



OPEN ACCESS

EDITED BY

Manal S. Fawzy,
Suez Canal University, Egypt

REVIEWED BY

Eman Toraih,
Tulane University, United States
Bailing Zhou,
Dezhou University, China

*CORRESPONDENCE

Minggao Guo,
guominggao203@163.com

SPECIALTY SECTION

This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 31 May 2022

ACCEPTED 21 September 2022

PUBLISHED 07 October 2022

CITATION

Wang S, Liu W, Ye Z, Xia X and Guo M (2022), Development of a joint diagnostic model of thyroid papillary carcinoma with artificial neural network and random forest.
Front. Genet. 13:957718.
doi: 10.3389/fgene.2022.957718

COPYRIGHT

© 2022 Wang, Liu, Ye, Xia and Guo. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Development of a joint diagnostic model of thyroid papillary carcinoma with artificial neural network and random forest

Shoufei Wang, Wenfei Liu, Ziheng Ye, Xiaotian Xia and Minggao Guo*

Department of Thyroid, Parathyroid, Breast, and Hernia Surgery, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China

Objective: Papillary thyroid carcinoma (PTC) accounts for 80% of thyroid malignancy, and the occurrence of PTC is increasing rapidly. The present study was conducted with the purpose of identifying novel and important gene panels and developing an early diagnostic model for PTC by combining artificial neural network (ANN) and random forest (RF).

Methods and results: Samples were searched from the Gene Expression Omnibus (GEO) database, and gene expression datasets (GSE27155, GSE60542, and GSE33630) were collected and processed. GSE27155 and GSE60542 were merged into the training set, and GSE33630 was defined as the validation set. Differentially expressed genes (DEGs) in the training set were obtained by "limma" of R software. Then, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis as well as immune cell infiltration analysis were conducted based on DEGs. Important genes were identified from the DEGs by random forest. Finally, an artificial neural network was used to develop a diagnostic model. Also, the diagnostic model was validated by the validation set, and the area under the receiver operating characteristic curve (AUC) value was satisfactory.

Conclusion: A diagnostic model was established by a joint of random forest and artificial neural network based on a novel gene panel. The AUC showed that the diagnostic model had significantly excellent performance.

KEYWORDS

papillary thyroid carcinoma, random forest, artificial neural network, early diagnostic, biomarker

Introduction

Thyroid cancer (TC), accounting for 3.4% of all cancers diagnosed annually, is the most common endocrine malignancy (Seib and Sosa, 2019). TC was divided into three types: differentiated thyroid cancer, representing over 90% of thyroid cancer, consists of papillary thyroid carcinoma (PTC) and follicular thyroid carcinoma (FTC). In addition, anaplastic thyroid cancer (ATC, 1%) and poorly differentiated thyroid cancer (PDTC, 5%) are rare tumors (Prete et al., 2020). With the popularization of physical examination, the incidence of PTC has increased rapidly (Wang and Sosa, 2018). PTC is usually an inert and curable tumor with a 10-year survival rate >90% (Alvarado et al., 2009). However, more than 25% of advanced-stage PTC patients are characterized by their invasiveness and metastasis; these traits usually result in poor prognosis (Huang et al., 2021). Among PTC patients, cervical lymph node metastasis (LNM) occurs in 50%–80% of patients, which is a tested risk factor for recurrence and a reduced survival rate (Ling et al., 2021). Distant metastasis occurs in 2% of advanced-stage PTC, which is the main reason for death. Lung is the most common metastatic site (53.4%), and 26.3% presented with the multiple-organ metastasis (Toraih et al., 2021). So, we intended to discover novel and important gene panels and develop a diagnostic model for PTC of early stage.

The availability of microarray technology and more precise RNA-sequencing technology improves the research of disease

pathogenesis (Xie et al., 2020). Discovering the most meaningful variables for classification is the primary question about developing a classification model using gene expression data. To resolve this, a variety of machine learning algorithm, including random forest (RF) (Kursa, 2014; Cai et al., 2015) and artificial neural network (ANN) (Chen et al., 2014), were used. Differing from common statistical methods, machine learning involves learning from cases (Van Calster et al., 2019). Therefore, RF and ANN were joined to develop a novel diagnostic model of PTC by learning from the training set and then testing the model in the validation set. The results of this study reveal a novel gene panel for early clinical diagnosis of PTC.

