Identification of Potential Key Genes in Anaplastic Thyroid Cancer Using Bioinformatics AnalysisIdentification of Potential Key Genes in Anaplastic Thyroid Cancer using Bioinformatics Analysis

Zhenzhen Li
Xi'an Jiaotong University Second Affiliated Hospital
Chaoliang Xiong
Xi'an Jiaotong University Second Affiliated Hospital
Jin Wei
Xi'an Jiaotong University Second Affiliated Hospital
Ping Chen
Institute of Endemic Diseases
Yanping Zhang
Xi'an Jiaotong University Second Affiliated Hospital
Huange Zhu
Xi'an Jiaotong University Second Affiliated Hospital
Yuqi Zhao
Xi'an Jiaotong University Second Affiliated Hospital
Wei Lu
Xi'an Jiaotong University Second Affiliated Hospital
Qian He
Xi'an Jiaotong University Second Affiliated Hospital
Yan Geng
Xi'an Jiaotong University Second Affiliated Hospital
Jianhong Zhu (✉ zjh13389212293@163.com)
Xinjiang Medical University Affiliated Second Hospital https://orcid.org/0000-0002-2383-5479

Research

Keywords: Anaplastic thyroid cancer, Gene Expression Omnibus database, Bioinformatics analysis, Differentially expressed genes, Potential key genes

Posted Date: March 23rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-328857/v1

Abstract

Background

Anaplastic thyroid cancer (ATC) has a high degree of malignancy and a poor prognosis. Its incidence accounts for approximately 10-15% of all thyroid cancers. The purpose of this study was to determine the differentially expressed genes (DEGs) of ATC through biometric analysis technology, clarify the potential interactions between them, and screen genes related to the prognosis of ATC.

Methods

The GSE29265, GSE65144, GSE33630, and GSE85457 expression profiles downloaded from the Gene Expression Omnibus database (GEO) contained a total of 117 tissue samples (81 normal thyroid tissue samples and 36 ATC samples). The four datasets were integrated and analyzed by the limma packages to obtain DEGs. With these DEGs, we performed gene ontology functional annotation and Kyoto Encyclopedia of Genes and Genomes pathway analyses using the Database for Annotation, Visualization and Integrated Discovery, protein-protein interaction (PPI) analysis using Cytoscape, and survival analysis using the Kaplan-Meier (KM) plotter.

Results.

After R integration analysis of the four datasets, 764 DEGs were obtained, i.e., 314 upregulated and 450 downregulated genes. Among the hub DEGs obtained in the PPI network, the expression levels of thymidylate synthase (TYMS), fibronectin 1, chordin-like 1, syndecan 2, integrin alpha 2, collagen type I alpha 1 chain, collagen type IX alpha 3 chain (COL9A3), and collagen type XXIII alpha 1 chain (COL23A1) were associated with ATC prognosis. These results showed that the overall survival and recurrence-free survival of TYMS, COL9A3, and COL23A1 were statistically significant in our KM plotter survival analysis; thus, these DEGs may be used as potential biomarkers of ATC.

Conclusion

This study identified several potential target genes and pathways that may affect the development of ATC. These findings provide new insights for the detection of novel diagnostic and therapeutic biomarkers for ATC.

Background

Thyroid cancer is the most common endocrine malignancy, accounting for 1.6% of all malignant tumors [1, 2]. In recent years, the incidence of thyroid cancer has rapidly increased[2]. The degree of malignancy of anaplastic thyroid cancer (ATC) is the highest among all types of thyroid cancers and is difficult to detect at an early stage[3]. The advanced stage (stage IV) can be divided into stages IVa-IVc [4]. Treatment of ATC is difficult as many patients have distant metastasis at the time of diagnosis, and effective treatment options are limited[5]. Given that patients with ATC still require long-term

chemotherapy and radiotherapy to prevent recurrence and deterioration even after thyroid resection and that the proposed target drugs for ATC have not yet been used clinically, we urgently need to identify molecular markers that can evaluate the prognosis and survival of ATC as well as speed up the research for its molecular mechanism.

With the rapid development of microarray analysis technology in recent years, the microarray data of many tumors have been obtained and uploaded to public databases by researchers. Gene Expression Omnibus (GEO) is an open database that stores large quantities of cancer microarray information (https://www.ncbi.nlm.nih.gov/geo/). Although the amount of data in GEO is rich, there is no definite relationship between the contents. Therefore, it is necessary to systematically analyze the ATC gene chip information in GEO using biological information analysis technology and screen the differentially expressed genes (DEGs) that are relevant to the pathogenesis and prognosis of ATC.

In this study, specific relationships between various DEGs and ATC were further clarified. Through a comprehensive utilization of various biological analysis technologies and databases, molecules and molecular pathways related to ATC development were screened, and biomarkers related to the prognosis and survival of ATC were identified.

Methods

Microarray data

The gene expression profiles (GSE29265, GSE65144, GSE33630, and GSE85457) analyzed in this study were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). All the above profiles were based on the GPL570 (Affymetrix Human Genome U133 Plus 2.0) platform.

The GSE29265 (including 20 normal thyroid tissue samples and 9 ATC samples), GSE65144 (including 13 normal thyroid tissue samples and 12 ATC samples), GSE33630 (including 45 normal thyroid tissue samples and 11 ATC samples) and GSE85457 (including 3 normal thyroid tissue samples and 4 ATC samples) gene expression profile matrix files were downloaded. The dataset information is provided in Table .

Microarray integration

All samples (n = 117) from the four profiles were separated into two groups, which included normal thyroid tissue samples (n = 81) and ATC samples (n = 36). The CEL files were first converted into probe expression values and preprocessed for background adjustment and quantile normalization using the robust multiarray average (RMA) algorithm with the ‘affy’ package in R (version 4.0.0; https://www.r-project.org/). In addition, the batch effects between different datasets were removed using the combat function with the ‘sva’ package.

The annotation file (HG-U133_Plus_2.na36.annot) of Affymetrix Human Genome U133 Plus 2.0 Array was downloaded from the official website (www.affymetrix.com/support/technical/byproduct.affx?

product=hg-u133-plus) to transform the probe-level data of genes into expression values. If a gene symbol was matched by multiple probes, then the average expression value was calculated for this gene.

The 'limma' package in R (version 4.0.0) was used to identify DEGs between ATC tissues and normal tissues.

Identification of DEGs

We defined |log fold change (FC)| ≥ 2 and adjusted *P*-value < 0.01 as the DEG screening criteria for the ATC samples from the four microarray datasets and created DEG volcano maps in R (version 4.0.0). In addition, a .txt file of DEG lists, including the upregulated and downregulated genes sorted by log (FC), was saved for further analysis. All R packages used in our research were arranged in R software (version 4.0.0).

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of DEGs

Database for Annotation, Visualization, and Integrated Discovery (DAVID) 6.8 (https://david.ncifcrf.gov) is an online bioinformatics database for gene functional analysis and annotation of biological functions for large-scale gene or protein lists. DAVID was used for GO and KEGG pathway enrichment analyses of DEGs. Differences were considered statistically significant at a *P*-value < 0.05.

