expressed genes (DEGs) were analyzed by R package “limma” and “RobustRankAggreg”. Then the hub genes, which had a higher degree, were identified by the protein-protein interaction (PPI) network. Gene expression analysis between the TC and normal data, the disease-free survival (DFS) analysis of TC patients from Gene Expression Profiling Interactive Analysis (GEPIA) database, function analysis, and immunohistochemistry (IHC) were performed to verify the importance of the hub genes.

Results

A total of 80 DEGs (34 upregulated and 46 downregulated) were identified. Then *FN1*, *TIMP1*, *ITGA2*, and *KIT* were considered hub genes, which had a high degree of connectivity in the PPI network. GEPIA identified that *FN1*, *TIMP1*, and *ITGA2* were upregulated, and *KIT* was downregulated. Upregulations of *FN1* expression (*P*=0.024) and *ITGA2* expression (*P*=0.029) and downregulation of *KIT* expression (*P*=0.012) increased risks of decreased DFS for patients. IHC showed that the expression of FN1, TIMP1, and ITGA2 protein were upregulated, while the expression of KIT protein was downregulated in the TC clinical specimens. Besides, five hub genes were enriched in the PI3K-Akt signaling pathway and ECM-receptor interaction.

Conclusions

In summary, these hub genes were potential biomarkers of diagnosis and prognosis of TC.

Background

Thyroid cancer (TC) is the most common endocrine malignancy. In the past 40 years, the incidence of TC has risen steadily in many countries around the world [1]. With the improvement of detection technology, the incidence of recent TC in China in 2014 was 12.40/100,000, making it one of the fastest-growing malignant tumors [2]. Although the incidence of TC is increasing year by year, the mortality rate of TC is stable or decreased slightly [3] due to early diagnosis [4]. Ultrasonography is the main method for stratifying the risk of early cancerous thyroid nodules. Patients with suspected TC could take a fine-needle aspiration biopsy (FNAB) or surgical resection, and followed by a pathological examination to assess the nature of cancer [5]. However, in some cases, due to the overlapping of cytological features of

malignant and benign thyroid nodules, approximately 15%-30% of biopsies are uncertain [6]. And the uncertain diagnosis may bring Inappropriate treatment leading to poor quality of life, such as the surgical removal of the thyroid will seriously affect the quality of life of patients. Ultrasound and FNAB can only be diagnosed and cannot respond to the question of whether to perform surgery. Therefore, supplementing and improving traditional diagnostic methods of TC patients is still an urgent problem to be addressed.

Gene expression profiling is an important aspect of bioinformatics, which has broad application prospects and powerful functions in oncology medicine. It has great clinical application potential in molecular diagnosis, tumor molecular screening, new target discovery, tumor response prediction, patient classification, and prognosis prediction [7]. More and more genes and their encoded proteins are involved in TC. Previous studies have shown that they have significant effects on cell proliferation, differentiation, apoptosis, and metastasis [8, 9]. However, the exact molecular mechanism of TC has not been determined. Recently, various studies have identified key genes related to TC through bioinformatics analysis. Zhang et al. [10] identified hundreds of DEGs in papillary thyroid cancer (PTC) based on the GEO database, and six genes of which were key genes. Besides, based on two GEO microarray datasets, 228 commonly changed DEGs (141 upregulated and 87 downregulated) were identified in anaplastic thyroid carcinoma (ATC), and seven of these were selected as hub genes [11]. However, a comprehensive transcriptomic profile analysis of TC has not yet been conducted.

In this study, through the analysis of existing genetic data, a transcriptomic profile of thyroid cancer was constructed, and further analysis was performed, as well as exploring the potential biomarkers for diagnosis and prognostic guidance.

Methods

Data collection

Data screening was completed by two researchers independently, and consistent comparison was conducted after completion. We searched in Gene Expression Omnibus (GEO) database [12] (http://www.ncbi.nlm.nih.gov/geo) for all published data related to gene expression in thyroid cancer until August 1, 2020, by using the following keywords: (thyroid cancer OR cancerous goiter OR carcinoma of thyroid OR thyroid carcinoma). Finally, a total of 4 thyroid cancer data sets from GEO database (GSE29265, GSE33630, GSE60542, and GSE65144) were included in this study. Datasets inclusion criteria: (1) they contained human TC samples and normal tissue samples. (2) their sample sizes at least 20. The information of four GEO datasets was shown in Table S1.

Integrated gene expression analysis

All data were processed using the R software (version 3.6.1). The limma package [13] was used for identifying the DEGs between the TC samples and normal samples. The adjust *P*-value < 0.05 and the

absolute fold change (FC) >= 1.5 were considered statistically significant. The RobustRankAggreg package[14], which is a method for robust rank aggregation, was used to integrate the different gene results of four data sets. In total, 80 DEGs with a score of < 0.01 were selected.

Functional enrichment analysis

Gene Ontology (GO) can annotate genes, gene products and sequences with potential biological phenomena in three aspects: biological process (BP), cellular component (CC), and molecular function (MF) [15]; Kyoto Encyclopedia of Genes and Genomes (KEGG) is a comprehensive database for understanding advanced functions and utilities of biological systems [16]. Metascape (http://metascape.org/) is an online tool that provides biologists with annotations and analysis of gene lists [17]. In this study, GO and KEGG analysis of the DEGs were performed by Metascape.

Hub genes selection and analyses

The online tool Search Tool for the Retrieval of Interacting Genes database (STRING) (https://string-db.org/) [18] was used for analyzing the interactive relationships of the DEGs. The combined score of ≥0.4 was the cut-off value. Cytoscape software (version 3.6.0) was then used to make PPI networks [19]. And the hub genes were identified, which degrees at least 4. Next, we took the clusterProfiler [20] package to get the enrichment analysis of these 5 genes.

Expression level analysis and survival analysis of hub genes

We demonstrated the expression of hub genes in TC tissues (512 samples) and normal ones (337 samples) on the website of Gene Expression Profiling Interactive Analysis (GEPIA) (http://gepia.cancer-pku.cn/index.html) [21], which is a novel tool containing the Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) data for gene expression analysis. Moreover, the disease-free survival (DFS) information of these hub genes was also shown by the GEPIA.

Patients and ethical approval

From October 2018 to October 2019, a total of 5 thyroid cancer specimens and paired normal specimens from the Shanghai Tong Ren Hospital were enrolled in the study. All diagnoses were based on pathological and/or cytological evidence. Ethical approval was obtained from the Ethics Committee of the Shanghai Tong Ren Hospital and informed consent was given by all patients before sample collection.

Immunohistochemistry (IHC)

Five human thyroid normal and tumor tissues were fixed in 4% formaldehyde and embedded in paraffin. Paraffin-embedded samples were cut into 6 μm thick sections, and then they were placed in xylene and ethanol solution for deparaffinating and rehydration in PBS. 3% hydrogen peroxide solution and 5% donkey serum were used to block endogenous peroxidase activity and nonspecific sites after samples were placed in the citric acid solution for antigen retrieval. Sections were stained with rabbit antibodies against FIBRONECTIN (Abcam, ab268021, 1:600), TIMP1 (Abcam, ab109125, 1:500), ITGA2 (Abcam, ab133557, 1:200), and KIT (Abcam, ab32363, 1:200) for incubating at 4℃ overnight, and then incubation in the second antibody (Abcam, ab97051, 1:1000) was carried out at room temperature for 2h. Antigen-antibody complexes were detected using ChemMateTM EnVisonTM/HRP complex (DAB) as a peroxidase substrate (GK500705, Gene Tech). Results were visualized under an optical microscope (OLYMPUS IX73, JAPAN). All pictures were taken under the same microscope camera (OLYMPUS DP80, JAPAN) with uniform parameters (OLYMPUS cellSens Standard 1.18) after sections were mounted with xylene sealant. Brown staining areas were selected with eyedropper tool after dropping colors represented by fewer than 10 pixels, and the mean option density were calculated in the manner of integrated option density (IOD) / area using Image Pro Plus 6.0 software.