Materials and methods

Research design

The study flowchart is shown in Figure 1. Three gene expression datasets (GSE27155, GSE60542 and GSE33630) were collected from the GEO database. GSE27155 and GSE60542 were integrated into the training set, and GSE33630 was selected as the validation set. Differentially expressed genes (DEGs) were defined by the R package “limma.” Gene Ontology (GO) function enrichment analyses and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were conducted through the “clusterProfiler” package of R based on the DEGs in the training set. Also, immune cell infiltration analysis was performed by R. We determined 10 important genes through random forest from the DEGs by the “randomForest” package. Further, an artificial neural network diagnostic model was established through the “neuralnet” package using the 10 important genes and was assessed by AUC. Finally, the validity of the ANN diagnostic model was validated with the performance of a validation set.

Data download and process

The GSE27155, GSE60542, and GSE33630 datasets were collected *via* the Gene Expression Omnibus database (GEO; <https://www.ncbi.nlm.nih.gov/geo/>). Then, gene names were obtained by transforming probe names by R software. If several probes could be matched to identical gene, the

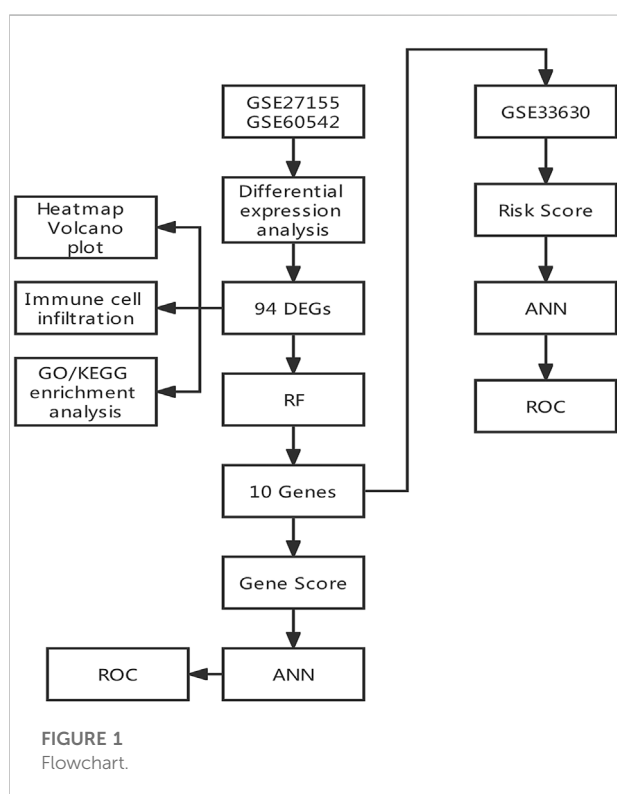


TABLE 1 Source of datasets.

Datasets	Platform	Normal	PTC
GSE27155	GPL96	4	51
GSE60542	GPL570	30	33
GSE33630	GPL570	45	49

expression data with gene was replaced by its mean expression value. Finally, 55 samples (4 normal samples and 51 PTC samples) in GSE27155, 63 samples (30 normal samples and 33 PTC samples) in GSE60542 and 94 samples (45 normal samples and 49 PTC samples) in GSE33630 were utilized in this research (Table 1). GSE27155 and GSE60542 were combined into a training set, and GSE33630 as a validation set.

Screening for differentially expressed genes

Differentially expressed genes (DEGs) were identified in 34 normal samples and 84 PTC samples of the training set through the R package “limma” using the classic Bayesian data analysis. $|\log_2FC| > 1.5$ and adjusted p Value < 0.05 were set as threshold. Then, we obtained 94 DEGs involving 53 up-regulated genes and 41 down-regulated genes. R package “pheatmap” and “ggplot2” were utilized to conduct the heatmap and volcano plot.

Gene Ontology function and Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis

To explore the biological significance of these DEGs, Gene Ontology (GO) enrichment analysis (adjusted p value < 0.05) categorizing genes into biological process (BP), cellular component (CC), and molecular function (MF) was performed. Meantime, Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was used to describe metabolic pathways (p value < 0.05). Enrichment analysis of DEGs was conducted through the R package “clusterProfiler.”