Protein-protein interaction (PPI) network construction, hub DEG identification, and module analysis

The STRING database (http://string-db.org) was used to analyze the direct (physical) and indirect (functional) associations between different genes to construct a PPI network. DEGs were imported into the STRING database to evaluate interactive relationships, with a combined score > 0.7 defined as significant. Subsequently, the PPI network was visualized using Cytoscape software (www.cytoscape.org/). Moreover, using a Cytoscape plugin called Molecular Complex Detection (MOCDE), we selected a MOCDE score > 7, and the DEGs from the selected MOCDE in the gene expression network were identified as the potential hub DEGs. Significantly different genes were selected as potential hub DEGs using the University of Alabama Cancer (UALCAN) database (http://ualcan.path.uab.edu/cgi-bin/TCGAExHeatMap2KK.pl).

Verification of hub DEGs in The Cancer Genome Atlas (TCGA) datasets

Gene Expression Profiling Interactive Analysis (GEPIA) is a web-based tool (http://gepia.cancer-pku.cn) that delivers fast and customizable functionalities based on TCGA data. In this study, we used GEPIA to verify the hub DEGs, and the expression levels of hub DEGs in TCGA datasets were demonstrated via boxplots. The following were set as the screening criteria: |log (FC)| ≥ 1 and *P* < 0.05.

Survival analysis of hub DEGs

To further investigate the prognostic values of hub DEGs in ATC prognosis, we used the Kaplan-Meier (KM) plotter (http://kmplot.com/analysis/) for survival and statistical analyses. Differences were statistically significant at $P < 0.05$. Because of their importance in the prognosis of ATC, these hub DEGs were classified as key genes.

Results

Identification of DEGs

Following integration of the GSE29265, GSE65144, GSE33630, and GSE85457 profiles, the datasets were unevenly distributed before standardization (Figure S1A), whereas after standardization, the dataset distribution was clearly homogenized (Figure S1B). As observed from the principal coordinate analysis (PCA) diagram, the group factor was the batch effect between different datasets (Figure S1C), which was removed using the Combat package (Figure S1D).

Identification of integrated DEGs

A total of 776 integrated DEGs including 315 upregulated genes and 461 downregulated genes were identified using the limma package. Figure 1 shows the DEGs.

GO function and KEGG pathway enrichment analyses of the DEGs

The DAVID dataset showed the GO function and KEGG pathway enrichment analyses for the DEGs. The enriched GO terms were divided into the biological process (BP), cell composition (CC), and molecular function (MF). As shown in Fig. 2 and Table , the upregulated DEGs were mainly enriched in BP (including extracellular matrix organization, mitotic nuclear division, and collagen catabolic processes), in CC (including collagen trimer, proteinaceous extracellular matrix, and extracellular space), and in MF (including integrin binding and microtubule binding). The downregulated DEGs were mainly enriched in BP (including extracellular matrix organization and axon genesis), in CC (including integral component of the membrane and extracellular exosome), and in MF (including selenium binding and glutathione transferase activity). The KEGG pathway analysis included upregulated and downregulated pathways. As shown in Fig. 3 and Table , the KEGG pathways of the upregulated DEGs were mainly extracellular matrix (ECM)-receptor interaction, protein digestion and absorption, and focal adhesion, while the KEGG pathways of the downregulated DEGs were mainly thyroid hormone synthesis, protein digestion and absorption, and gastric acid secretion.

PPI analysis of DEGs and hub gene identification

In this study, the PPI network of DEGs was constructed using Cytoscape software and the STRING database. The 411 nodes and 3001 edges of the PPI network of DEG expression (confidence score of ≥ 0.7) are presented in Fig. 4A. Four significant modules were screened from the PPI network of DEGs using the MCODE (MCODE ≥ 7) plugin of Cytoscape (Fig. 4B). After being identified by the UALCAN database,

19 hub DEGs including 13 upregulated and 6 downregulated DEGs with statistical significance ($P < 0.05$) were screened from the four MCODE modules (Fig. 5A and Table).

Verification of hub DEGs in TCGA data

To validate the reliability of the key genes, we used GEPIA. The hub DEGs were differentially expressed between normal thyroid tissue samples and thyroid cancer samples (Fig. 6). The expression levels of thymidylate synthase (TYMS), collagen type I alpha 1 chain (COL1A1), fibronectin 1 (FN1), integrin alpha 2 (ITGA2), collagen type IX alpha 3 chain (COL9A3), collagen type XXIII alpha 1 chain (COL23A1), syndecan 2 (SDC2), and chordin-like 1 (CHRDL1) in our study were consistent with those in the UALCAN database.

Survival analysis of hub DEGs

To investigate the prognostic values of these hub DEGs, we used the KM plotter bioinformatics analysis platform. It was found that recurrence-free survival (RFS) was lower in the high TYMS ($P = 0.00076$), COL1A1 ($P = 0.047$), FN1 ($P = 7e-05$), and ITGA2 ($P = 0.0027$) expression groups than in the respective low expression groups, while the low COL9A3 ($P = 0.047$), COL23A1 ($P = 0.0094$), SDC2 ($P = 5.1e-05$), and CHRDL1 ($P = 0.018$) expression groups had higher RFS than the respective high expression groups.

As shown in Fig. 7, the other hub DEGs were not significantly associated with the prognosis of thyroid cancer ($P > 0.05$). Hence, we derived eight key genes related to the prognosis of thyroid cancer: TYMS, COL1A1, FN1, ITGA2, COL9A3, COL23A1, SDC2, and CHRDL1. However, only the overall survival (OS) analysis values of TYMS ($P = 0.0057$), COL9A3 ($P = 0.05$), and COL23A1 ($P = 0.027$) were statistically significant (Fig. 8). It can be seen from the above-mentioned results that the OS and RFS of TYMS, COL9A3, and COL23A1 were statistically significant; thus, these DEGs may be used as potential biomarkers of ATC.

Discussion

As the primary subtype of thyroid cancer, ATC remains one of the deadliest diseases[6]. Because of the variability in ATC genes, ATC can easily metastasize and has a lack of valid therapeutic targets[7]. Therefore, at present, chemotherapy combined with thyroid surgery is the primary adjunct therapy for ATC[8]. However, these therapies have some limitations and tend to result in poor prognosis[9]. Therefore, it is necessary to reevaluate the etiology and molecular mechanisms of ATC as well as explore new potential diagnostic, therapeutic, and prognostic targets using bioinformatics analysis techniques.

Based on four GEO databases for gene expression, a total of 764 DEGs including 314 upregulated and 450 downregulated genes were screened using R integration analysis. In order to investigate the biological roles of these DEGs, we performed GO and KEGG pathway analyses. GO term enrichment analysis showed that upregulated DEGs mainly participated in cell division, while the downregulated DEGs were enriched in integral components of the membrane. KEGG pathway analysis indicated that

upregulated DEGs mainly participated in ECM-receptor interaction and protein digestion and absorption, whereas the downregulated DEGs were enriched in thyroid hormone synthesis. Next, we constructed a PPI network using the DEGs and screened potential hub DEGs using MOCDE (MOCDE ≥ 7). After UALCAN ($P < 0.05$) screening and verification by the GEPIA website, we identified several hub DEGs including TYMS, FN1, CHRDL1, SDC2, ITGA2, COL1A1, COL9A3, and COL23A1. Finally, we used the KM plotter tool to predict the prognosis of hub DEGs in patients with ATC.

