Original Research

Identification of key biomarkers for thyroid cancer by integrative gene expression profiles

Jinyi Tian, Yizhou Bai, Anyang Liu and Bin Luo

Department of General Surgery, School of Clinical Medicine, Tsinghua University, Beijing Tsinghua Changgung Hospital, Beijing 102218, China

Corresponding author: Bin Luo. Email: lba01097@btch.edu.cn

Impact statement

We identified 594 significantly differentially expressed genes (DEGs) (277 up-regulated genes together with 317 down-regulated genes) from the tumor and normal tissue samples. A diagnostic seven-gene signature was established using a logistic regression model with the area under the receiver operating characteristic curve (AUC) of 0.967. Seven robust candidate biomarkers predictive of thyroid cancer were identified, and the obtained seven-gene signature may serve as a useful marker for thyroid cancer diagnosis and prognosis.

Abstract

Thyroid cancer is a frequently diagnosed malignancy and the incidence has been increased rapidly in recent years. Despite the favorable prognosis of most thyroid cancer patients, advanced patients with metastasis and recurrence still have poor prognosis. Therefore, the molecular mechanisms of progression and targeted biomarkers were investigated for developing effective targets for treating thyroid cancer. Eight chip datasets from the gene expression omnibus database were selected and the inSilicoDb and inSilicoMerging R/Bioconductor packages were used to integrate and normalize them across platforms. After merging the eight gene expression omnibus datasets, we obtained one dataset that contained the expression profiles of 319 samples (188 tumor samples plus 131 normal thyroid tissue samples). After screening, we identified 594 significantly differentially expressed genes (277 up-regulated genes plus 317 down-regulated genes) between the tumor and normal tissue samples. The differentially expressed genes exhibited enrichment in multiple signaling pathways, such as p53 signaling. By building a protein–protein interaction network and module analysis, we confirmed seven hub genes, and they were all differentially expressed at all the clinical stages of thyroid cancer. A diagnostic seven-gene signature was established using a logistic regression model with the area under the receiver operating characteristic curve (AUC) of 0.967. Seven robust candidate biomarkers predictive of thyroid cancer were identified, and the obtained seven-gene signature may serve as a useful marker for thyroid cancer diagnosis and prognosis.

Keywords: Thyroid cancer, gene expression omnibus, multi-platform, biomarkers, bioinformatics

Experimental Biology and Medicine **2021; 246: 1617–1625. DOI: 10.1177/15353702211008809**

Introduction

Thyroid cancer is an endocrine malignancy and the incidence keeps increasing around the world.[1,2] There were approximately 64,300 new cases of thyroid cancer diagnosed in 2016, resulting in nearly 1980 deaths.[3] Thyroid cancer is sub-classified into papillary, follicular, medullary, poorly differentiated, and anaplastic thyroid cancers according to specific histopathological characteristics. The prognosis of the cancer largely depends on the cancer type and stage, so it is particularly necessary to understand the molecular mechanisms that underlie cancer development and progression. The key biomarkers as well as targets for thyroid cancer treatment shall be identified crucially.

High-throughput technologies have allowed the discovery and integration of many gene expression data, which provide a basis for more deeply understanding the molecular mechanisms regarding thyroid cancer. Multiple datasets have been integrated and screened to detect genetic markers for thyroid cancer. Li *et al.* identified 423 DEGs related to papillary thyroid cancer (PTC) using three sets of gene expression omnibus (GEO) expression profile datasets, and mined and screened 21 core PTC-related genes in a protein interaction network.[4] Zhao *et al.* [5] identified *FN1* and *TRAF6* as two key genes associated with thyroid cancer

by analyzing five GEO microarray datasets. Qu *et al.* [6] used two GEO microarray datasets to screen a large number of DEGs and identified potential regulatory pathways of *FN1* and *SERPINA1* in PTC. Together, these findings helped us to more deeply understand the regulatory mechanisms underlying thyroid cancer development as well as opened new directions for discovering PTC biomarkers and therapeutic targets.

Integrating and analyzing multiple datasets that were obtained using different study cohorts, different sequencing platforms, and different sample processing methods can be hindered by noisy data and batch effects.[7] Various algorithms have been designed to solve these problems. Cross-platform data integration is effective in screening disease markers and can greatly improve the reproducibility and robustness of the results. Multi-dataset integration can greatly enlarge the sample size and enhance the reliability of the results [8] by reducing the effects of individual datasets.[9]

In this study, eight GEO datasets were integrated and 594 DEGs were identified, which were enriched mainly in signaling pathways closely related to p53, a tumor suppressor protein, and other tumors. We constructed a protein–protein interaction (PPI) network, meanwhile identifying seven hub genes (*NMU*, *COL1A2*, *LPAR5*, *CXCL8*, *COL5A2*, *COL11A1*, *COL5A1*). A seven-gene signature with diagnostic value was established using a logistic regression model with AUC of 0.967. These seven genes can be considered robust candidate biomarkers predictive of thyroid cancer, and the seven-gene signature is a candidate marker for diagnosis and prognosis prediction of thyroid cancer.

Materials and methods

GEO microarray dataset selection and processing for thyroid cancer

Eight gene expression profiling datasets were screened from the GEO database (https://www.ncbi.nlm.nih.gov/geo/): GSE3467 [10], GSE3678 (no citation), GSE6004 [11], GSE29265 (no citation), GSE33630 [12], GSE53157 [13], GSE60542 [14], and GSE65144 [15]. There were three inclusion criteria:

1. The dataset contained data for both normal and thyroid tumor samples.
2. The detection platform was an Affymetrix gene chip
3. Sample sizes were >10.