Immune cell infiltration analysis

As a crucial component of the tumor environment, immune cells impacted the development and prognosis of tumors, their composition as well as function in various types of tumors were different. On the one hand, some immune cells have the roles to be favorable targets for immunotherapy. On the other hand, some may react negatively, even resulting in drug resistance. Given these causes, figuring out the ingredient and possible influence of immune cells in PTC was beneficial to identify valuable therapeutic targets. We downloaded a reference including human gene expression in immune cells. Then, we exploited the reference and a file including the gene expression of each sample to complete immune cell infiltration analysis and obtain the result involving the expression of immune cells of each sample by R software. Based on the result, we

performed correlation analysis of immune cell by the “corrplot” package and difference analysis of immune cells by the “vioplot” package.

Random forest analysis to determine important genes

The important genes were identified by a random forest classifier *via* “randomForest” package of R. Firstly, the parameter “ntree” was set as 500 to find the best number of trees. By calculating the error of cross validation, 15 was selected as the best number of trees represented the minimum error of cross validation. Then, parameter 15 was used to reconduct random forest and the importance score of genes was obtained. Genes with an importance score greater than 2 were seen as PTC significantly related genes. The R package “heatmap” was utilized to draw a heatmap based on important genes.

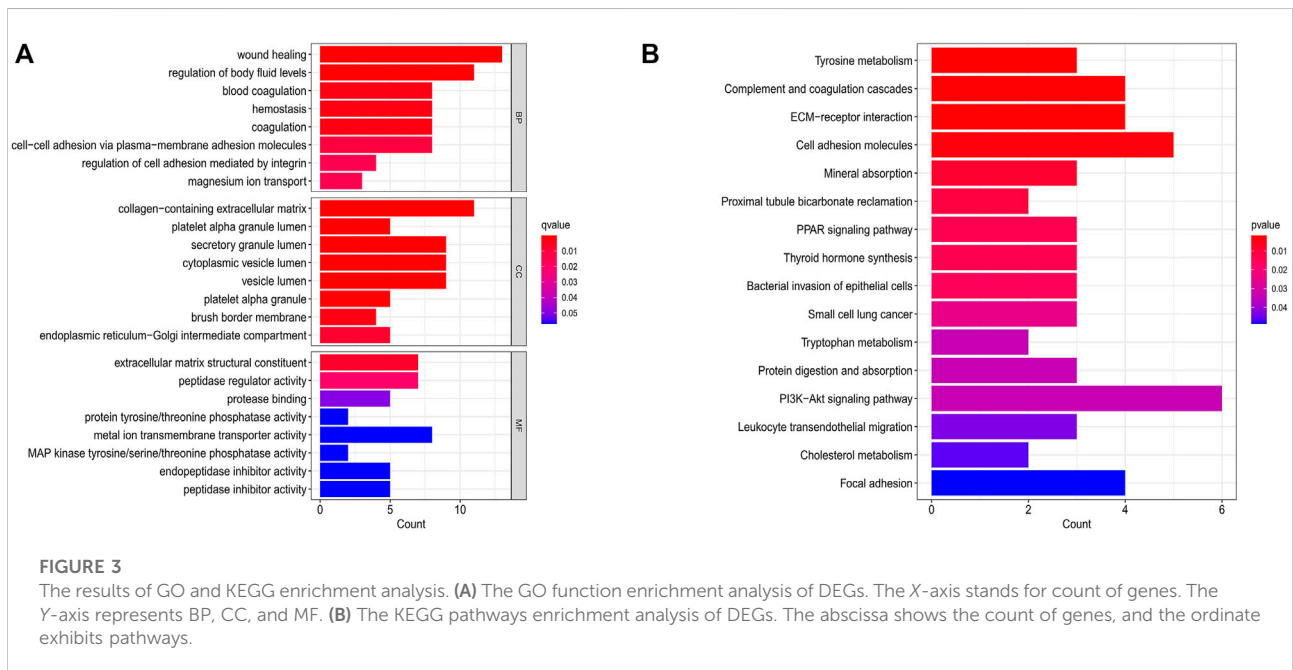
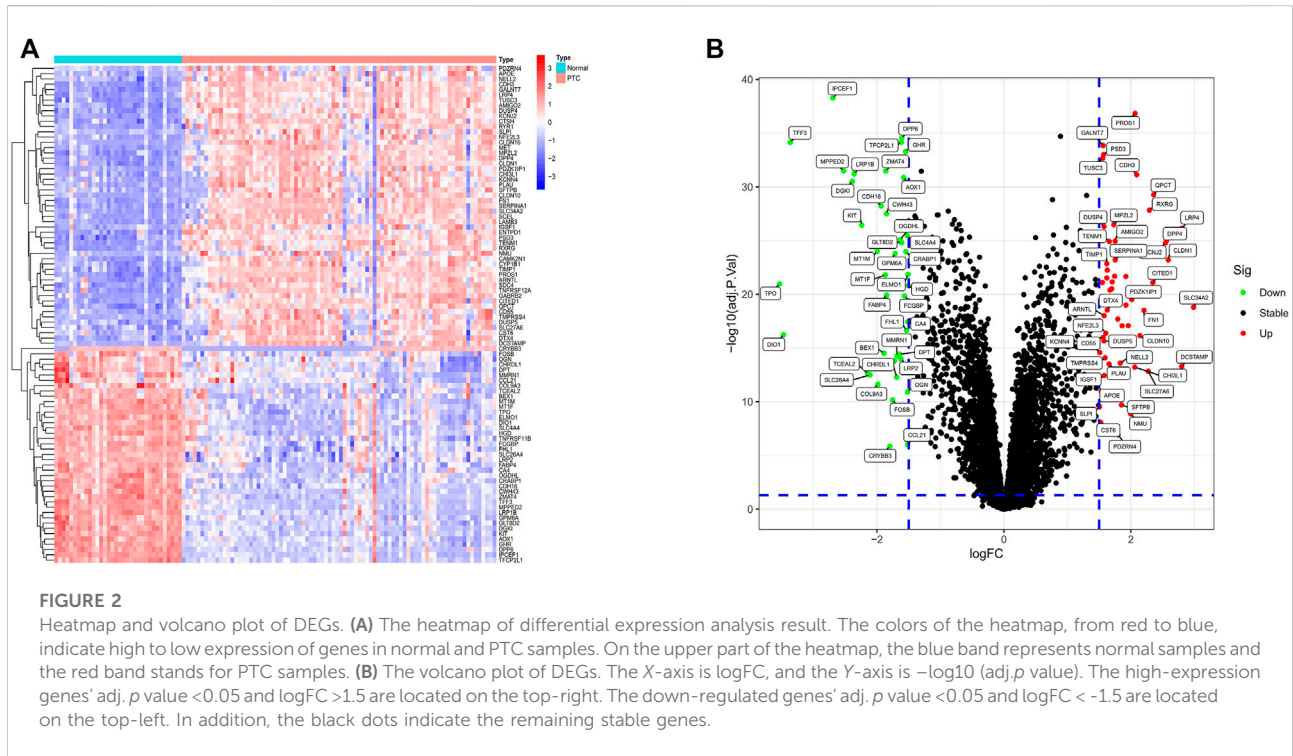
Development and validation of an artificial neural network model