Based on the KM plotter, high expression levels of TYMS, FN1, CHRDL1, and SDC2 as well as low expression levels of ITGA2, COL1A1, COL9A3, and COL23A1 were associated with RFS in patients with ATC (Fig. 7); however, only TYMS, COL9A3, and COL23A1 were significant for OS (Fig. 8). Therefore, we found that a high expression level of TYMS and low expression levels of COL9A3 and COL23A1 were related to favorable prognostic factors in patients with ATC.

TYMS is a protein coding gene that contributes to the de novo mitochondrial thymidylate biosynthesis pathway [10–12]. This function maintains the thymidine-5-prime monophosphate pool critical for DNA replication and repair [10]. Among its related pathways are one carbon pool by folate and the E2F transcription factor network [13]. GO annotations related to this gene include protein homodimerization activity and mRNA binding. Microarray and immunohistochemical studies have shown that the expression of this enzyme is significantly upregulated in a variety of tumors, including breast, bladder, cervical, kidney, lung, and gastrointestinal cancers [13–17]. The high expression of TYMS was also associated with poor clinical prognosis of these cancers, suggesting that TYMS may act as an oncogene. In fact, ectopic expression of TYMS in xenograft models has been shown to confer transformed and tumorigenic phenotypes on normal cells. It is worth noting that the elevated expression level of TYMS also showed greater invasion and metastasis ability in these cells. Protein enzymes have been of interest as targets for cancer chemotherapeutic agents. ATC is also a metastatic and aggressive cancer. TYMS belongs to the hub DEGs in our ATC data module 1. Therefore, TYMS is likely to have a strong relationship with aggressive and metastatic ATC.

COL9A3 encodes one of the three alpha chains of type IX collagen, which is the major collagen component of hyaline cartilage [18]. Diseases associated with COL9A3 include multiple epiphyseal dysplasia type 3 and intervertebral disc disease [19, 20]. Among its related pathways are the integrin pathway and gastric cancer network 2[20]. COL9A3 also participates in the protein digestion and absorption pathway [21]. Studies have shown that COL9A3 is highly expressed in the brain, retina, salivary glands, and thyroid gland but remains low in ATC [21, 22]. Therefore, we infer that COL9A3 has a certain relationship with ATC that can be used as a target for ATC diagnosis or prognosis prediction.

COL23A1 (Collagen Type XXIII Alpha 1 Chain) is a Protein Coding gene,which is a member of the transmembrane collagens, a subfamily of the nonfibrillar collagens that contain a single pass hydrophobic transmembrane domain[23]. It is a new transmembrane collagen found in metastatic tumor cells and highly expressed in thyroid and cardiovascular systems [24, 25]. Among its related pathways are the integrin pathway and degradation of the extracellular matrix. From the analysis results in this

study, it can be seen that COL23A1 is derived from module 2. COL23A1 has an important paralog of this gene is COL5A1.The KEGG results showed that COL5A1 is related to the ECM-receptor interaction, protein digestion and absorption, focal adhesion, amoebiasis, and phosphoinositide 3-kinase (PI3K)-protein kinase B (Akt) signaling pathways. These cellular pathways indicate that the body is in a state of accelerated metabolism, and the PI3K-Akt signaling pathway is a typical tumor metabolism pathway [26, 27]. In recent years, several studies have shown that COL23A1 is closely related to renal cell carcinoma, head and neck cancer, and so on [28, 29]. Based on the data mining results in this study, we believe that COL23A1 is closely related to the occurrence and prognosis of ATC.

Considering all the existing research results, we speculate that the eight hub DEGs have significant expression levels or distinct pro-cancer effects in different cancers. Detecting the levels of these genes may be useful in the early diagnosis or prognosis assessment of some cancers.

Based on multiple datasets and thorough bioinformatics analysis, an eight-gene cohort shows a superior prediction of prognosis and survival of ATC. However, there are limitations existing in the present study that require acknowledgment. Such as, lack of experimental verification, which will be the focus of our later work.”

Conclusions

Here, we used a series of bioinformatics analysis methods to screen TYMS, COL9A3, and COL23A1 as key genes and identified pathways involved in ATC initiation and progression from four gene expression profiles consisting of normal and ATC samples. Our results provide a more detailed molecular mechanism for the development of ATC, shedding light on potential biomarkers and therapeutic targets. However, the interaction mechanism and functions of the various genes need to be confirmed via additional experiments.

Abbreviations

ATC: Anaplastic thyroid cancer;

DEGs: differentially expressed genes;

GEO: Gene Expression Omnibus database;

KEGG: Kyoto Encyclopedia of Genes and Genomes;

GO: gene ontology;

PPI: protein-protein interaction;

DAVID: Database for Annotation, Visualization and Integrated Discovery;

KM: Kaplan-Meier;

TYMS: expression levels of thymidylate synthase;

COL9A3: collagen type IX alpha 3 chain;

COL23A1: collagen type XXIII alpha 1 chain;

RMA: robust multiarray average;

MOCDE: Molecular Complex Detection;

UALCAN: University of Alabama Cancer;

GEPIA: Gene Expression Profiling Interactive Analysis;

BP: biological process;

CC: cell composition;

MF: molecular function;

COL1A1: collagen type I alpha 1 chain;

FN1: fibronectin 1;

ITGA2: integrin alpha 2;

SDC2: syndecan 2;

CHRDL1: chordin-like 1;

RFS: recurrence-free survival;

OS: overall survival

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analyzed during the present study are available in the Gene Expression Omnibus repository (https://www.ncbi.nlm.nih.gov/geo/).

Competing interests

The authors declare that they have no conflict of interest.

Funding

This work was supported by the National Natural Science Foundation of China (Program No. 81870298). Prof. Jin Wei, as the founder of the National Natural Science Foundation of China, was responsible for study design as well as drafting of the paper.

Authors' contributions

LZZ contributed substantially to the concept and design of the study. ZYP, XCL, and ZHG, WJ, were responsible for study design as well as drafting of the paper. LW and ZYQ contributed to data collection and collation. HQ, GY, CP, ZJH, contributed to the study design. All authors have read and approved the refined version of the manuscript.

Acknowledgments

Not applicable.

Authors' information

[1]Department of Clinical Laboratory, The Second Affiliated Hospital of Xi'an Jiaotong University, 157 Xiwu Road, Xi'an, China; [2]Department of Cardiology, The Second Affiliated Hospital of Xi'an Jiaotong University, 157 Xiwu Road, Xi'an, China; [3]Clinical Research Center for Endemic Disease of Shaanxi province, 5 Jianqiang Road, Xi'an, China; [4]Department of Dermatology, Shaanxi Provincial Institute of Dermatology and Venereology, No. 391, Lianhu Road, Xi'an 710003, China.