Results

Identification of differentially expressed genes in TC and data integration

There were 1956 DEGs in GSE29265, 3115 DEGs in GSE33630, 2110 DEGs in GSE60542, 5535 DEGs in GSE65144. Among the DEGs, 907, 1608, 1148, and 2598 genes were upregulated while 1049, 1507, 962, and 2937 genes were downregulated in GSE29265, GSE33630, GSE60542, and GSE65144, respectively (Fig. 1A-D). After integrated analyses of the four GEO datasets, 80 genes (34 upregulated and 46 downregulated) were identified as the DEGs, and the cluster heatmap of these DEGs is shown in Fig. 1E.

Significant functions and pathway enrichment analysis

For 34 upregulated DEGs list and 46 downregulated DEGs list, a pathway and process enrichment analysis were carried out respectively by Metascape (Fig. 2A-D). Terms with an adjusted *P*-value <0.05 were collected. Functional enrichment analyses demonstrated that upregulated DEGs were mainly related to cell adhesion, extracellular matrix (ECM), and glucose metabolism. Downregulated DEGs were mainly related to immune, and tyrosine degradation.

Hub genes identification in the PPI network

We used the STRING database and Cytoscape to construct a PPI network to further analyze the interaction between the DEGs (Fig. 3). There were 69 nodes and 31 edges in the network. As a result, fibronectin 1 (*FN1*), the tissue inhibitor of metalloproteinase 1 (*TIMP1*), integrin alpha 2 (*ITGA2*), and

receptor tyrosine kinase (*KIT*), which have a high degree of connectivity, were identified as hub genes (Table S2).

Enrichment analysis, expression level, and survival analysis of hub genes

In the present study, we contrasted enrichment analysis of hub genes by the clusterProfiler package, and adjusted *P*-value <0.05 was a cut-off value. As a result, 4 hub genes were significantly enriched in the collagen binding, protease binding, and proteoglycan binding by GO analysis (Fig. 4A), while they were significantly enriched in the PI3K-Akt signaling pathway and ECM-receptor interaction by KEGG analysis (Fig. 4B). Then, we used GEPIA to verify the expression level of hub genes between TC tissues and normal ones, which contained expression data from TCGA and GETx. The result identified that *FN1*, *TIMP1*, and *ITGA2* were upregulated in the TC tissues, while *KIT* was downregulated in the TC tissues (Fig. 5A-E). Otherwise, we also tested 4 genes for DFS analyses in TC patients on the GEPIA. It showed that upregulations of *FN1* expression (*P*=0.024) and *ITGA2* expression (*P*=0.029) and downregulation of *KIT* expression (*P*=0.012) increased risks of decreased DFS for patients (Fig. 6A-E).

Immunohistochemistry (IHC)

To confirm the reliability of the hub genes, the IHC was used to verify the protein expression levels of the 4 hub genes. Consistent with mRNA expression, the IHC results showed that FN1, TIMP1, and ITGA2 protein were upregulated while KIT protein was downregulated in the TC tissues (Fig. 7).

Discussion

In this study, we comprehensively analyzed four microarray datasets from GEO, which were all published data related to gene expression in thyroid cancer and normal ones in GEO database until August 1, 2020. A total of 80 DEGs (34 upregulated DEGs and 46 downregulated) were identified between TC tissues and normal ones. Functional enrichment analyses demonstrated that upregulated DEGs were mainly related to cell adhesion, extracellular matrix (ECM), and glucose metabolism. Downregulated DEGs were mainly related to immune, and tyrosine degradation. We finally screened out *FN1*, *TIMP1*, *ITGA2*, and *KIT* were hub genes from the PPI network, and they were significantly enriched in the PI3K-Akt signaling pathway and ECM-receptor interaction. Then, we identified that *FN1*, *TIMP1*, and *ITGA2* were upregulated in the TC tissues, while *KIT* was downregulated in the TC tissues by GEPIA. Otherwise, the DFS analyses on the GEPIA showed that upregulations of *FN1* expression (*P*=0.024) and *ITGA2* expression (*P*=0.029) and downregulation of *KIT* expression (*P*=0.012) increased risks of decreased DFS for patients. Besides, it was verified in clinical samples that the protein expression levels of the 4 hub genes were consistent with the mRNA expression pattern.

Fibronectin 1 (*FN1*) participates in cell adhesion and migration processes and is expressed in multiple cell types [22]. *FN1* has also turned out to be associated with the ECM changes [23]. Numerous studies demonstrated that it plays an important role in various cancers, such as oral squamous cell carcinoma [24], nasopharyngeal carcinoma [25], ovarian cancer [26], renal cancer [27], and thyroid cancer [28-33]. In the thyroid cancer, Sponziello M et al demonstrated that the overexpression of *FN1* mediated thyroid tumor cell migration and invasion by 86 patients [28]. Shaohua Zhan et al identified *FN1* as a novel prognostic biomarker associated with sporadic medullary thyroid cancer pathophysiological changes [29]. Furthermore, some studies found *FN1* was highly expressed in TC from a series of data sets, but they did not offer a survival analysis of *FN1* [30-33]. In this study, we demonstrated *FN1* could be associated with TC occurrence and worse DFS, which may be through the PI3K-Akt signaling pathway and ECM-receptor interaction pathway.

The tissue inhibitor of metalloproteinase 1 (*TIMP1*) has a role in matrix remodeling by inhibiting matrix metalloproteinase (MMP). Moreover, some studies have shown *TIMP1* is closely linked to cancers. Nieuwesteeg MA et al [34] showed that overexpression of *TIMP1* can decrease the invasiveness ability of pancreatic cancer cells, and it plays the same role in human lung carcinoma cell lines [35]. However, *TIMP1* has also been recorded to serve as an oncogene in endometrial cancer [36], breast cancer [37], and brain cancer [38]. The discrepancies may be explained by the complicated functions of *TIMP1* in different cancers. *TIMP1* can not only suppress cancer by inhibiting the expression and activation of MMP, but also promote cancer by angiogenesis, promoting cell growth and tumor inflammation [39]. The present study showed that *TIMP1*, which was identified as a hub gene, was upregulated in the TC tissues.

Integrin alpha 2 (*ITGA2*) is an alpha subunit, which often combines with the beta subunit to form a heterodimer α2β1, and then participates in the adhesion of platelets and other cells to the extracellular matrix [40-42]. *ITGA2* is highly expressed in many tumors, including pancreatic cancer, gastric cancer [43], liver cancer [44], prostate cancer [45], and breast cancer [46]. More studies suggested that *ITGA2* might be closely related to tumor cell migration, invasion, and metastasis [47, 48]. In thyroid cancer, one study identified that *ITGA2* expression was upregulated in PTC tissues but not in BRAF-positive samples [49], and another study showed that PTC patients with high *ITGA2* expression had poorer relapse-free survival than PTC patients with low *ITGA2* expression [50]. In our study, we also identified *ITGA2* was a hub gene, and the higher expression of this gene increased risks of decreased DFS for TC patients by the PI3K-Akt signaling pathway and ECM-receptor interaction pathway.

Receptor tyrosine kinase (*KIT*), a form of myeloid receptor that binds the stem cell factor, plays a major role in cancer occurrence [51]. Papers are showing that *KIT* is highly expressed in small cell lung cancer [52], leukemia cells [53], colon cancer [54], and neuroblastoma [55]. But *KIT* expression is lower in breast cancer [56] and melanoma [57]. Numerous studies have investigated the expression of *KIT* in TC [58-60], suggesting its role in thyroid epithelial cell differentiation and growth control. Therefore, these inconsistent results in turn indicate that *KIT* may be crucial in TC and validate our results.

The phosphoinositide 3-kinase–protein kinase B/AKT (PI3KPKB/AKT) pathway is one of the most prominent molecular signaling pathways implicated in the cellular growth, proliferation, apoptosis, and metabolism [61]. Activation of the PI3K-AKT signaling pathway is critical to the occurrence and development of thyroid cancer [62-64]. In this study, we found *FN1*, *ITGA2*, and *KIT* were enriched in the PI3K-Akt signaling pathway. The extracellular matrix–receptor (ECM-receptor) interaction pathway plays an important role in the process of tumor shedding, adhesion, movement, degradation, and hyperplasias, such as prostate cancer [65], gastric cancer [66], breast cancer [67], and anaplastic thyroid cancer [68]. Our study showed that *FN1* and *ITGA2* could influence the occurrence and development of thyroid cancer by participating in the ECM-receptor interaction pathway.