Based on 10 important genes identified by RF, an artificial neural network model was developed *via* the R package “neuralnet.” At first, the expression of 10 important genes was converted into “gene tag” based on their expression levels. In the case of a certain sample, the expression level of a specific gene was compared to the median of all sample expression level. Among the up-regulated genes, if the expression level is higher than the median, it will be valued as 1, otherwise 0. Among the down-regulated gene, if expression level is higher than median, it will be valued as 0, otherwise 1. Then, we finished a “gene tag” sheet. Next, the hidden layers of ANN were set as 5 to obtain a gene weight calculated by “gene tag”. Finally, the ANN diagnostic model was established. We assessed the model in the training set. Also, we validated the model in the validation set, and its diagnostic performance was evaluated by AUC.

Results

Differential expression analysis

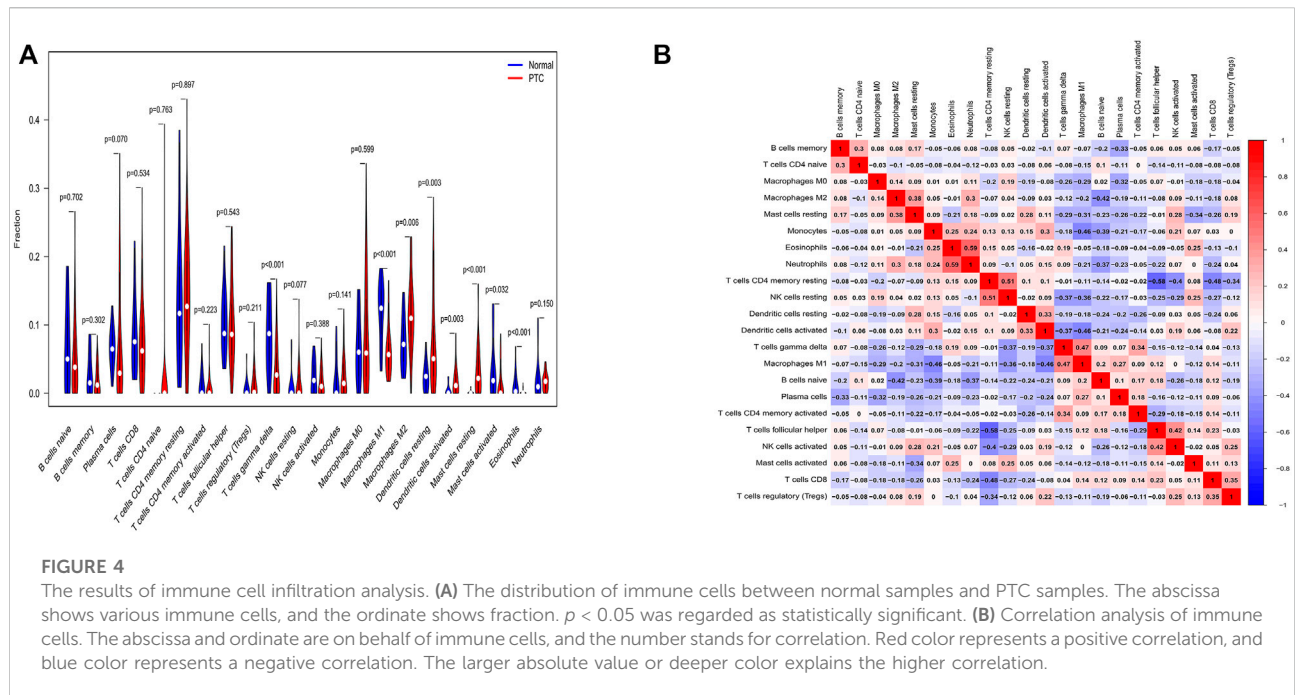
In total, 53 significantly up-regulated genes and 41 significantly down-regulated genes were determined between normal samples and PTC samples based on $|\log_2FC| > 1.5$ and adjusted p value < 0.05 as threshold in training sets GSE27155 and GSE60542. Heatmaps (Figure 2A) and volcano plots (Figure 2B) of DEGs showed favorable discrimination of gene expression.



Gene Ontology/Kyoto Encyclopedia of Genes and Genomes enrichment analysis

To reveal the biological importance of DEGs in the mechanism of PTC, we conducted GO and KEGG enrichment

analysis using the package “clusterProfiler” of R software. Amid GO enrichment analysis results (adjusted p value cutoff = 0.05), concerning BP included wound healing, regulation of body fluid levels, blood coagulation and hemostasis; CC involved collagen-containing extracellular matrix; MF contained extracellular



matrix structural constituents and other important functions (Figure 3A). KEGG pathway enrichment analysis (p value cutoff = 0.05) indicated that the DEGs were significantly related to tyrosine metabolism, complement and coagulation cascades, ECM-receptor interaction, and cell adhesion molecules (Figure 3B).

Immune cell infiltration analysis

We utilized a reference including human gene expression in immune cells and a file involving gene expression of each sample to perform immune cell infiltration analysis. Then, we conducted difference analysis of immune cells by the “vioplot” package and correlation analysis of immune cells by the “corrplot” package. Compared with normal samples, T cells gamma delta, macrophages M2, dendritic cells resting, dendritic cells activated and mast cells resting were higher in PTC samples ($p < 0.05$). Conversely, macrophages M1, mast cells activated, and eosinophils were lower in PTC ($p < 0.05$) (Figure 4A). And macrophages M1 were correlated with dendritic cells activated and T cells gamma delta (correlation, 0.46 and 0.47) (Figure 4B).