Tables

Table 1
Statistics of the three microarray databases derived from the GEO database

Dataset ID	ATC	Normal	Total number
GSE29265	9	20	29
GSE65144	12	13	25
GSE33630	11	45	56
GSE85457	4	3	7

GEO, Gene Expression Omnibus; ATC, anaplastic thyroid cancer

Table 2
GO enrichment analysis and functional annotation of DEGs

Category	Term	Description	Count	*P*-Value
A, upregulated				
BP	GO:0030198	extracellular matrix organization	35	0.0015
BP	GO:0007067	mitotic nuclear division	34	3.07E-12
BP	GO:0030574	collagen catabolic process	32	5.98E-14
BP	GO:0051301	cell division	30	3.78E-19
BP	GO:0007155	cell adhesion	29	2.27E-15
BP	GO:0007062	sister chromatid cohesion	24	1.60E-07
BP	GO:0030199	collagen fibril organization	23	3.12E-04
BP	GO:0007059	chromosome segregation	21	5.21E-06
BP	GO:0006954	inflammatory response	21	3.95E-05
BP	GO:0000082	G1/S transition of mitotic cell cycle	19	8.73E-04
CC	GO:0005581	collagen trimer	107	0.002953
CC	GO:0005615	extracellular space	80	1.21E-04
CC	GO:0005578	proteinaceous extracellular matrix	68	2.96E-17
CC	GO:0000775	chromosome, centromeric region	67	6.06E-04
CC	GO:0030496	Midbody	63	6.34E-11
CC	GO:0005819	Spindle	34	1.44E-18
CC	GO:0005788	endoplasmic reticulum lumen	29	3.87E-15
MF	GO:0005178	integrin binding	185	1.48E-06
MF	GO:0008017	microtubule binding	24	0.003269
MF	GO:0005515	protein binding	22	0.009986
MF	GO:0048407	platelet-derived growth factor binding	17	3.81E-07
B, downregulated				
BP	GO:0030198	extracellular matrix organization	13	0.001586
BP	GO:0007409	Axon genesis	9	0.0016
CC	GO:0016021	integral component of membrane	145	6.56E-04
CC	GO:0005886	plasma membrane	116	0.002979

Category	Term	Description	Count	*P*-Value
CC	GO:0070062	extracellular exosome	106	1.85E-08
CC	GO:0005887	integral component of plasma membrane	51	6.65E-04
CC	GO:0016323	basolateral plasma membrane	19	1.06E-07
MF	GO:0008430	selenium binding	28	0.003789
MF	GO:0004364	glutathione transferase activity	12	0.006795
MF	GO:0008201	heparin binding	5	0.001155

Table 3
KEGG pathway Enrichment Analysis of DEGs

Term	Description	Count	P Value < 0.05	Genes
A, Upregulated				
hsa04512	ECM-receptor interaction	20	5.79E-05	IBSP, TNC, COL3A1, ITGA2, COL5A2, COL5A1, HMMR, LAMB3, COL6A3, COL6A2, COL1A2, COL6A1, COL1A1, THBS1, THBS2, COL11A1, SPP1, FN1
hsa04974	Protein digestion and absorption	18	8.65E-13	COL13A1, COL3A1, MME, COL5A2, COL5A1, COL6A3, COL6A2, COL1A2, COL12A1, CPA3, COL6A1, COL1A1, COL11A1, COL10A1
hsa04510	Focal adhesion	18	7.06E-07	IBSP, TNC, COL3A1, ITGA2, ACTN1, COL5A2, COL5A1, LAMB3, COL6A3, COL6A2, COL1A2, COL6A1, COL1A1, THBS1, THBS2, COL11A1, SPP1, FN1
hsa04110	Cell cycle	14	1.75E-08	CCNE2, CDK1, CDC6, CDKN2A, MAD2L1, CCNB2, CDKN2B, BUB1, TTK, BUB1B, CHEK1, CDC20, CCNA2
hsa04115	p53 signaling pathway	13	6.63E-06	CCNE2, CDK1, CDKN2A, CCNB2, RRM2, SERPINE1, CHEK1, PMAIP1, THBS1, GTSE1
hsa05146	Amoebiasis	12	8.33E-06	LAMB3, COL3A1, COL1A2, TLR2, CXCL8, ACTN1, COL1A1, COL11A1, COL5A2, CD14, COL5A1, FN1
hsa04151	PI3K-Akt signaling pathway	12	2.09E-04	IBSP, TNC, COL3A1, TLR2, ITGA2, COL5A2, COL5A1, CCNE2, LAMB3, EIF4EBP1, COL6A3, COL1A2, COL6A2, COL6A1, COL1A1, THBS1, THBS2, COL11A1, SPP1, FN1
hsa04610	Complement and coagulation cascades	10	6.85E-06	C1QA, C3AR1, C1QB, C5AR1, F13A1, SERPINE1, C1QC, PLAU, PLAUR
hsa04145	Phagosome	10	0.006526	MRC1, MSR1, NCF2, TUBB2A, CD209, TLR2, ITGA2, CLEC7A, THBS1, THBS2, FCGR3B, CD14
hsa05150	Staphylococcus aureus infection	9	6.72E-05	C1QA, C3AR1, C1QB, C5AR1, FPR3, FCGR3B, C1QC
B, downregulated				
hsa04918	Thyroid hormone synthesis	12	4.75E-07	SLC26A4, TG, ATP1B1, PLCB4, DUOXA2, PAX8, GPX3, TPO, GNAS, LRP2, TSHR, IYD

Term	Description	Count	P Value < 0.05	Genes
hsa04960	Aldosterone-regulated sodium reabsorption	8	0.045918	ATP1B1, NR3C2, SCNN1G, SCNN1A, IRS1
hsa04974	Protein digestion and absorption	7	0.015841	COL4A4, COL4A3, ATP1B1, COL14A1, COL9A3, SLC1A1, KCNJ13
hsa04971	Gastric acid secretion	6	0.026392	KCNJ16, KCNJ15, ATP1B1, PLCB4, SLC26A7, GNAS
hsa04514	Cell adhesion molecules (CAMs)	5	0.012053	NCAM1, CLDN8, ITGA9, OCLN, CADM1, CLDN3, CDH1, SDC2
hsa00350	Tyrosine metabolism	4	0.046241	MAOA, TPO, ADH1B, HGD

Table 4
The list of hub DEGs after screening

DEGs	Gene names						
Upregulated	TYMS	KIAA0101	UBE2C	COL1A1	CDH2	CCL13	ADM
	TNC	SPP1	FN1	LGALS1	ITGA2	TIMP1	
Downregulated	COL9A3	COL23A1	SDC2	CCL21	CHRDL1	CCL28	