We downloaded the raw data files about the eight GEO datasets from the GEO website. A standard robust multi-array average (RMA) method served for data preprocessing.[16] The processed data were converted to IDs and transformed to gene symbols. In the case of multiple probes corresponding to one gene, we used the median gene expression value. Probes that corresponded to multiple genes were removed. Finally, each dataset was quantile normalized by virtue of the affy package in R (Supplementary Figure 1).

Merging the eight microarray datasets

The inSilicoMerging served for dataset normalization.[7] The batch effects across platforms were removed by inSilicoDb[17] and adjusted using ComBat.[18] ComBat is an empirical Bayesian method that estimates the parameters representing the batch effect by summarizing the information among genes in each batch for reducing the batch effect parameters to the overall estimated average.[19] After merging the eight GEO datasets, one dataset that incorporated a total of 319 samples (188 tumor and 131 normal thyroid tissues) was obtained.

Screening of DEGs

We used the limma package in R [20] to identify DEGs related to thyroid cancer in the merged dataset. $|\log2FC| > 1$ and corrected *P*-values <0.05 were considered as the threshold.

Functional enrichment analyses

We conducted gene ontology (GO) enrichment analysis using the DAVID online analytical tool (https://david.ncifcrf.gov/), for figuring out the functional role played by DEGs.[21] Enriched GO terms under the three main categories, namely molecular function (MF), biological process (BP), and cellular component (CC), and false discovery rate (FDR) <0.05 were considered significant. A Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was carried out using the KOBAS server;[22] FDRs <0.05 exhibited significance.

PPI network construction and module mining

STRING (version 11.2) (https://string-db.org/) assisted in building a network between known and predicted proteins.[23] An interaction score of 0.9 was considered as the threshold for obtaining reliable PPI data, and local Cytoscape (version 3.6.2) software (https://www.cytoscape.org/) served for visualizing the network. The topological properties of the network were analyzed, and module mining was carried out using the MCODE plug-in.

Identification of hub genes in the PPI network

The Cytoscape software plug-in cytoHubba [24] with maximum neighborhood component (MNC) and maximal clique centrality (MCC) assisted in identifying significant hub genes. The top 10 hub genes were screened and a Venn diagram was created to select mutual hub genes.

Validation of key gene using an external dataset

We downloaded the FPKM data of gene expression profiles derived from 502 thyroid cancer and 58 paracancer samples from The Cancer Genome Atlas (TCGA) database [25]using gdc-api and used the data for verifying the expressions of key genes. The GEPIA2 database (https://gepia2.cancer-pku.cn/) [26] was used to verify the prognostic differences among the genes. A ROC was plotted and a logistic regression model was adopted for evaluating the diagnostic performance owned by the seven-gene signature for thyroid cancer.

Results

Identification of DEGs related to thyroid cancer

An integrative multi-platform analysis was conducted against the gene expression profiles in the merged GEO dataset to identify key biomarkers of thyroid cancer (Figure 1). Five hundred ninety-four DEGs were identified, 277 up-regulated, and 317 down-regulated, with corrected *P*-values <0.05 and $|\log2FC| > 1$. Figure 2 displays Volcano plots and a heatmap of the DEGs. Table 1 lists the top five DEGs.

GO term and KEGG pathway function enrichment analysis of the DEGs

DAVID helped to identify 12 enriched GO terms in total (20). Under molecular function, the up-regulated genes presented significant enrichment in serine-type endopeptidase activity, extracellular matrix structural constituent, and protease binding (Figure 3(a)). Under biological process, the up-regulated genes presented enrichment in the collagen catabolic process, extracellular matrix organization, and cell adhesion. Under cellular component, the up-regulated genes presented enrichment mainly in extracellular space, proteinaceous extracellular matrix, and collagen trimer. Under biological process, the down-regulated genes exhibited enrichment in the BMP signaling pathway and thyroid hormone generation. Under cellular component, the down-regulated genes exhibited enrichment mainly in extracellular space, proteinaceous extracellular matrix, and extracellular exosome (Figure 3(b)).

KOBAS helped to identify 14 significantly enriched KEGG pathways in total.[21] The up-regulated genes played a role in protein digestion and absorption, p53 signaling, complement and coagulation cascades, and pathways related to cancer (Figure 3(c)). The down-regulated genes showed significant enrichment in thyroid hormone synthesis, TGF-beta signaling, as well as tyrosine metabolism (Figure 3(d)). These results indicate the close involvement of DEGs in biological functions and pathways associated with cancer.

PPI network construction and module mining

STRING (22) was employed for constructing a PPI network that had a total of 203 nodes comprising 117 up-regulated genes plus 86 down-regulated genes (Figure 4(a)). Four network modules were identified using the MCODE plug-in (Figure 4(b) to (e)). We explored the significantly enriched pathways in each module using KOBAS.[21] In module 1, the DEGs showed significant enrichment in chemokine signaling, cAMP signaling pathway, and cytokine-cytokine receptor interaction. In module 2, the DEGs showed enrichment in the protein digestion and absorption, ECM-receptor interaction, and focal adhesion. In module 3, the DEGs showed enrichment in *Staphylococcus aureus* infection, complement and coagulation cascades, as well as bacterial invasion regarding epithelial cells. In module 4, the DEGs showed enrichment mainly in neuroactive ligand-receptor interaction, calcium signaling pathway, and renin-angiotensin system. These results confirm the involvement of the DEGs in the four network modules in different biological processes in the development of thyroid cancer.