In this study, we found 4 hub genes, which were potential biomarkers of diagnosis and prognosis of TC, but our observations still have some limitations. As the cumulation of TC samples, subtype analysis was required to perform and experiments based on our wide transcriptomic analysis can be designed and conducted.

Conclusions

In conclusion, this study used existing data to construct a more comprehensive transcriptomic profile of TC and found potential biomarkers for TC diagnosis and prognosis. However, it is necessary to further study the specific role of these genes in TC.

Abbreviations

TC: thyroid cancer; GEO: Gene Expression Omnibus; DEGs: differentially expressed genes; PPI: protein-protein interaction; DFS: disease-free survival; GEPIA: Gene Expression Profiling Interactive Analysis; IHC: immunohistochemistry; FNAB: fine-needle aspiration biopsy; PTC: papillary thyroid cancer; ATC: anaplastic thyroid carcinoma; FC: fold change; GO: Gene Ontology; BP: biological process; CC: cellular component; MF: molecular function; KEGG: Kyoto Encyclopedia of Genes and Genomes; STRING: Search Tool for the Retrieval of Interacting Genes; TCGA: the Cancer Genome Atlas; GTEx: the Genotype-Tissue Expression; *FN1*: fibronectin 1; *TIMP1*: the tissue inhibitor of metalloproteinase 1; *ITGA2*: integrin alpha 2; *KIT*: receptor tyrosine kinase; ECM: extracellular matrix; MMP: matrix metalloproteinase.

Declarations

Ethics approval and consent to participate

The informed consent of all patients has been obtained before the operation, and the procedures for organizing collection have been approved by the Ethics Committee of the Shanghai Tong Ren Hospital.

Consent for publication

Not applicable.

Availability of data and materials

The data used to support the findings of this study are included in the article.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by Startup Fund for Youngman Research at SJTU (to TF, 17X100040015), Shanghai Jiao Tong University Medical and Industrial Cross Project (to TF, YG2017QN70), and The Scientific Project of Shanghai Municipal Health Commission (to BY, Grant No.2018ZHYL0202).

Authors' contributions

QL and SJ analyzed the data and helped draft the manuscript. TNF edited and revised the manuscript. TTZ and BYQ formulated the main research plan. The authors read and approved the final manuscript.

Acknowledgments

Thank you very much for the data provided by the NCBI-GEO database.

References

1. Siegel RL, Miller KD, Jemal A: **Cancer statistics, 2020**. *CA: A Cancer Journal for Clinicians* 2020, **70**(1):7-30.
2. Chen W, Sun K, Zheng R, Zeng H, Zhang S, Xia C, Yang Z, Li H, Zou X, He J: **Cancer incidence and mortality in China, 2014**. *Chinese journal of cancer research* 2018, **30**(1):1.
3. Jegerlehner S, Bulliard J-L, Aujesky D, Rodondi N, Germann S, Konzelmann I, Chiolero A, Group NW: **Overdiagnosis and overtreatment of thyroid cancer: a population-based temporal trend study**. *PloS one* 2017, **12**(6).
4. Maniakas A, Davies L, Zafereo ME: **Thyroid disease around the World**. *Otolaryngologic Clinics of North America* 2018, **51**(3):631-642.
5. Li X, Zhang S, Zhang Q, Wei X, Pan Y, Zhao J, Xin X, Qin C, Wang X, Li J: **Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective,**

multicohort, diagnostic study. *The Lancet Oncology* 2019, **20**(2):193-201.

6. Abooshahab R, Gholami M, Sanoie M, Azizi F, Hedayati M: **Advances in metabolomics of thyroid cancer diagnosis and metabolic regulation**. *Endocrine* 2019:1-14.
7. Wadlow R, Ramaswamy S: **DNA microarrays in clinical cancer research**. *Current molecular medicine* 2005, **5**(1):111-120.
8. Poller DN, Glaysher S: **Molecular pathology and thyroid FNA**. *Cytopathology* 2017, **28**(6):475-481.
9. Valerio L, Pieruzzi L, Giani C, Agate L, Bottici V, Lorusso L, Cappagli V, Puleo L, Matrone A, Viola D: **Targeted therapy in thyroid cancer: state of the art**. *Clinical oncology* 2017, **29**(5):316-324.
10. Zhang S, Wang Q, Han Q, Han H, Lu P: **Identification and analysis of genes associated with papillary thyroid carcinoma by bioinformatics methods**. *Bioscience reports* 2019, **39**(4).
11. Gao X, Wang J, Zhang S: **Integrated bioinformatics analysis of hub genes and pathways in anaplastic thyroid carcinomas**. *International journal of endocrinology* 2019, **2019**.
12. Edgar R, Domrachev M, Lash AE: **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository**. *Nucleic acids research* 2002, **30**(1):207-210.
13. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies**. *Nucleic acids research* 2015, **43**(7):e47-e47.
14. Kolde R, Laur S: **RobustRankAggreg: Methods for robust rank aggregation**. *R package version* 2013, **1**:254.
15. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT: **Gene ontology: tool for the unification of biology**. *Nature genetics* 2000, **25**(1):25-29.
16. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K: **KEGG: new perspectives on genomes, pathways, diseases and drugs**. *Nucleic acids research* 2017, **45**(D1):D353-D361.
17. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK: **Metascape provides a biologist-oriented resource for the analysis of systems-level datasets**. *Nature communications* 2019, **10**(1):1-10.
18. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP: **STRING v10: protein–protein interaction networks, integrated over the tree of life**. *Nucleic acids research* 2015, **43**(D1):D447-D452.
19. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks**. *Genome research* 2003, **13**(11):2498-2504.
20. Yu G, Wang L-G, Han Y, He Q-Y: **clusterProfiler: an R package for comparing biological themes among gene clusters**. *Omics: a journal of integrative biology* 2012, **16**(5):284-287.
21. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z: **GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses**. *Nucleic acids research* 2017, **45**(W1):W98-W102.
22. Pankov R, Yamada KM: **Fibronectin at a glance**. *Journal of cell science* 2002, **115**(20):3861-3863.