Figures

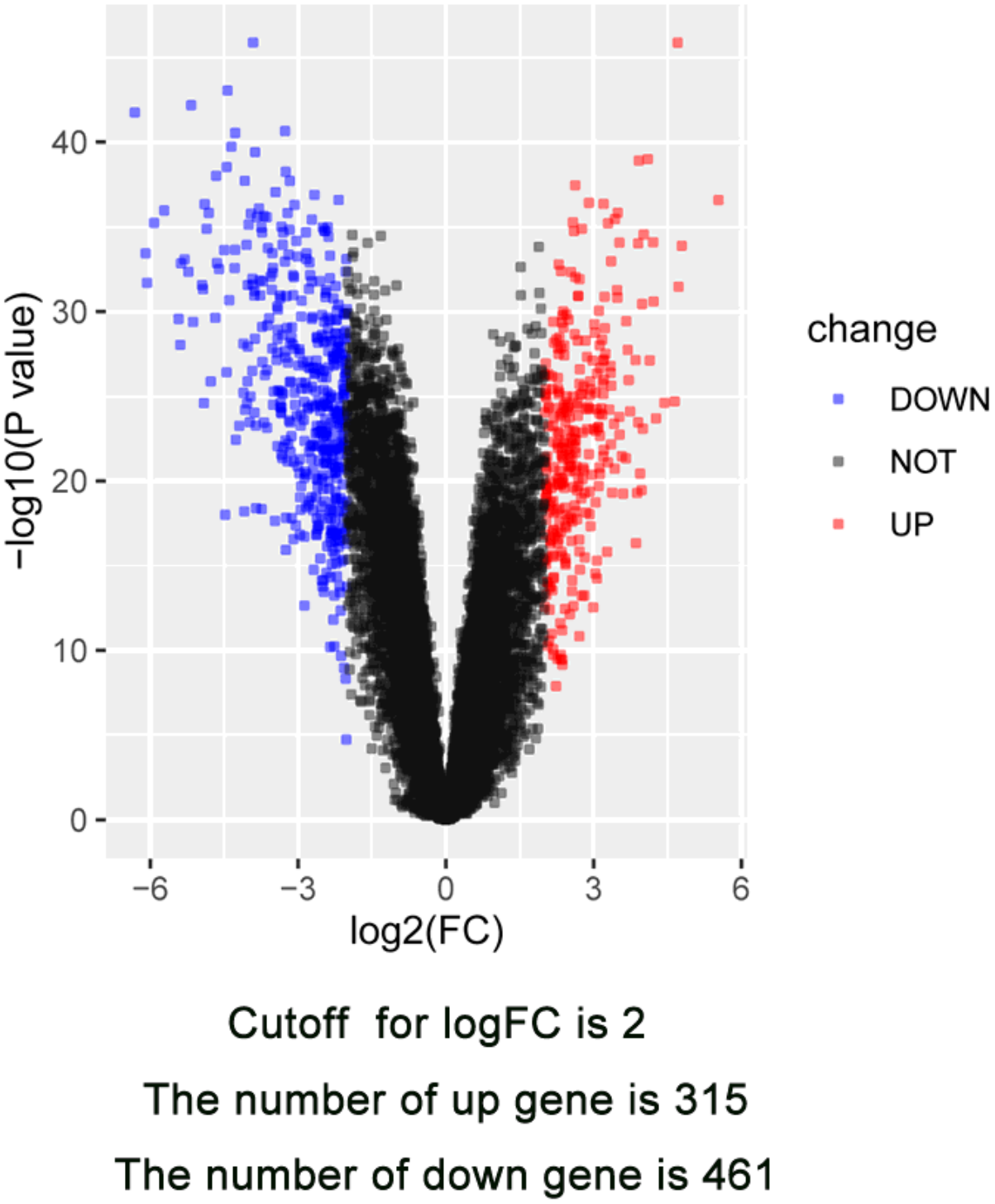

Figure 1

The volcano plot of DEGs in 4 datasets. Red dots represent upregulated genes, gray dots represent undifferentiated genes, and blue dots represent downregulated genes.

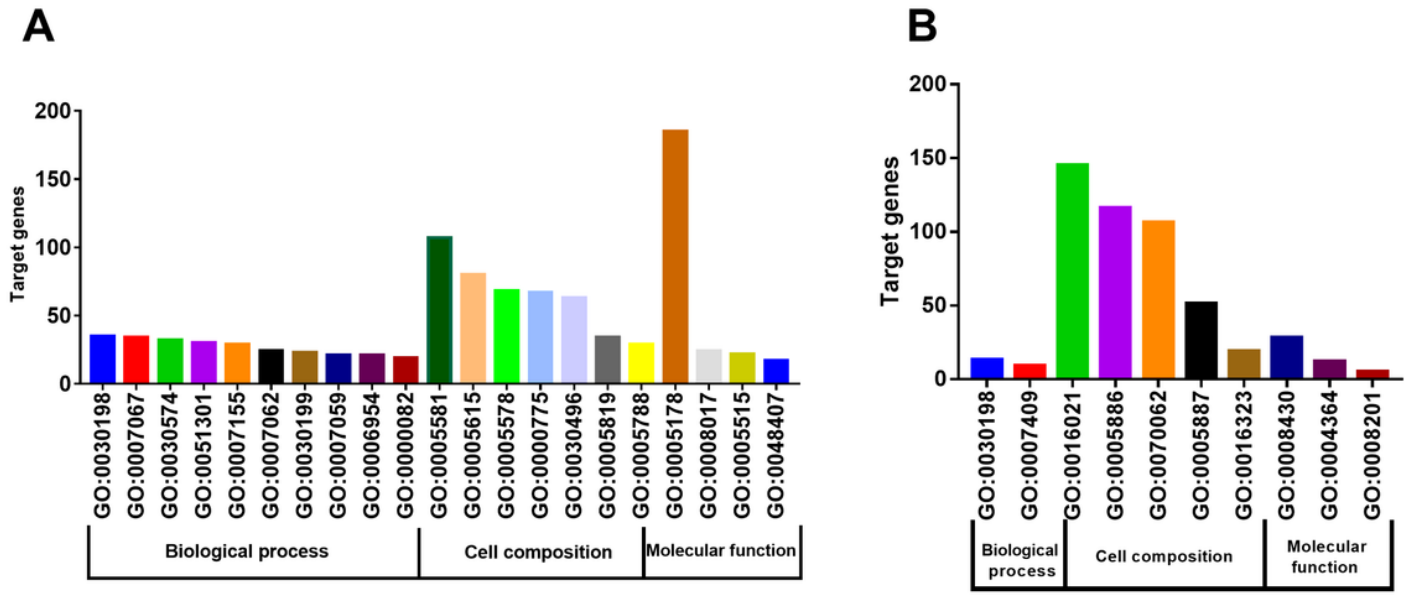

Figure 2

(A) GO functional enrichment analysis of upregulated DEGs based on the three types of sub-ontologies (BP, CC, MF); (B) GO functional enrichment analysis of downregulated DEGs based on the three types of sub-ontologies. GO, gene ontology; BP, Biological Processes; MF, Molecular Function; CC, Cellular Component. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