Identification of network hub genes

The Cytoscape plug-in cytoHubba [24] helped to detect key hub genes. A Venn diagram about the top ten hub genes was plotted to select shared hub genes. Seven genes—*NMU*, *COL1A2*, *LPAR5*, *CXCL8*, *COL5A2*, *COL11A1*, and *COL5A1*—were identified (Figure 5(a)). These genes may

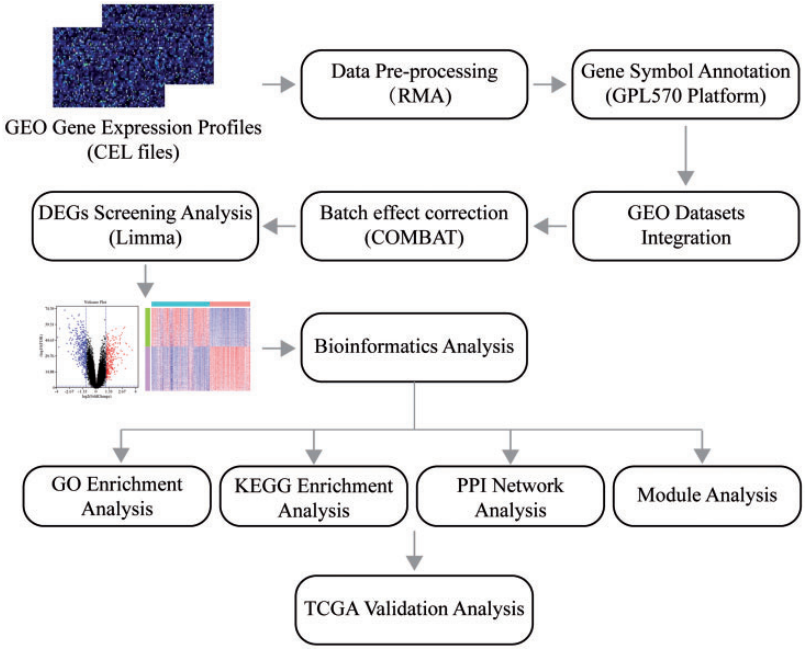

Figure 1. Flow chart of the integrative multi-platform analysis conducted in this study (A color version of this figure is available in the online journal.)

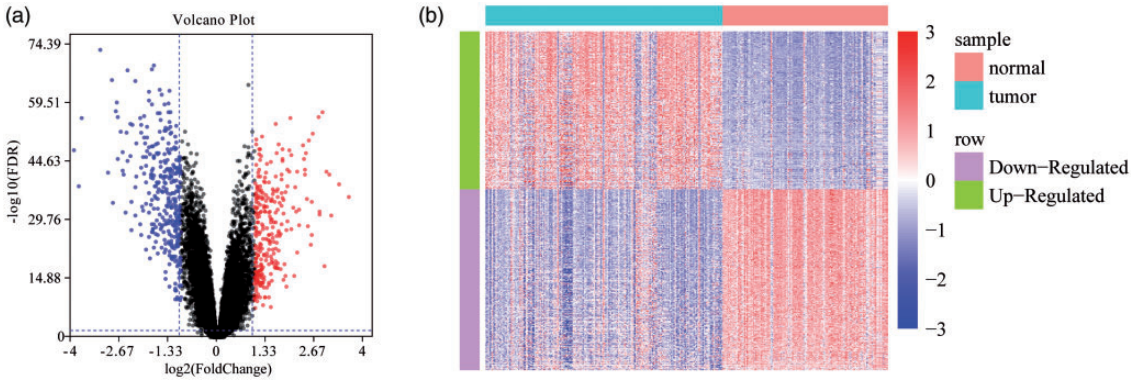

Figure 2. DEGs related to thyroid cancer identified in the study. (a) Volcano graph of the 594 identified DEGs. The horizontal axis indicates gene expression multiples; the vertical axis indicates significant *P*-values. (b) Heatmap of the 594 identified DEGs. (A color version of this figure is available in the online journal.)

Table 1. Top 10 DEGs of thyroid cancer.

Gene symbol	Gene name	Log2(Fold-Change)	Corrected *P*-value	Regulation
ZCCHC12	Zinc finger CCHC-type containing 12	3.598	1.45746539074895e-36	Up
PRR15	Proline-rich 15	3.316	6.6507222979452e-40	Up
CHI3L1	Chitinase 3-like 1	3.094	8.35633170653537e-32	Up
LRP4	LDL receptor-related protein 4	3.061	3.40155936718936e-42	Up
GABRB2	Gamma-aminobutyric acid type A receptor beta2 subunit	2.976	5.29826988266041e-43	Up
TFF3	Trefoil factor 3	−4.117	2.56113340224372e-70	Down
TPO	Thyroid peroxidase	−3.943	1.47946458958149e-48	Down
DIO1	Iodothyronine deiodinase 1	−3.804	2.83170255583346e-39	Down
PKHD1L1	PKHD1 like 1	−3.718	1.22427913897898e-56	Down
IPCEF1	Interaction protein for cytohesin exchange factors 1	−3.217	5.52506500130381e-74	Down