23. Cai X, Liu C, Zhang TN, Zhu YW, Dong X, Xue P: **Down-regulation of FN1 inhibits colorectal carcinogenesis by suppressing proliferation, migration, and invasion**. *Journal of cellular biochemistry* 2018, **119**(6):4717-4728.
24. Yen C-Y, Huang C-Y, Hou M-F, Yang Y-H, Chang C-H, Huang H-W, Chen C-H, Chang H-W: **Evaluating the performance of fibronectin 1 (FN1), integrin α4β1 (ITGA4), syndecan-2 (SDC2), and glycoprotein CD44 as the potential biomarkers of oral squamous cell carcinoma (OSCC)**. *Biomarkers* 2013, **18**(1):63-72.
25. Lai CJ, Tay BH: **Pharmacophore-based screening targeted at upregulated FN1, MMP-9, APP reveals therapeutic compounds for nasopharyngeal carcinoma**. *Computers in biology and medicine* 2016, **69**:158-165.
26. Ren H, Tan Z-P, Zhu X, Crosby K, Haack H, Ren J-M, Beausoleil S, Moritz A, Innocenti G, Rush J: **Identification of anaplastic lymphoma kinase as a potential therapeutic target in ovarian cancer**. *Cancer research* 2012, **72**(13):3312-3323.
27. Waalkes S, Atschekzei F, Kramer MW, Hennenlotter J, Vetter G, Becker JU, Stenzl A, Merseburger AS, Schrader AJ, Kuczyk MA: **Fibronectin 1 mRNA expression correlates with advanced disease in renal cancer**. *BMC cancer* 2010, **10**(1):503.
28. Sponziello M, Rosignolo F, Celano M, Maggisano V, Pecce V, De Rose RF, Lombardo GE, Durante C, Filetti S, Damante G: **Fibronectin-1 expression is increased in aggressive thyroid cancer and favors the migration and invasion of cancer cells**. *Molecular and cellular endocrinology* 2016, **431**:123-132.
29. Zhan S, Li J, Wang T, Ge W: **Quantitative proteomics analysis of sporadic medullary thyroid cancer reveals fn1 as a potential novel candidate prognostic biomarker**. *The oncologist* 2018, **23**(12):1415-1425.
30. Liang W, Sun F: **Identification of key genes of papillary thyroid cancer using integrated bioinformatics analysis**. *Journal of endocrinological investigation* 2018, **41**(10):1237-1245.
31. Zhao H, Li H: **Network-based meta-analysis in the identification of biomarkers for papillary thyroid cancer**. *Gene* 2018, **661**:160-168.
32. Qiu J, Zhang W, Zang C, Liu X, Liu F, Ge R, Sun Y, Xia Q: **Identification of key genes and miRNAs markers of papillary thyroid cancer**. *Biological research* 2018, **51**(1):45.
33. Liu L, He C, Zhou Q, Wang G, Lv Z, Liu J: **Identification of key genes and pathways of thyroid cancer by integrated bioinformatics analysis**. *Journal of cellular physiology* 2019, **234**(12):23647-23657.
34. Nieuwesteeg M, Willson J, Cepeda M, Fox M, Damjanovski S: **Functional characterization of tissue inhibitor of metalloproteinase-1 (TIMP-1) N-and C-terminal domains during Xenopus laevis development**. *The Scientific World Journal* 2014, **2014**.
35. Jee BK, Park KM, Surendran S, Lee WK, Han CW, Kim YS, Lim Y: **KAI1/CD82 suppresses tumor invasion by MMP9 inactivation via TIMP1 up-regulation in the H1299 human lung carcinoma cell line**. *Biochemical and biophysical research communications* 2006, **342**(2):655-661.
36. Honkavuori M, Talvensaari-Mattila A, Puistola U, Turpeenniemi-Hujanen T, Santala M: **High serum TIMP-1 is associated with adverse prognosis in endometrial carcinoma**. *Anticancer research* 2008,

28(5A):2715-2719.

37. Dechaphunkul A, Phukaoloun M, Kanjanapradit K, Graham K, Ghosh S, Santos C, Mackey JR: **Prognostic significance of tissue inhibitor of metalloproteinase-1 in breast cancer**. *International journal of breast cancer* 2012, **2012**.
38. Aaberg-Jessen C, Christensen K, Offenberg H, Bartels A, Dreehsen T, Hansen S, Schrøder HD, Brünner N, Kristensen BW: **Low expression of tissue inhibitor of metalloproteinases-1 (TIMP-1) in glioblastoma predicts longer patient survival**. *Journal of neuro-oncology* 2009, **95**(1):117-128.
39. Jackson HW, Defamie V, Waterhouse P, Khokha R: **TIMPs: versatile extracellular regulators in cancer**. *Nature Reviews Cancer* 2017, **17**(1):38.
40. Adorno-Cruz V, Liu H: **Regulation and functions of integrin α2 in cell adhesion and disease**. *Genes & diseases* 2019, **6**(1):16-24.
41. Plow E, Haas T, Zhang LL: **J. and Smith, JW 2000**. *Ligand binding to*.
42. Tuckwell D, Calderwood DA, Green LJ, Humphries MJ: **Integrin alpha 2 I-domain is a binding site for collagens**. *Journal of cell science* 1995, **108**(4):1629-1637.
43. Chuang Y-C, Wu H-Y, Lin Y-L, Tzou S-C, Chuang C-H, Jian T-Y, Chen P-R, Chang Y-C, Lin C-H, Huang T-H: **Blockade of ITGA2 induces apoptosis and inhibits cell migration in gastric cancer**. *Biological procedures online* 2018, **20**(1):10.
44. Zhang L, Huang Y, Ling J, Zhuo W, Yu Z, Luo Y, Zhu Y: **Is integrin subunit alpha 2 expression a prognostic factor for liver carcinoma? A validation experiment based on bioinformatics analysis**. *Pathology & Oncology Research* 2019, **25**(4):1545-1552.
45. Chin SP, Marthick JR, West AC, Short AK, Chuckowree J, Polanowski AM, Thomson RJ, Holloway AF, Dickinson JL: **Regulation of the ITGA2 gene by epigenetic mechanisms in prostate cancer**. *The Prostate* 2015, **75**(7):723-734.
46. Pan H, Wanami LS, Dissanayake TR, Bachelder RE: **Autocrine semaphorin3A stimulates alpha2 beta1 integrin expression/function in breast tumor cells**. *Breast cancer research and treatment* 2009, **118**(1):197.
47. Gong J, Lu X, Xu J, Xiong W, Zhang H, Yu X: **Coexpression of UCA1 and ITGA2 in pancreatic cancer cells target the expression of miR-107 through focal adhesion pathway**. *Journal of cellular physiology* 2019, **234**(8):12884-12896.
48. Zhao X, Wu Y, Lv Z: **miR-128 modulates hepatocellular carcinoma by inhibition of ITGA2 and ITGA5 expression**. *American journal of translational research* 2015, **7**(9):1564.
49. Chernaya G, Mikhno N, Khabalova T, Svyatchenko S, Mostovich L, Shevchenko S, Gulyaeva L: **The expression profile of integrin receptors and osteopontin in thyroid malignancies varies depending on the tumor progression rate and presence of BRAF V600E mutation**. *Surgical oncology* 2018, **27**(4):702-708.
50. Zhai T, Muhanhali D, Jia X, Wu Z, Cai Z, Ling Y: **Identification of gene co-expression modules and hub genes associated with lymph node metastasis of papillary thyroid cancer**. *Endocrine* 2019, **66**(3):573-584.

51. Ashman LK: **The biology of stem cell factor and its receptor C-kit**. *The international journal of biochemistry & cell biology* 1999, **31**(10):1037-1051.
52. Matsumura Y, Umemura S, Ishii G, Tsuta K, Matsumoto S, Aokage K, Hishida T, Yoshida J, Ohe Y, Suzuki H: **Expression profiling of receptor tyrosine kinases in high-grade neuroendocrine carcinoma of the lung: a comparative analysis with adenocarcinoma and squamous cell carcinoma**. *Journal of cancer research and clinical oncology* 2015, **141**(12):2159-2170.
53. Yu G, Yin C, Jiang L, Zheng Z, Wang Z, Wang C, Zhou H, Jiang X, Liu Q, Meng F: **Amyloid precursor protein cooperates with c-KIT mutation/overexpression to regulate cell apoptosis in AML1-ETO-positive leukemia via the PI3K/AKT signaling pathway**. *Oncology reports* 2016, **36**(3):1626-1632.
54. Chen EC, Karl TA, Kalisky T, Gupta SK, O'Brien CA, Longacre TA, van de Rijn M, Quake SR, Clarke MF, Rothenberg ME: **KIT signaling promotes growth of colon xenograft tumors in mice and is up-regulated in a subset of human colon cancers**. *Gastroenterology* 2015, **149**(3):705-717. e702.
55. Lau S, Hansford L, Chan W, Chan GC, Wan TS, Wong KK, Kaplan D, Tam PK, Ngan ES: **Prokineticin signaling is required for the maintenance of a de novo population of c-KIT+ cells to sustain neuroblastoma progression**. *Oncogene* 2015, **34**(8):1019-1034.
56. Tramm T, Kim J-Y, Leibl S, Moinfar F, Tavassoli FA: **Expression of C-KIT, CD24, CD44s, and COX2 in benign and non-invasive apocrine lesions of the breast**. *Virchows Archiv* 2016, **469**(3):285-295.
57. Dahl C, Abildgaard C, Riber-Hansen R, Steiniche T, Lade-Keller J, Guldberg P: **KIT is a frequent target for epigenetic silencing in cutaneous melanoma**. *Journal of Investigative Dermatology* 2015, **135**(2):516-524.
58. Broecker-Preuss M, Sheu S-Y, Worm K, Feldkamp J, Witte J, Scherbaum W, Mann K, Schmid K, Schott M: **Expression and mutation analysis of the tyrosine kinase c-kit in poorly differentiated and anaplastic thyroid carcinoma**. *Hormone and metabolic research* 2008, **40**(10):685-691.
59. Murakawa T, Tsuda H, Tanimoto T, Tanabe T, Kitahara S, Matsubara O: **Expression of KIT, EGFR, HER-2 and tyrosine phosphorylation in undifferentiated thyroid carcinoma: Implication for a new therapeutic approach**. *Pathology international* 2005, **55**(12):757-765.
60. TANAKA T, UMEKI K, YAMAMOTO I, KOTANI T, SAKAMOTO F, NOGUCHI S, OHTAKI S: **c-Kit Proto-oncogene Is More Likely to Lose Expression in Differentiated Thyroid Carcinoma Than Three Thyroid-specific Genes**. *Endocrine journal* 1995, **42**(5):723-728.
61. Hemmings BA, Restuccia DF: **Pi3k-pkb/akt pathway**. *Cold Spring Harbor perspectives in biology* 2012, **4**(9):a011189.
62. Motti ML, Califano D, Troncone G, De Marco C, Migliaccio I, Palmieri E, Pezzullo L, Palombini L, Fusco A, Viglietto G: **Complex regulation of the cyclin-dependent kinase inhibitor p27kip1 in thyroid cancer cells by the PI3K/AKT pathway: regulation of p27kip1 expression and localization**. *The American journal of pathology* 2005, **166**(3):737-749.
63. Larson SD, Jackson LN, Riall TS, Uchida T, Thomas RP, Qiu S, Evers BM: **Increased incidence of well-differentiated thyroid cancer associated with Hashimoto thyroiditis and the role of the PI3k/Akt pathway**. *Journal of the American College of Surgeons* 2007, **204**(5):764-773.