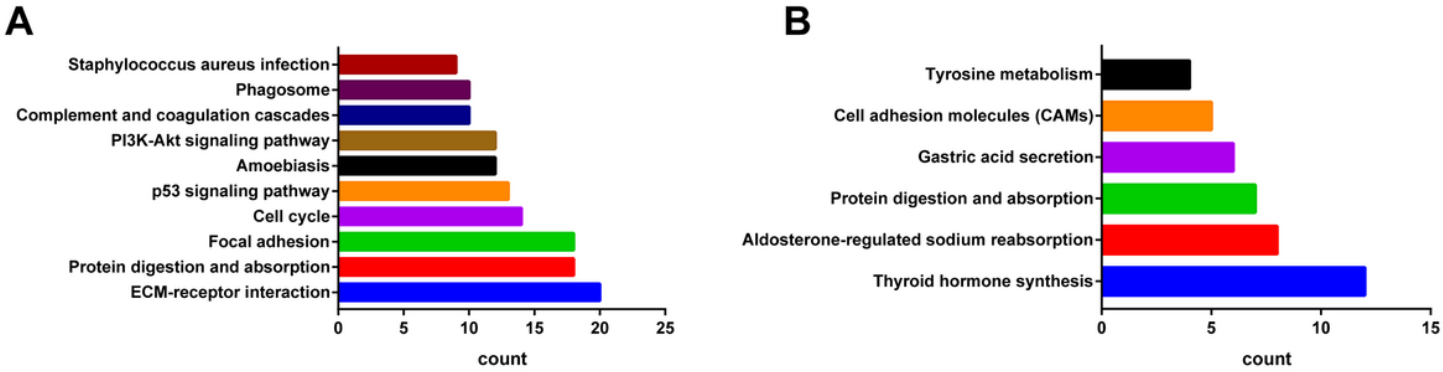

Figure 3

KEGG pathway enrichment analysis of DEGs. KEGG functional enrichment analysis of (A) upregulated and (B) downregulated DEGs were presented, based on P-value.

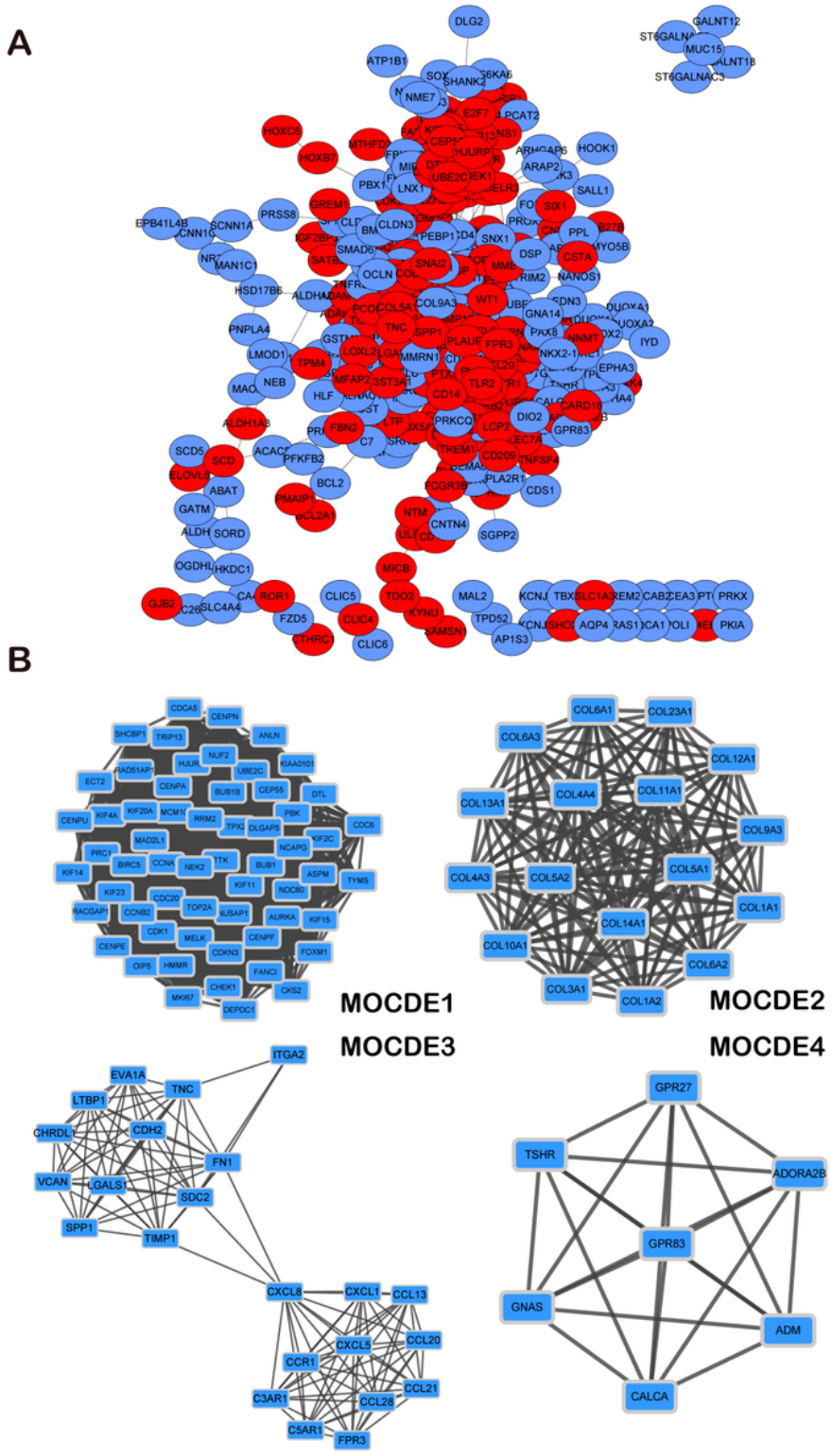

Figure 4

Protein-Protein Interaction network of identified DEGs and the meaningful MOCDEs by the Cytoscape software. (A) Protein-Protein Interaction network of identified DEGs. Blue nodes represent the downregulated genes, while red nodes represent the upregulated genes. (B) The meaningful MOCDEs (MCODE≥7) by the Cytoscape software. DEGs, differentially expressed genes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