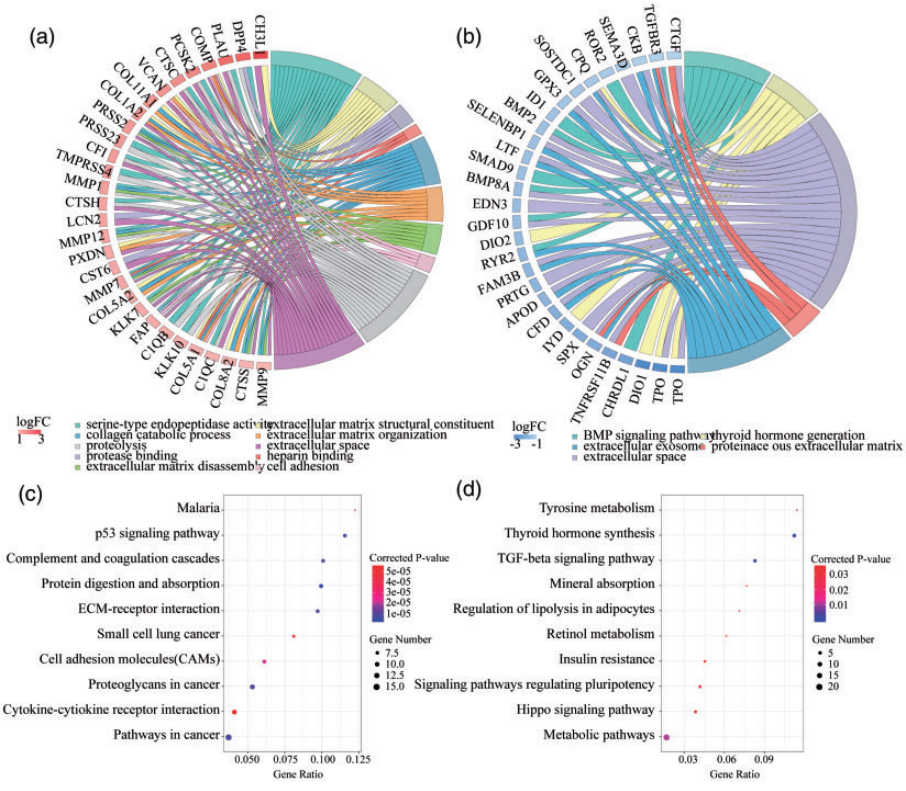

Figure 3. Function enrichment analysis on DEGs related to thyroid cancer. (a, b) String diagrams of the enriched gene ontology (GO) terms for (a) the up-regulated genes and (b) the down-regulated genes. Genes are on the left; GO terms are on the right; colors indicate the difference in multiples. (c, d) Bubble charts of the enriched KEGG pathways specific to (c) the up-regulated genes and (d) the down-regulated genes. Colors indicate the significance; dot size indicates the number of DEGs. (A color version of this figure is available in the online journal.)

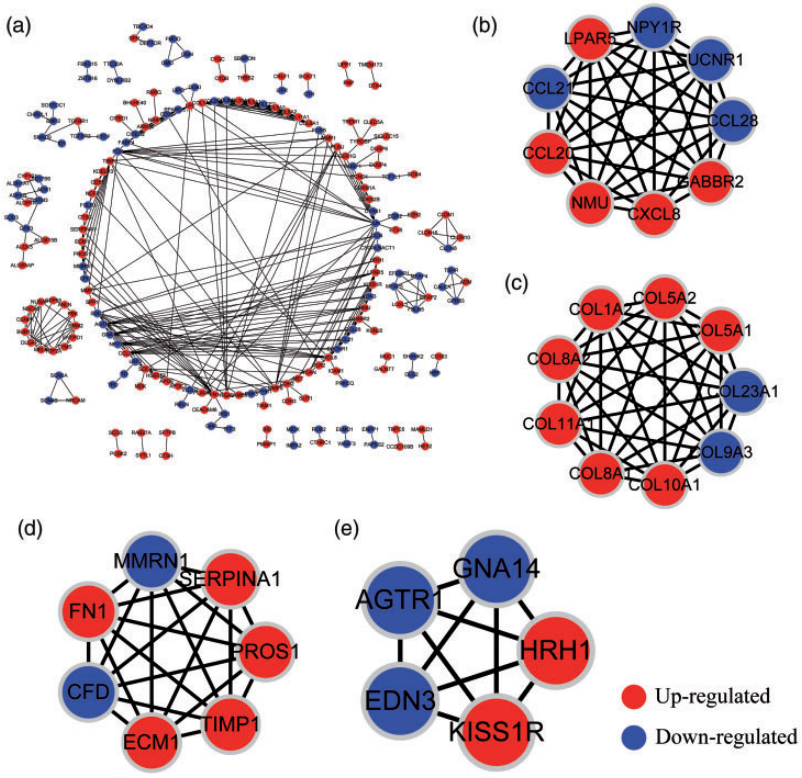

Figure 4. PPI network and module analysis. (a) Proposed PPI network of the proteins encoded by the differentially expressed genes. (b–e) Network modules identified with the MCODE plug-in: (b) module 1 (c) module 2, (d) module 3, and (e) module 4. (A color version of this figure is available in the online journal.)

potentially serve for thyroid cancer treatment as novel biomarkers.

Robustness of gene signatures

For validating our results, we extracted the expression data about the seven hub genes from the TCGA gene expression dataset of thyroid cancer. We analyzed the differences in their expression profiles between the carcinoma and paracancer samples (Figure 5(a)), and found that the significant difference was in line with our results (Figure 5(c)). How these seven genes affect the prognosis of thyroid cancer was analyzed using GEPIA2 (24) and found that only *LPAR5* showed significant differences (log-rank P-value = 0.0049, Figure 5(d)). Additionally, all seven hub genes showed differential expression throughout the different stages of thyroid cancer (Supplementary Figure 2(a) to (g)). On the basis of these results, we established a seven-gene signature and evaluated its diagnostic performance. A combined ROC was drawn using a logistic regression model, and the combined AUC was 0.967 (Figure 5(e)). In order to further verify the seven gene diagnosis performance, we download a set of thyroid cancer data from GEO database validation set GSE35570,[2] the seven gene expression profiles were extracted, genes expression differences were detected in the cancer samples and paracancerous samples. As expected, they are significantly highly expressed in tumor samples (Figure 6(a)). Furthermore, we obtained the expression profile and prognosis information of seven genes from TCGA, and observed the prognostic significance of these seven genes from univariate and multivariate aspects, respectively. Univariate analysis showed that LPAR5, COL5A2, and COL5A1 had significant prognostic differences. Multivariate analysis of significant LPAR5 had a significant prognostic difference (Figure 6(b)), which followed the results from the GEPIA database.