Diagnosis-related important genes with random forest

We performed a random forest based on 94 DEGs to determine important genes. Considering the relationship plot

between error of cross validation and the number of decision trees, we chose 15 trees as the parameter of the final model, which revealed a minimum error of cross validation in the model (Figure 5A). Eventually, 10 important genes were determined on the condition that their importance score was greater than 2 for the following analysis (Figure 5B). Amid the 10 important genes, GALNT7 was the most important. With heatmap, 10 important genes could be differed from normal samples and PTC samples. Among them, CITED1, AMIGO2, PSD3, GALNT7, and PROS1 were down-regulated in normal samples and up-regulated in PTC samples. TFF3, SLC4A4, AOX1, IPCEF1, and TFCEP2L1 were up-regulated in normal samples and down-regulated in PTC samples (Figure 5C).

Development and validation of an artificial neural network model

Expression of important genes determined by RF was transformed into “gene tag” marked as 0/1. The weight of all genes was calculated for optimal discrimination between normal samples and PTC samples. Then, an ANN diagnostic model based on gene weight was established (Figure 6). Performance of the model had an AUC of 0.988 (Figure 7A) in the training set and 0.968 (Figure 7B) in the validation set, indicating that the model was very satisfactory in diagnosing PTC. The results demonstrated that we had developed a precise diagnostic model between PTC and normal samples.

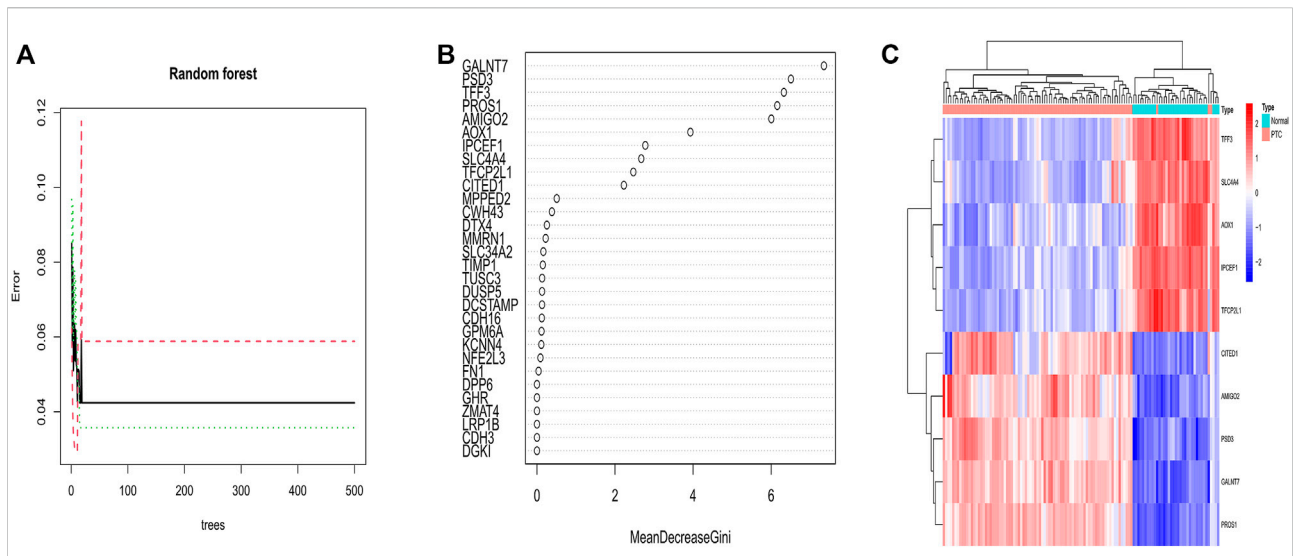


FIGURE 5 (A) Correlation between the number of decision trees and the error of cross validation. The number of decision trees is shown on the abscissa; the error of cross validation is exhibited on the ordinate. The best number of decision trees was 15 for which the error of cross validation was minimum. (B) The X-axis stands for the importance score of genes calculated by the Gini coefficient method. The Y-axis represents the names of genes. (C) The heatmap of 10 important genes determined by random forest. Red color represents up-regulated genes in both samples, and blue color is on behalf of down-regulated genes in both samples. Above the picture, PTC samples are showed as a red band, and blue band indicates normal samples.