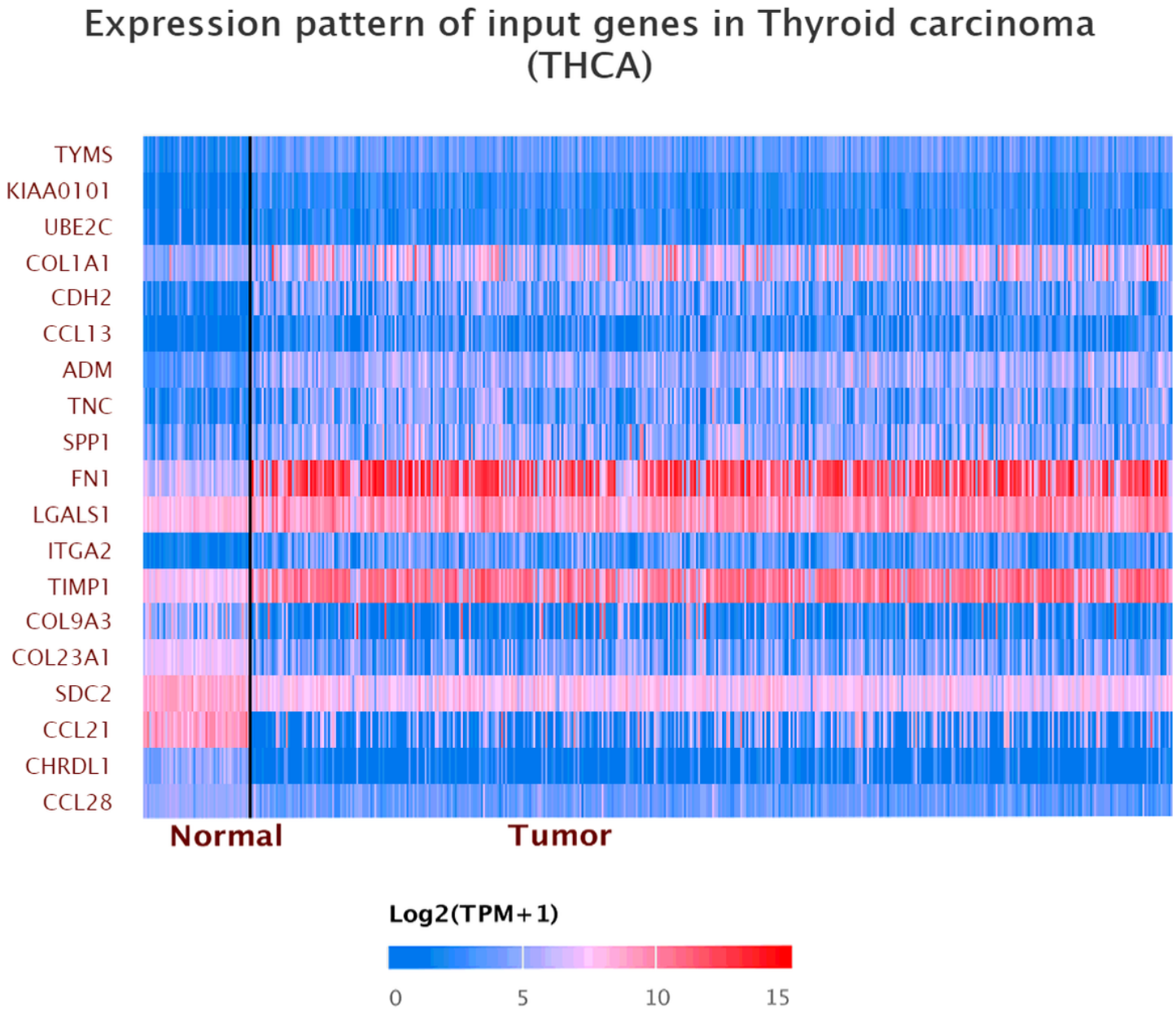

Figure 5

The heatmap of verified Hub DEGs from UALCAN database. The blue color represents lower gene expression, and the red color represents higher gene expression.

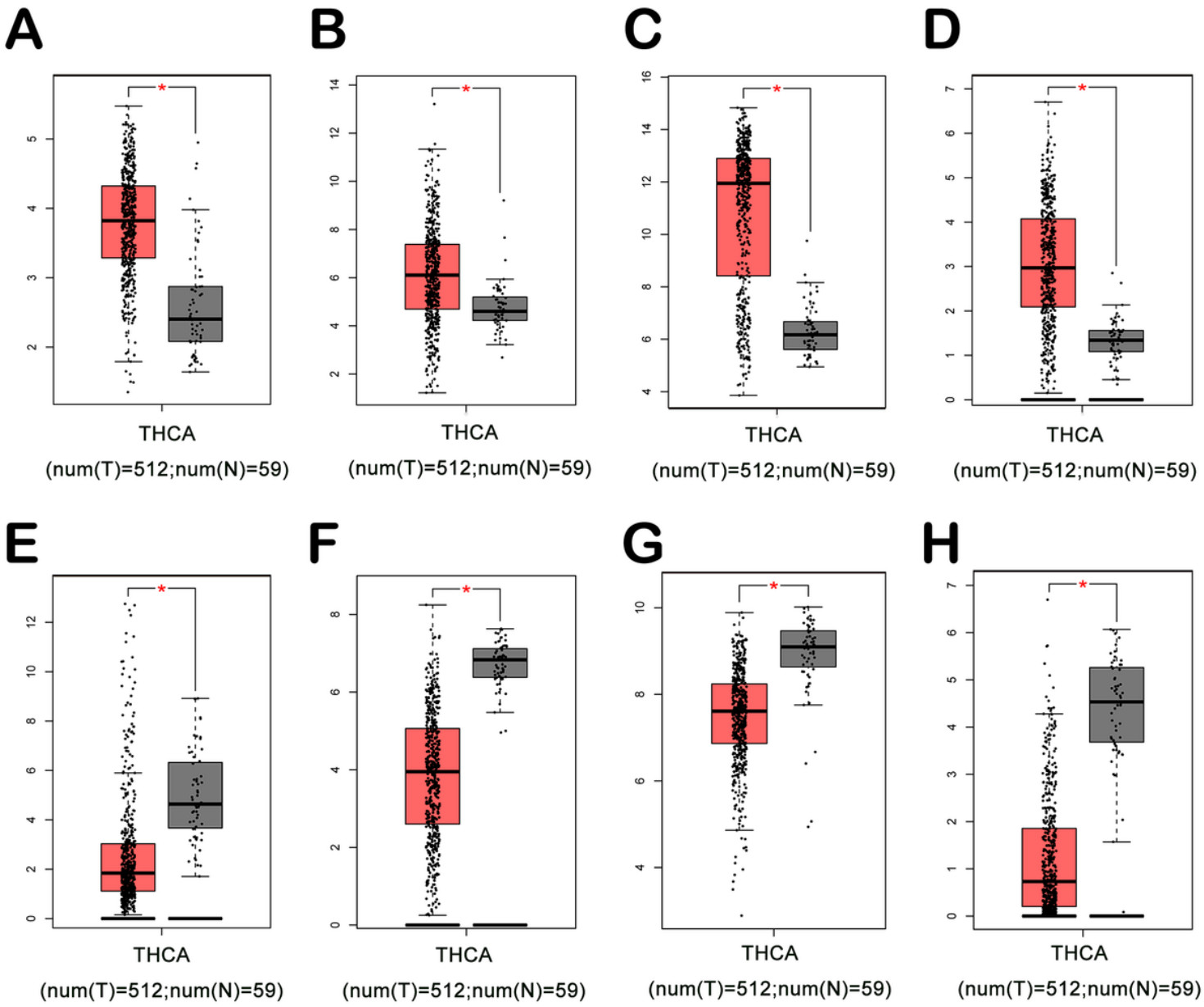

Figure 6

Verification of hub DEGs in TCGA. (A, TYMS; B, COL1A1; C, FN1; D, ITGA2; E, COL9A3; F, COL23A1; G, SDC2; H, CHRDL1. * P< 0.05).