Discussion

An integrative multi-platform analysis was conducted over the gene expression profiles to identify key biomarkers for thyroid cancer. The gene expression profiles of eight GEO datasets were integrated after removing batch effects among the datasets from different sources. The integrated dataset contained a total of 319 samples; 188 tumor and 131 normal thyroid tissue samples. GO terms and KEGG pathways were used to annotate significant DEGs between tumor samples and normal samples, a PPI network was constructed, and the network modules were analyzed. The key DEGs were validated using a TCGA gene expression dataset. Seven hub genes—*NMU*, *COL1A2*, *LPAR5*, *CXCL8*, *COL5A2*, *COL11A1*, and *COL5A1*—were identified as potential and novel biomarkers for thyroid cancer.

The rapid increases in thyroid cancer cases worldwide and the availability of high-throughput gene expression data encouraged us to study human malignancies, including thyroid cancer. Pan *et al.* [27] identified four down-regulated microRNAs (miRNAs) associated with thyroid cancer using the GEO and ArrayExpress databases, and suggested these miRNAs may influence tumorigenesis by regulating critical pathways. Furthermore, by analyzing TCGA, ArrayExpress, and GEO databases, it was

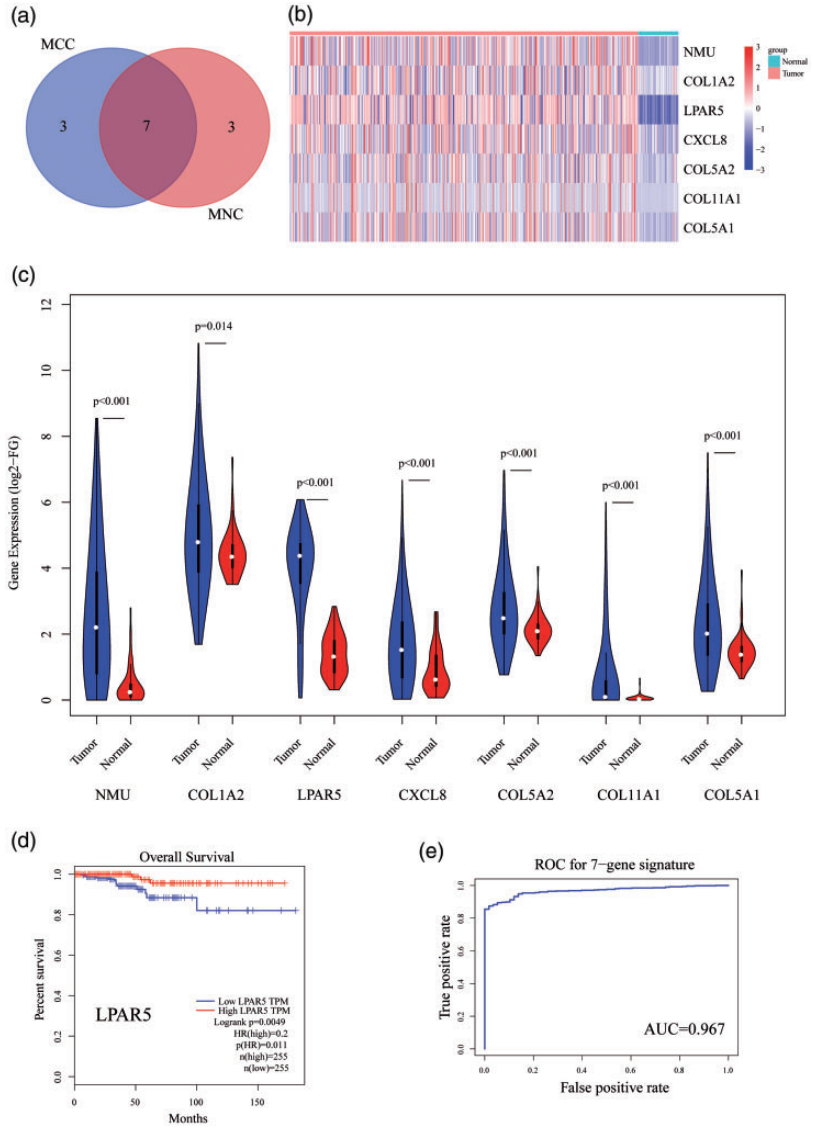

Figure 5. Verification of key genes using the TCGA gene expression dataset of thyroid cancer. (a) Venn diagram of hub genes identified with MCC and MNC algorithms. (b) Heatmap of hub genes in TCGA database. (c) Expression differences of the seven hub genes between the carcinoma and paracancer samples in TCGA. (d) Kaplan–Meier curves of prognosis differences in samples with highly or lowly expressed *LPAR5*. (e) ROC of the seven-gene signature for diagnosis of thyroid cancer. (A color version of this figure is available in the online journal.)