64. Hou P, Liu D, Shan Y, Hu S, Studeman K, Condouris S, Wang Y, Trink A, El-Naggar AK, Tallini G: **Genetic alterations and their relationship in the phosphatidylinositol 3-kinase/Akt pathway in thyroid cancer**. *Clinical cancer research* 2007, **13**(4):1161-1170.
65. Andersen MK, Rise K, Giskeødegård GF, Richardsen E, Bertilsson H, Størkersen Ø, Bathen TF, Rye M, Tessem M-B: **Integrative metabolic and transcriptomic profiling of prostate cancer tissue containing reactive stroma**. *Scientific reports* 2018, **8**(1):1-11.
66. Hu K, Chen F: **Identification of significant pathways in gastric cancer based on protein-protein interaction networks and cluster analysis**. *Genetics and molecular biology* 2012, **35**(3):701-708.
67. Bao Y, Wang L, Shi L, Yun F, Liu X, Chen Y, Chen C, Ren Y, Jia Y: **Transcriptome profiling revealed multiple genes and ECM-receptor interaction pathways that may be associated with breast cancer**. *Cellular & molecular biology letters* 2019, **24**(1):38.
68. Huang Y, Tao Y, Li X, Chang S, Jiang B, Li F, Wang ZM: **Bioinformatics analysis of key genes and latent pathway interactions based on the anaplastic thyroid carcinoma gene expression profile**. *Oncology letters* 2017, **13**(1):167-176.

Figures

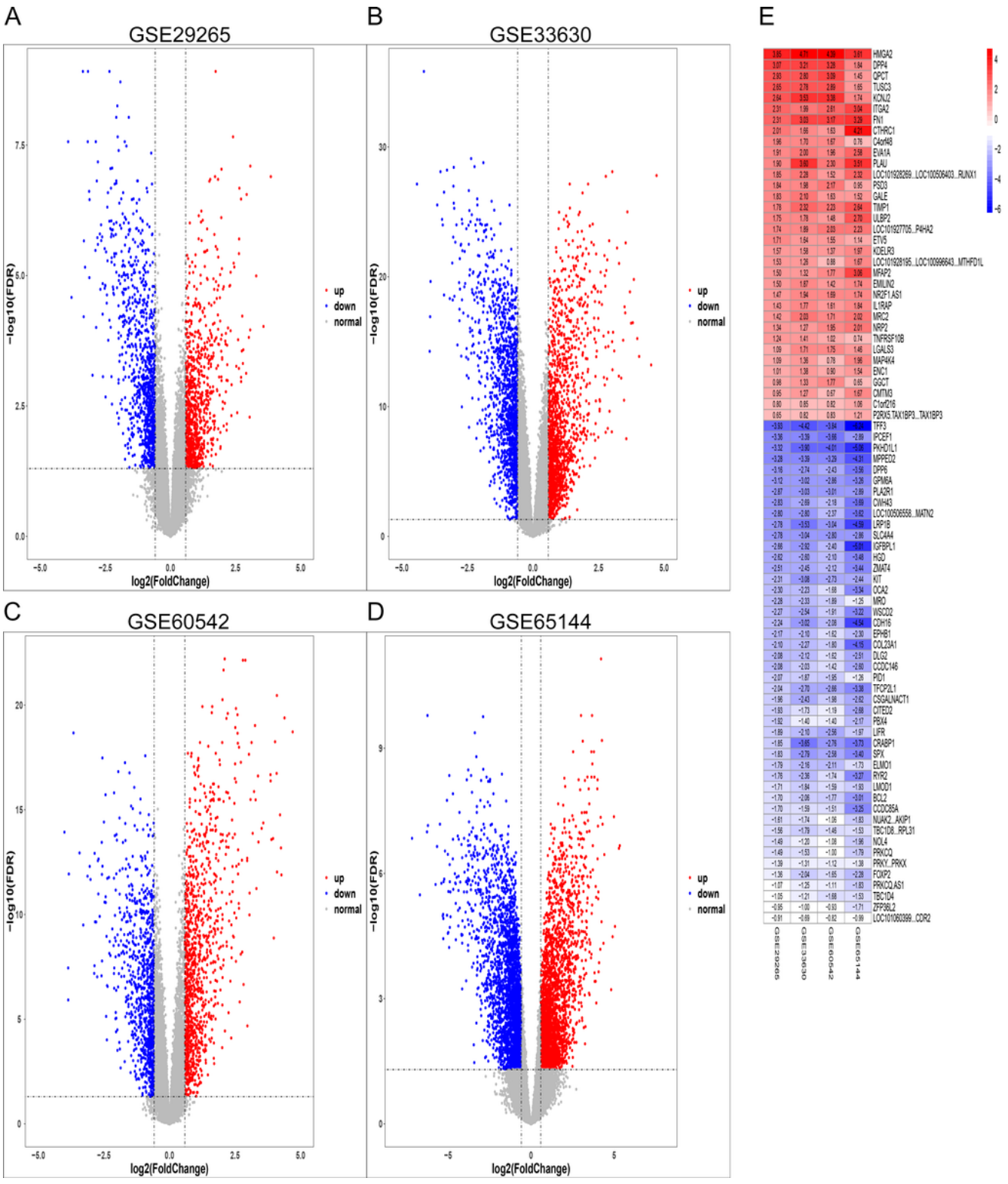

Figure 1

Identification of the DEGs in TC and data integration. (A) Volcano plot of GSE29265, (B) volcano plot of GSE33630, (C) volcano plot of GSE60542, (D) volcano plot of GSE65144, and (E) the cluster heatmap of the DEGs. Red, white, and blue represent higher expression levels, no expression differences, and lower expression levels, respectively. The rows and columns of the heatmap represent genes and datasets,

respectively, and the numbers in it represent the logFC between TC tissue and normal ones from each GEO dataset.