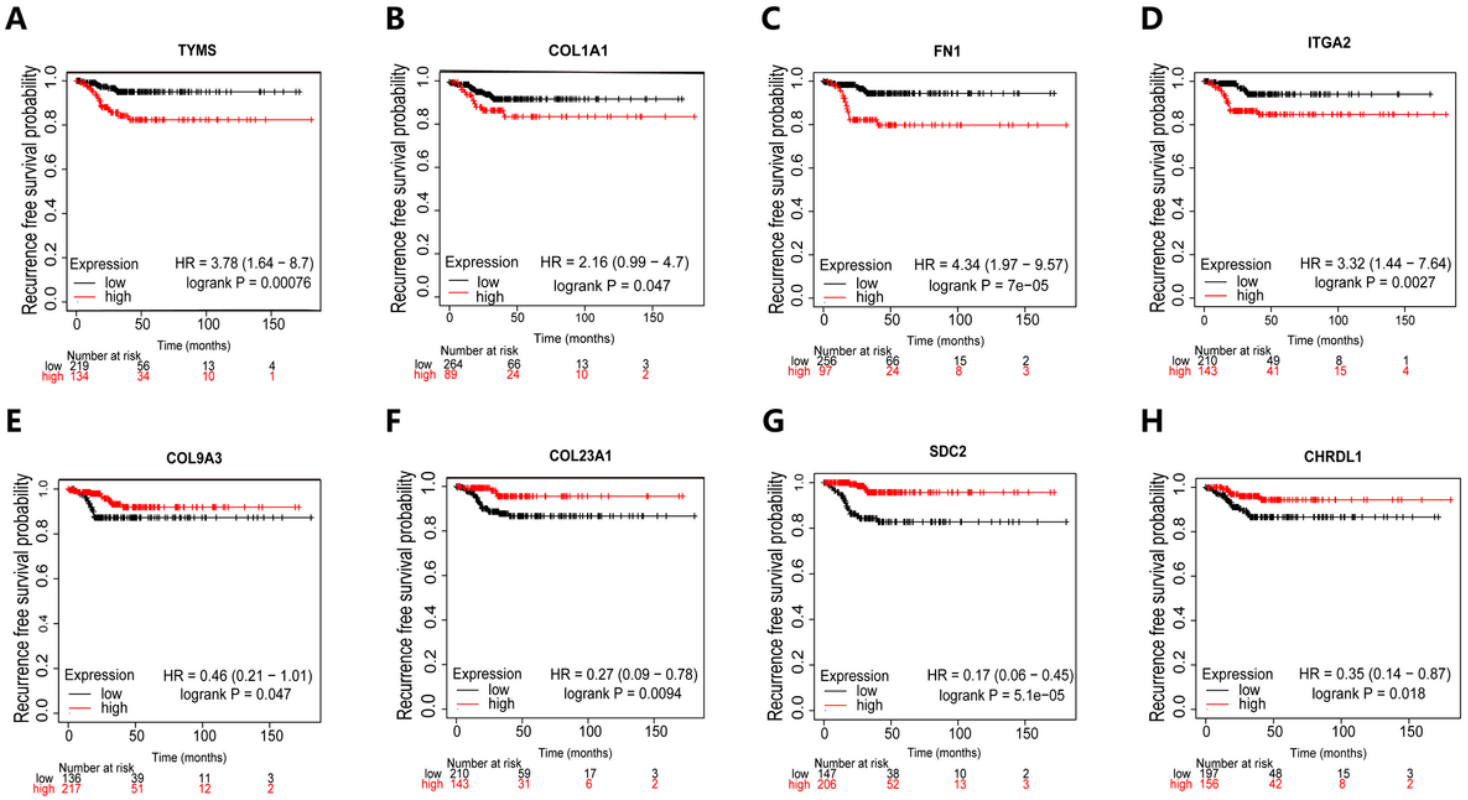

Figure 7

Kaplan–Meier plots of recurrence free survival for hub DEGs in ATC patients.

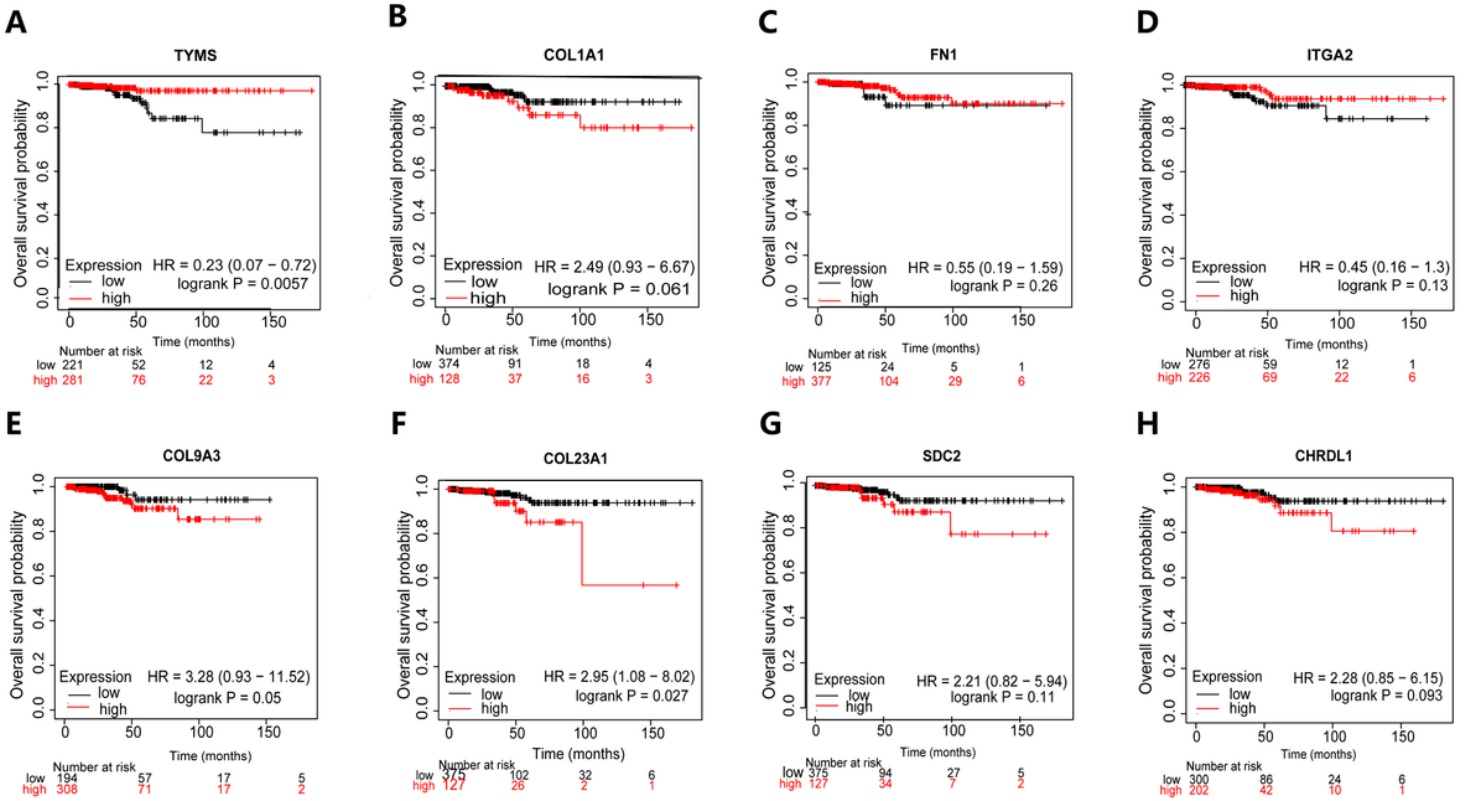

Figure 8

Kaplan–Meier plots of overall survival for the hub DEGs in ATC patients.

Supplementary Files

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

- FigureS1.tif