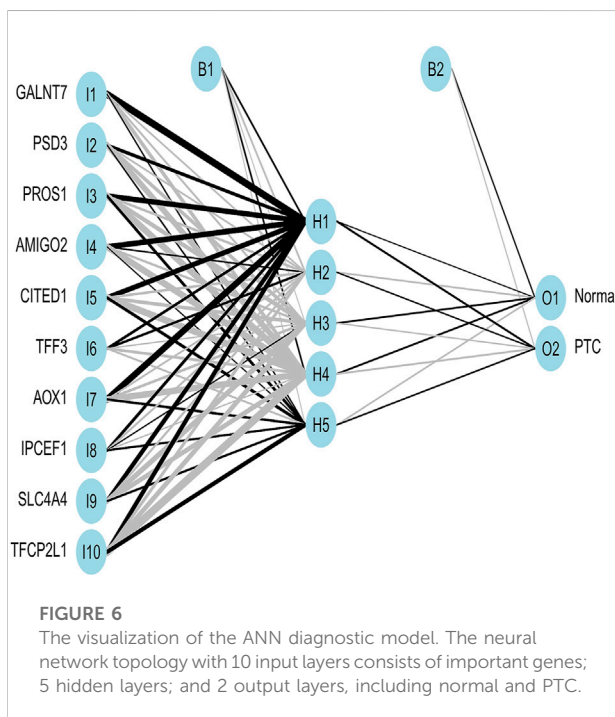


FIGURE 6 The visualization of the ANN diagnostic model. The neural network topology with 10 input layers consists of important genes; 5 hidden layers; and 2 output layers, including normal and PTC.

Discussion

To date, employment of machine learning algorithms and wide application of gene expression data in public databases offer methods to find biomarkers for cancer diagnosis in a list of fields (Wang et al., 2018; Tabl et al., 2019). The integration of RF and ANN can be employed to establish a stable diagnostic model for some diseases, like ulcerative colitis and abdominal aortic aneurysm (Li et al., 2020; Duan et al., 2022). Papillary thyroid carcinoma is characterized by slow progression and good prognosis. Early-stage PTC patients have a high postoperative survival rate, but advanced-stage PTC patients still have the risk of lymph node metastasis and distant metastasis, which seriously affect treatment and prognosis. Though ultrasound is viewed as primary screening approach for PTC, the diagnostic accuracy relying on node size is changeable, especially for nodules ≤ 10 mm (Sutherland et al., 2021). Other factors such as equipment, scan gain, dynamic range, frequency, and doctors also significantly impact the accuracy of ultrasound diagnosis (Gong et al., 2022). Owing to the absence of a perfect diagnostic approach and a lack of potential biomarkers that can be utilized in clinical practice, it is critical to establish a model for early diagnosis and screening of PTC.

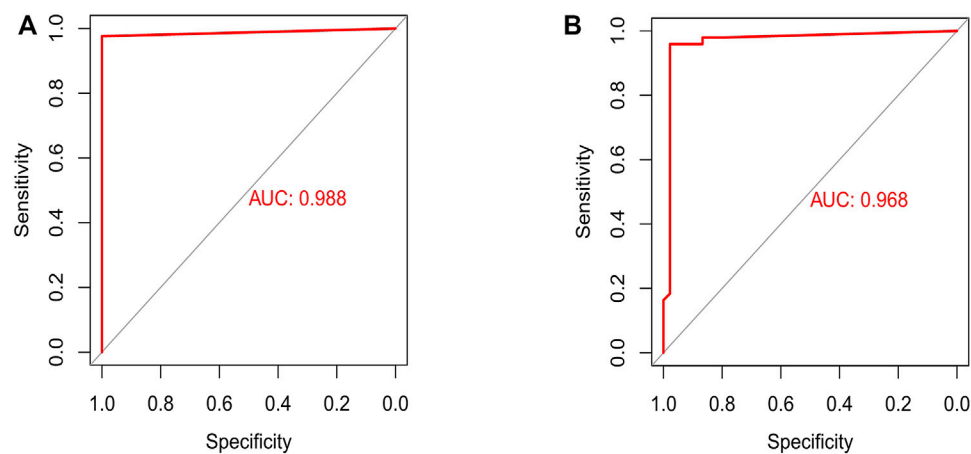


FIGURE 7

Assessment and testification of the ANN diagnostic model by the ROC curve. (A) The assessment result of the training set; (B) the testification result of the validation set.