discovered that miR-486-5p often suffered down-regulation in PTC compared with its expression in normal tissues, and that miR-486-5p expression was associated with cancer stage, pathological lymph node status, metastasis, tumor, as well as recurrence.[28] Luo *et al.* [29] analyzed the expression profiles regarding long non-coding RNAs (lncRNAs) in TCGA RNA-sequencing datasets, and established a three-lncRNA (AC079630.2, CRNDE, CTD-2171N6.1) survival signature that was confirmed to be able to independently predict PTC patients' survival. Other studies have also provided novel and promising diagnostic, prognostic, and therapeutic markers for thyroid cancer.[6,30,31] Liu *et al.* identified five hub genes (*LPAR5*, *NMU*, *FN1*, *NPY1R*, *CXCL12*) by constructing a PPI network.[30] Griffith *et al.* conducted a comprehensive meta-analysis on the thyroid cancer biomarkers from 21 studies and proposed a method for identifying biomarkers based on gene ranking.[31] Although various genetic markers for thyroid cancer have been proposed, so far no clinical genetic markers are available; therefore, more studies are needed to identify genetic markers that can be used by clinicians and experimentalists.

Limitations, such as small sample size and the combination of different platforms or groups, often result in inaccurate data. The potential diagnostic and prognostic applications of microarray data analysis have contributed to developing useful methods for data integration from various microarray platforms.[8] Cross-platform data integration can improve the reproducibility and robustness of genetic biomarkers by increasing sample size and statistical power to draw more general and reliable conclusions.[9] We used ComBat to remove batch effects because it reduces computational costs and is independent of sample size.[32] This helped to avoid the influence of different batch

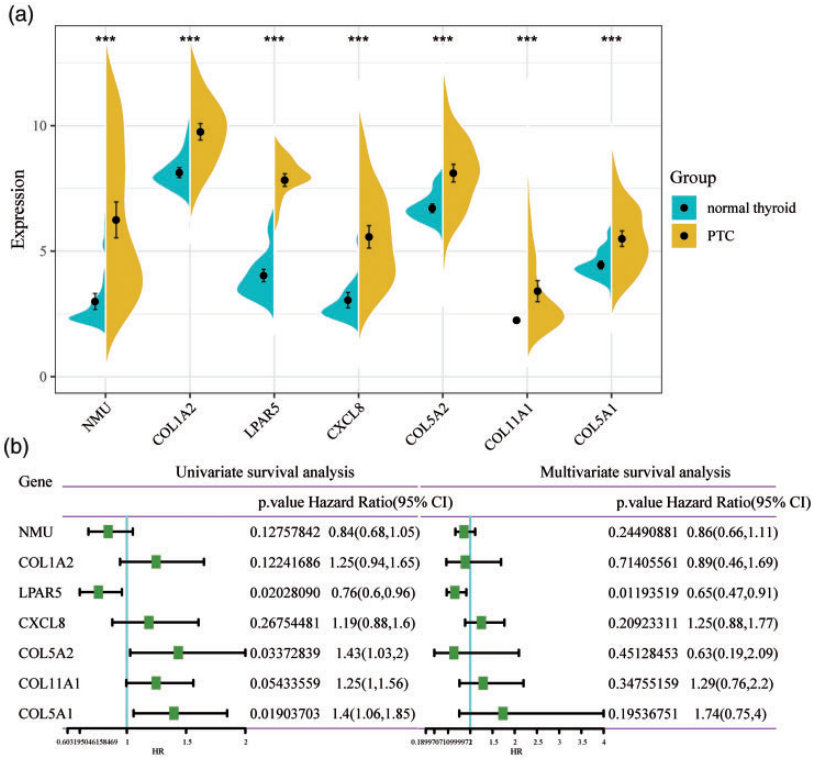

Figure 6. Validation of external data sets. (a) Expression difference of seven genes in GSE35570 data set. (b) Prognostic difference of seven genes between univariate and multivariate in the TCGA data set. (A color version of this figure is available in the online journal.)

processing parameters across platforms. We also compared our model with previously published models,[33–35] and found that our model had a higher AUC (Supplementary Figure 3). In this study, we used the inSilicoMerging and inSilicoDb packages in R to merge the eight selected GEO datasets into one integrated expression dataset. The merged dataset contained 319 samples (188 tumor samples plus 131 normal thyroid tissue samples) that were further examined.

According to the KEGG pathway enrichment analysis, the up-regulated genes showed significant enrichment in several cancer-related pathways, namely p53 signaling, pathways in cancer, ECM-receptor interaction, proteoglycans in cancer, and PI3K-Akt signaling, whereas the down-regulated genes experienced enrichment mainly in thyroid hormone synthesis and TGF-beta signaling. TGF-β1 was shown to be overexpressed in anaplastic thyroid cancer, and therapies targeting the TGF-β1 pathway were found to be effective in preventing primary tumor formation.[36]

The PPI network together with the module analysis assisted in identifying seven hub genes (*NMU*, *COL1A2*, *LPAR5*, *CXCL8*, *COL5A2*, *COL11A1*, *COL5A1*) among which *NMU* and *LPAR5* had been identified previously in a PPI network reported by Tang *et al.*.[37] A bioinformatics analysis identified *NMU* as a potential gene triggering alectinib resistance in non-small cell lung cancer (NSCLC),[38] and down-regulation of *LPAR5* was found to contribute to aberrant lysophosphatidic acid (LPA) signaling in nasopharyngeal carcinoma associated with Epstein–Barr virus (EBV).[39] An integrated analysis of microarray datasets revealed *COL1A2* as a candidate biomarker for cholangiocarcinoma diagnosis and prognosis;[40] however, there are no reports of *COL1A2* in thyroid cancer. As found, the aberrant expression of *CXCL8* affected the thyroid cancer pathogenesis.[41–43] Up-regulation of *COL11A1* is a candidate marker for cancer, including PTC; particularly, the T alleles of rs1763347 and rs2229783 reduced the risk of PTC in a Korean population.[44] Qiu *et al.* used five paired PTC tissues and RNA-sequencing and found a set of collagen-encoding genes with *COL5A1* as a hub node in a PPI network for thyroid cancer.[45] However, no studies have been conducted on *COL5A2* in thyroid cancer.