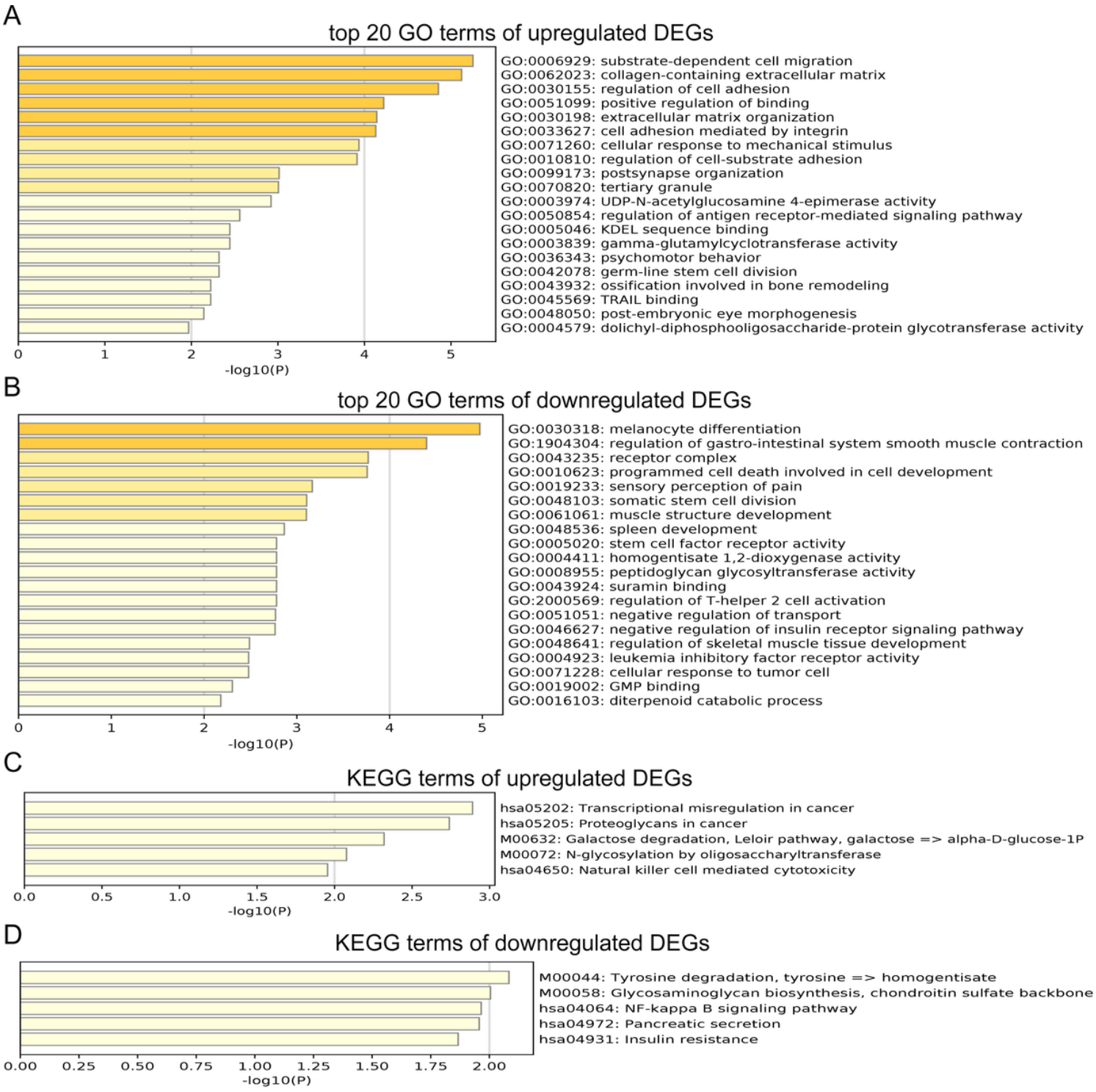

Figure 2

GO and KEGG pathway enrichment analysis. (A) Bar plot for top 20 GO terms of upregulated DEGs, (B) bar plot for top 20 GO terms of downregulated DEGs, (C) bar plot for KEGG terms of upregulated DEGs, and (D) bar plot for KEGG terms of downregulated DEGs.

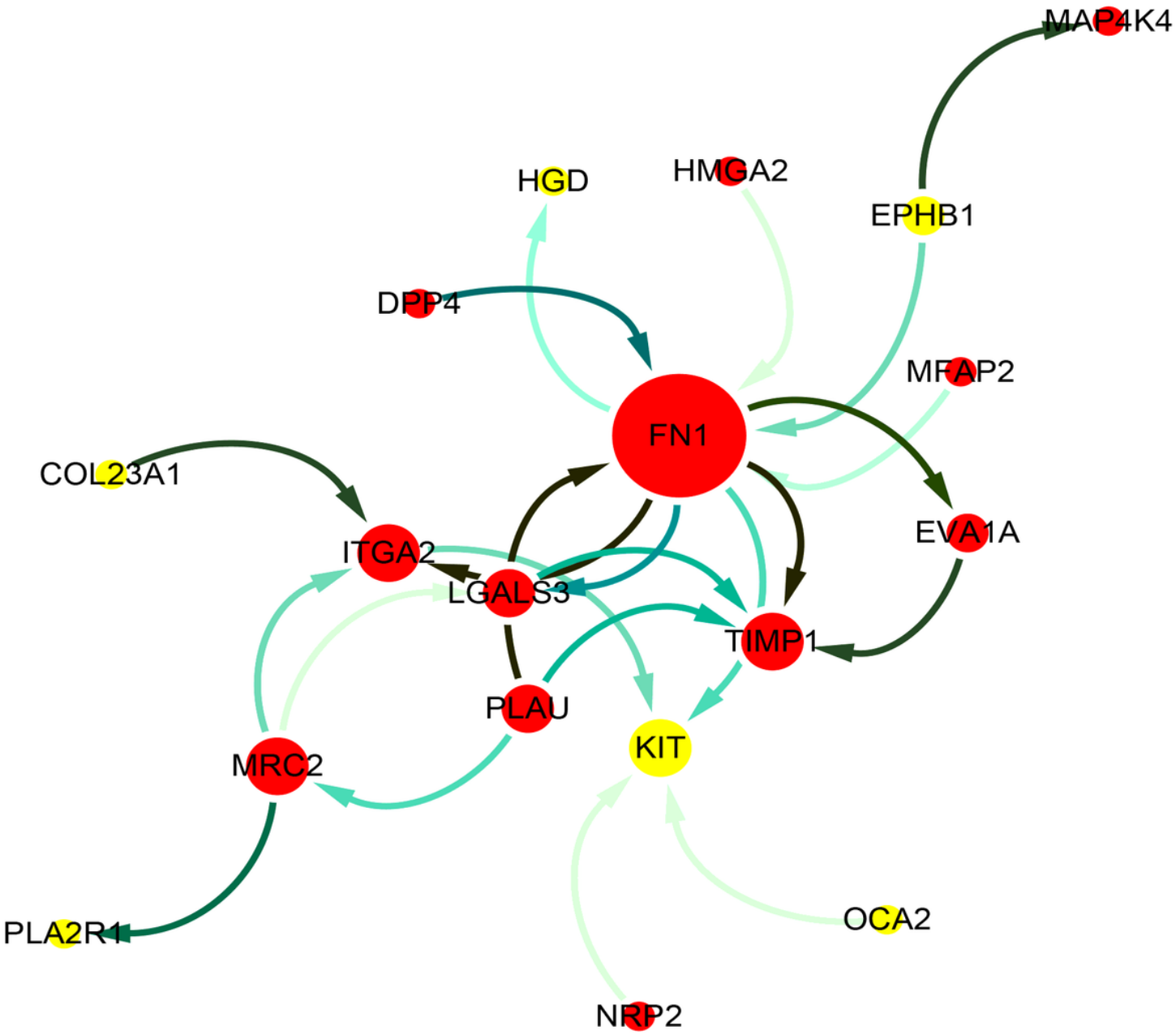

Figure 3

The PPI network of the DEGs. Red nodes and yellow nodes represent upregulated genes and downregulated genes, respectively. The magnitude of the degree is positively correlated with the magnitude of a node, and the size of the interaction score between genes was positively correlated with the depth of edge color.

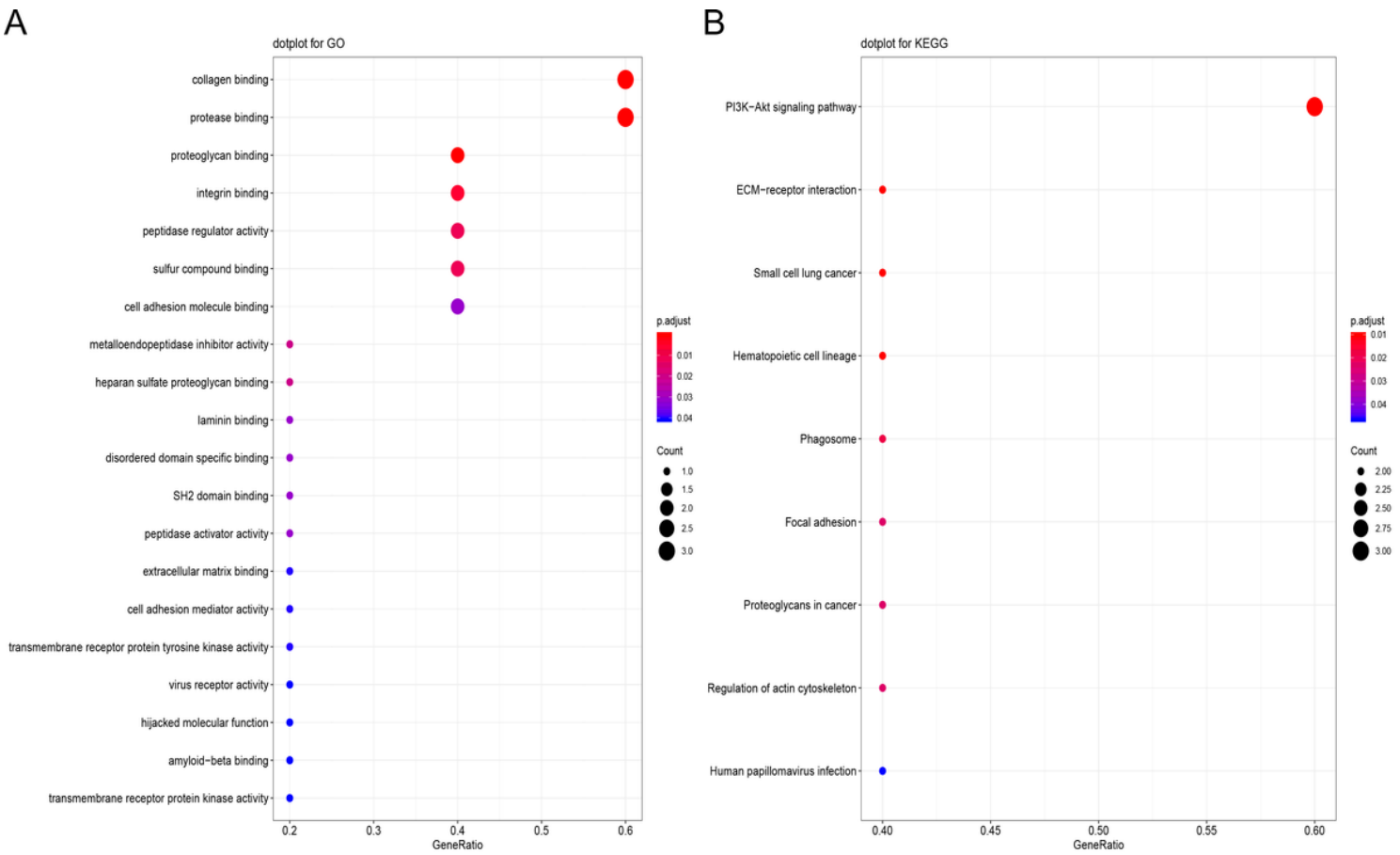

Figure 4

Enrichment analysis of 5 hub genes. (A) Dot plot for top 10 GO terms of hub genes, (B) dot plot for KEGG terms of hub genes.

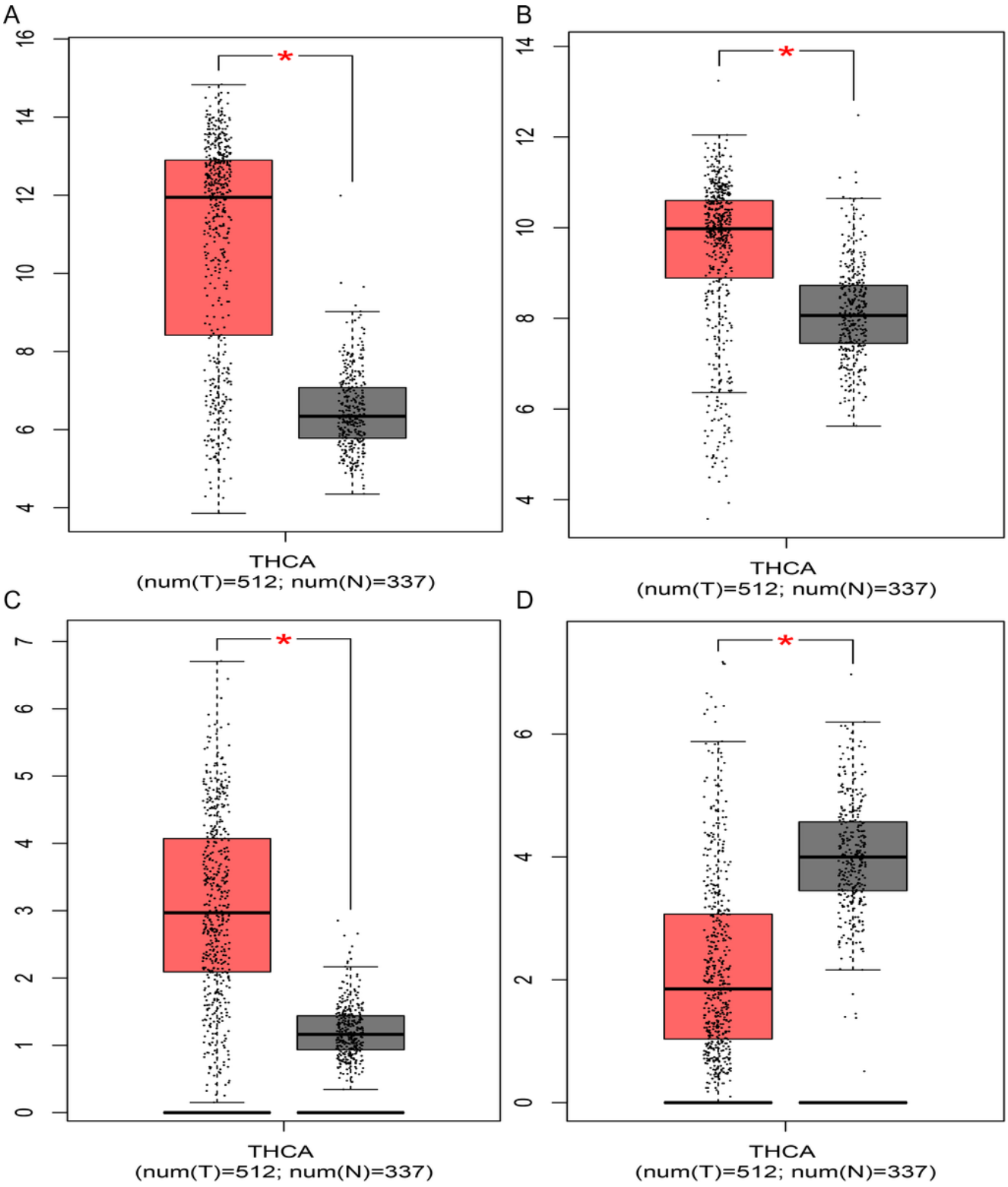

Figure 5

Expression level analysis of 4 hub genes. Red represents cancer and gray represents normal. (A) FN1, (B)TIMP1, (C)ITGA2, and (D) KIT. *P<0.01.