Conclusions

In conclusion, the identified seven hub genes can be considered as novel biomarkers, reliably used for the thyroid cancer diagnosis, prognosis prediction, as well as targeted therapy design.

AUTHORS' CONTRIBUTIONS

All authors participated in the design, interpretation of the studies and analysis of the data and review of the manuscript. TJY designed the study. TJY, BYZ, and LAY performed the study. TJY wrote the initial draft of the manuscript. LB reviewed and edited the manuscript. All authors read and approved the manuscript.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was conducted with the support of the Beijing Tsinghua Changgung Hospital Fund (No. 12017C1024).

ORCID iD

Bin Luo https://orcid.org/0000-0002-8721-4460

REFERENCES

1. Girardi FM, Barra MB, Zettler CG. Analysis of pattern of occurrence of thyroid carcinoma between 2001 and 2010. *Braz J Otorhinolaryngol* 2015;**81**:541–8
2. Carling T, Udelsman R. Thyroid cancer. *Annu Rev Med* 2014;**65**:125–37
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin* 2016;**66**:7–30
4. Li S, Yin Y, Yu H. Genetic expression profile-based screening of genes and pathways associated with papillary thyroid carcinoma. *Oncol Lett* 2018;**16**:5723–32
5. Zhao H, Li H. Network-based meta-analysis in the identification of biomarkers for papillary thyroid cancer. *Gene* 2018;**661**:160–8
6. Qu T, Li YP, Li XH, Chen Y. Identification of potential biomarkers and drugs for papillary thyroid cancer based on gene expression profile analysis. *Mol Med Rep* 2016;**14**:5041–8
7. Taminau J, Meganck S, Lazar C, Steenhoff D, Coletta A, Molter C, Duque R, de Schaetzen V, Weiss Solis DY, Bersini H, Nowe A. Unlocking the potential of publicly available microarray data using inSilicoDb and inSilicoMerging R/bioconductor packages. *BMC Bioinformatics* 2012;**13**:335
8. Walsh CJ, Hu P, Batt J, Santos CC. Microarray meta-analysis and cross-platform normalization: integrative genomics for robust biomarker discovery. *Microarrays* 2015;**4**:389–406
9. Taminau J, Lazar C, Meganck S, Nowe A. Comparison of merging and meta-analysis as alternative approaches for integrative gene expression analysis. *ISRN Bioinform* 2014;**2014**:345106
10. He H, Jazdzewski K, Li W, Liyanarachchi S, Nagy R, Volinia S, Calin GA, Liu CG, Franssila K, Suster S, Kloos RT, Croce CM, de la Chapelle A. The role of microRNA genes in papillary thyroid carcinoma. *Proc Natl Acad Sci U S A* 2005;**102**:19075–80
11. Vasko V, Espinosa AV, Scouten W, He H, Auer H, Liyanarachchi S, Larin A, Savchenko V, Francis GL, de la Chapelle A, Saji M, Ringel MD. Gene expression and functional evidence of epithelial-to-mesenchymal transition in papillary thyroid carcinoma invasion. *Proc Natl Acad Sci U S A* 2007;**104**:2803–8
12. Tomas G, Tarabichi M, Gacquer D, Hebrant A, Dom G, Dumont JE, Keutgen X, Fahey TJ 3rd, Maenhaut C, Detours V. A general method to derive robust organ-specific gene expression-based differentiation indices: application to thyroid cancer diagnostic. *Oncogene* 2012;**31**:4490–8
13. Pita JM, Banito A, Cavaco BM, Leite V. Gene expression profiling associated with the progression to poorly differentiated thyroid carcinomas. *Br J Cancer* 2009;**101**:1782–91
14. Tarabichi M, Saiselet M, Tresallet C, Hoang C, Larsimont D, Andry G, Maenhaut C, Detours V. Revisiting the transcriptional analysis of primary tumours and associated nodal metastases with enhanced biological and statistical controls: application to thyroid cancer. *Br J Cancer* 2015;**112**:1665–74
15. von Roemeling CA, Marlow LA, Pinkerton AB, Crist A, Miller J, Tun HW, Smallridge RC, Copland JA. Aberrant lipid metabolism in anaplastic thyroid carcinoma reveals stearoyl CoA desaturase 1 as a novel therapeutic target. *J Clin Endocrinol Metab* 2015;**100**:E697–709
16. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003;**4**:249–64
17. Taminau J, Steenhoff D, Coletta A, Meganck S, Lazar C, de Schaetzen V, Duque R, Molter C, Bersini H, Nowe A, Weiss SD. inSilicoDb: an R/bioconductor package for accessing human affymetrix expert-curated datasets from GEO. *Bioinformatics* 2011;**27**:3204–5
18. Krabben K, Orth D, van der Kamp J. Combat as an interpersonal synergy: an ecological dynamics approach to combat sports. *Sports Med* 2019;**49**:1825–36
19. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007;**8**:118–27
20. 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 Res* 2015;**43**:e47
21. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;**4**:44–57
22. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res* 2011;**39**:W316–22
23. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 2017;**45**: D362–D68
24. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and Sub-networks from complex interactome. *BMC Syst Biol* 2014;**8(Suppl4)**:S11
25. Wang Z, Jensen MA, Zenklusen JC. A practical guide to the cancer genome atlas (TCGA). *Methods Mol Biol* 2016;**1418**:111–41
26. 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 Res* 2017;**45**:W98–W102
27. Pan D, Lin P, Wen D, Wei Y, Mo Q, Liang L, Chen G, He Y, Chen J, Yang H. Identification of down-regulated microRNAs in thyroid cancer and their potential functions. *Am J Transl Res* 2018;**10**:2264–76
28. Wen DY, Pan DH, Lin P, Mo QY, Wei YP, Luo YH, Chen G, He Y, Chen JQ, Yang H. Downregulation of miR4865p in papillary thyroid carcinoma tissue: a study based on microarray and miRNA sequencing. *Mol Med Rep* 2018;**18**:2631–42
29. Luo YH, Liang L, He RQ, Wen DY, Deng GF, Yang H, He Y, Ma W, Cai XY, Chen JQ, Chen G. RNA-sequencing investigation identifies an effective risk score generated by three novel lncRNAs for the survival of papillary thyroid cancer patients. *Oncotarget* 2017;**8**:74139–58
30. Song EL, Xing L, Wang L, Song WT, Li DB, Wang Y, Gu YW, Liu MM, Ni WJ, Zhang P, Ma X, Zhang X, Yao J, Chen Y, An RH. LncRNA ADAMTS9-AS2 inhibits cell proliferation and decreases chemoresistance in clear cell renal cell carcinoma via the miR-27a-3p/FOXO1 axis. *Aging* 2019;**11**:5705–25
31. Griffith OL, Melck A, Jones SJ, Wiseman SM. Meta-analysis and meta-review of thyroid cancer gene expression profiling studies identifies important diagnostic biomarkers. *J Clin Oncol* 2006;**24**:5043–51
32. Lazar C, Meganck S, Taminau J, Steenhoff D, Coletta A, Molter C, Weiss-Solis DY, Duque R, Bersini H, Nowe A. Batch effect removal methods for microarray gene expression data integration: a survey. *Brief Bioinform* 2013;**14**:469–90
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. *J Cell Physiol* 2019;**234**:23647–57
34. Liang W, Sun F. Identification of key genes of papillary thyroid cancer using integrated bioinformatics analysis. *J Endocrinol Invest* 2018;**41**:1237–45
35. Zhang S, Wang Q, Han Q, Han H, Lu P. Identification and analysis of genes associated with papillary thyroid carcinoma by bioinformatics methods. *Biosci Rep* 2019;**39**:BSR20190083
36. Yin Q, Liu S, Dong A, Mi X, Hao F, Zhang K. Targeting transforming growth Factor-Beta1 (TGF-beta1) inhibits tumorigenesis of anaplastic thyroid carcinoma cells through ERK1/2-NFkappakB-PUMA signaling. *Med Sci Monit* 2016;**22**:2267–77