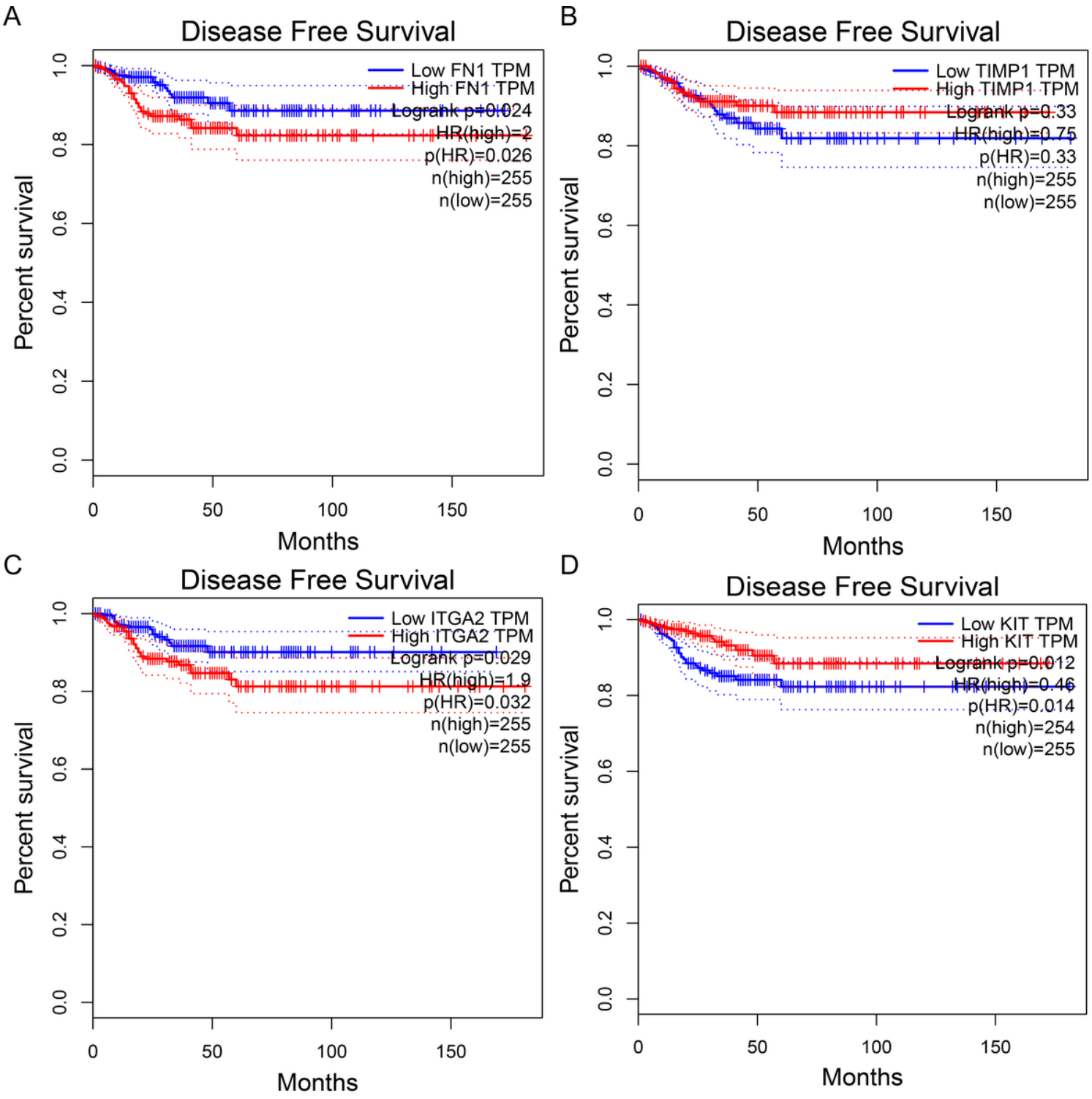

Figure 6

Survival analysis of 4 hub genes. TC patients were subdivided into high/low gene expression groups based on the median expression level of each gene in TC tissues. (A) DFS analysis of FN1, (B) DFS analysis of TIMP1, (C) DFS analysis of ITGA2, and (D) DFS analysis of KIT.

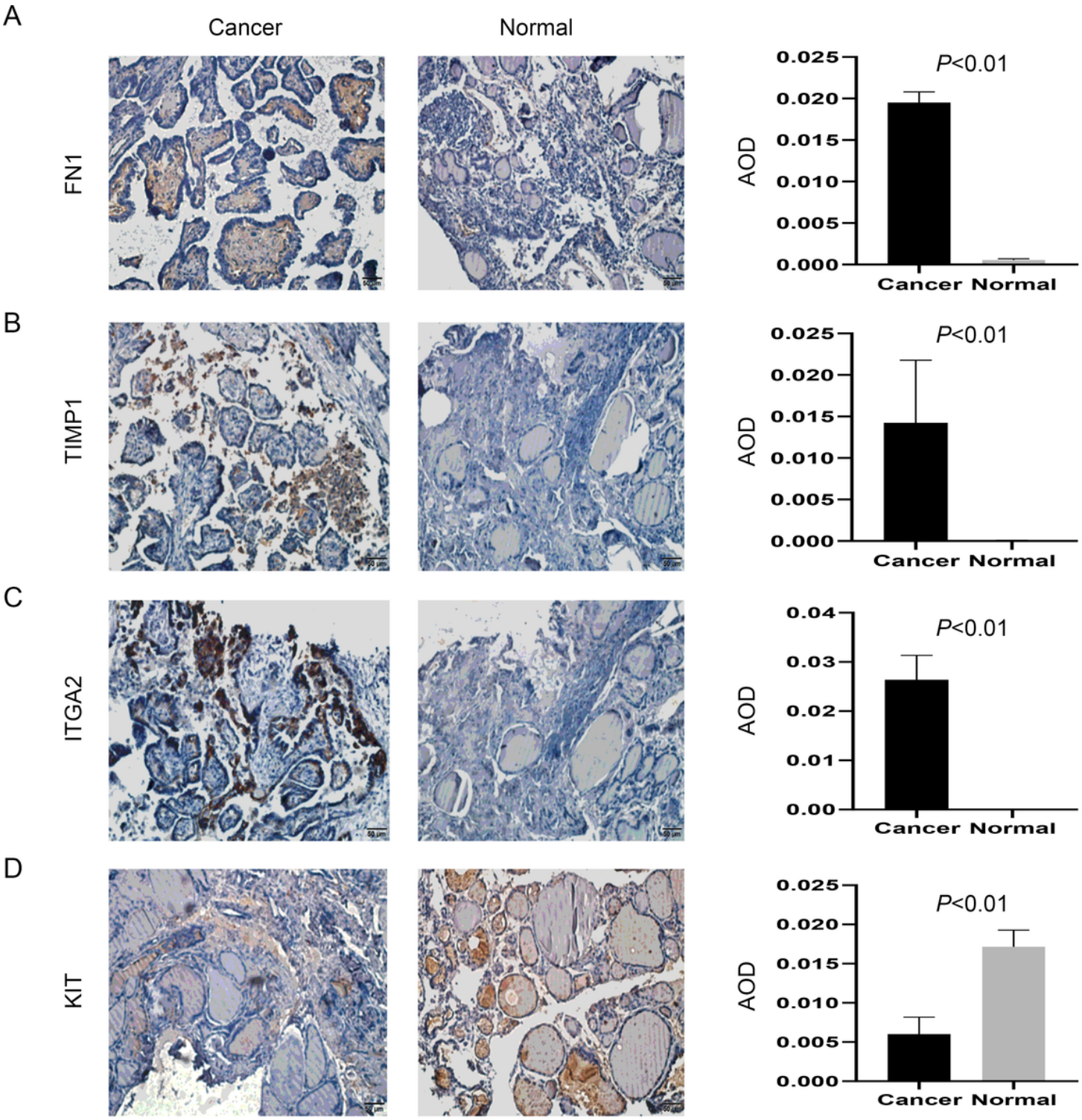

Figure 7

Immunohistochemical staining results of the protein expressions of hub genes from the clinical specimens. Brown staining represents expression of 4 protein and blue staining represents nucleus. Bar: 50um. (A) IHC of FN1, (B) IHC of TIMP1, (C) IHC of ITGA2, and (D) IHC of KIT.

Supplementary Files

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

- SupplementalTableS2.docx
- SupplementalTableS1.docx