37. Tang J, Kong D, Cui Q, Wang K, Zhang D, Yuan Q, Liao X, Gong Y, Wu G. Bioinformatic analysis and identification of potential prognostic microRNAs and mRNAs in thyroid cancer. *PeerJ* 2018;**6**:e4674
38. You S, Gao L. Identification of NMU as a potential gene conferring alectinib resistance in non-small cell lung cancer based on bioinformatics analyses. *Gene* 2018;**678**:137–42
39. Yap LF, Velapasamy S, Lee HM, Thavaraj S, Rajadurai P, Wei W, Vrzalikova K, Ibrahim MH, Khoo AS, Tsao SW, Paterson IC, Taylor GS, Dawson CW, Murray PG. Down-regulation of LPA receptor 5 contributes to aberrant LPA signalling in EBV-associated nasopharyngeal carcinoma. *J Pathol* 2015;**235**:456–65
40. Huang QX, Cui JY, Ma H, Jia XM, Huang FL, Jiang LX. Screening of potential biomarkers for cholangiocarcinoma by integrated analysis of microarray data sets. *Cancer Gene Ther* 2016;**23**:48–53
41. Song YS, Kim MJ, Sun HJ, Kim HH, Shin HS, Kim YA, Oh BC, Cho SW, Park YJ. Aberrant thyroid-stimulating hormone receptor signaling increases VEGF-A and CXCL8 secretion of thyroid cancer cells, contributing to angiogenesis and tumor growth. *Clin Cancer Res* 2019;**25**:414–25
42. Rotondi M, Coperchini F, Latrofa F, Chiovato L. Role of chemokines in thyroid cancer microenvironment: is CXCL8 the main player? *Front Endocrinol* 2018;**9**:314
43. Galdiero MR, Varricchi G, Loffredo S, Bellevicine C, Lansione T, Ferrara AL, Iannone R, di Somma S, Borriello F, Clery E, Triassi M, Troncone G, Marone G. Potential involvement of neutrophils in human thyroid cancer. *PLoS One* 2018;**13**:e0199740
44. Park HJ, Choe BK, Kim SK, Park HK, Kim JW, Chung JH, Hong IK, Chung DH, Kwon KH. Association between collagen type XI alpha1 gene polymorphisms and papillary thyroid cancer in a Korean population. *Exp Ther Med* 2011;**2**:1111–6
45. Qiu J, Zhang W, Xia Q, Liu F, Li L, Zhao S, Gao X, Zang C, Ge R, Sun Y. RNA sequencing identifies crucial genes in papillary thyroid carcinoma (PTC) progression. *Exp Mol Pathol* 2016;**100**:151–9

(Received January 24, 2021, Accepted March 20, 2021)