Among 10 important genes identified by RF, CITED1, AMIGO2, PSD3, GALNT7, and PROS1 were down-up-regulated in normal samples and up-regulated in PTC samples, TFF3, SLC4A4, AOX1, IPCEF1, and TFCP2L1 were up-regulated in normal samples and down-regulated in PTC samples. Then, an ANN diagnostic model was established, and the performance was assessed by AUC (0.988). Also, we testified the diagnostic ability in the validation set and the AUC was 0.968, which had great efficiency. Together, the developed diagnostic model could offer a novel perspective on our research of the mechanism of PTC.

In a previous study, 9 genes have been verified among 10 important genes related to PTC. The incidence and progression of a series of malignancies are relevant to the high expression of Polypeptide N-Acetylgalactosaminyltransferase7 (GALNT7) and its family members (Nakagawa et al., 2017; Detarya et al., 2020). Wang argued that GALNT7 by activating EGFR/PI3K/AKT kinase pathway to promote cell proliferation and invasion of papillary thyroid cancer (Wang et al., 2021a). In recent years, various human Pleckstrin and sec7 domain-containing 3 (PSD3) have been believed to be related to some tumors, like acute myeloid leukemia (Walker et al., 2021), breast cancer metastasis (Thomassen et al., 2009), astrocytoma progression (van den Boom et al., 2006). PSD3 inhibits apoptosis in papillary thyroid cancer by promoting proliferation, migration, invasion and G1/S transition (Jin et al., 2021). Abnormal expression of Protein S (PROS1) affects human papillary thyroid cancer progression, especially associated with lymph node metastasis (Wang et al., 2021b). Low expression of SLC4A4 affects invasion, metastasis, and the MAPK signaling pathway in PTC. Huang argued the

down-regulation of SLC4A4 might be on account of the excessive iodine intakes of patients (Huang et al., 2021). Glu/Asp rich carboxy-terminal domain 1 (CITED1) via Wnt/ β -catenin signaling pathway results in the development of PTC (Wang et al., 2019). Li showed that CITED1 increased phosphorylation of pRb as well as E2F1 transcriptional activity when p21 and p27 were expressed at low levels, and verified CITED1 was involved in PTC cell proliferation and tumorigenesis (Li et al., 2018). Yu reported that aldehyde oxidase 1 (AOX1) protein level in blood plasma was lower in patients with PTC, which indicated that AOX1 level in blood plasma had the potential to differ in PTC from healthy humans. Furthermore, low levels of AOX1 were highly related to poor survival of PTC (Yu et al., 2021). IPCEF1 was viewed as a significant biomarker for PTC. Moreover, a study showed that the hsa_circ_IPCEF1/hsa-miR-3619-5p axis was associated with the mechanism of PTC, which offers a new idea for further diagnosis and treatment of PTC (Guo et al., 2021). Several researchers revealed noticeable differences in Trefoil factor 3 (TFF3) between benign thyroid nodules and thyroid malignancy (Krause et al., 2008; Karger et al., 2012). Low expression of TFCP2L1 can promote the progression of PTC and CircHACE1 curbs PTC development by upregulating TFCP2L1 through adsorbing miR-346 (Li et al., 2021). Interestingly, we identified another important gene, Adhesion Molecule With Ig Like Domain 2 (AMIGO2), which has never been reported to be associated with PTC.

In addition, the research also has several limitations. Firstly, though we have validated 10 significantly PTC-related genes, the sample size is relatively small. Secondly, the ANN diagnostic model was conducted using datasets from the GEO database, so it should be tested in laboratory experiments and clinical practice.

Conclusion

In this study, 10 genetic biomarkers related to PTC were determined and the ANN model established by the 10-gene panel displayed satisfactory performance when diagnosing PTC. Moreover, the present research offers a useful basis for early screening of PTC and promotes further study for development of PTC as well as provides potential genes as targets for clinical treatment. In conclusion, our finding has a certain clinical value that can be valuable for early diagnosis of PTC.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

Author contributions

SW and MG conceived of and designed this study. WL, ZY, and XX made literature searches. SW performed figures and tables. WL and ZY conducted data analysis. SW wrote the manuscript, and MG critically reviewed it. MG supervised the

research. All authors have read and approved the final manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.957718/full#supplementary-material>

References

- Alvarado, R., Sywak, M. S., Delbridge, L., and Sidhu, S. B. (2009). Central lymph node dissection as a secondary procedure for papillary thyroid cancer: Is there added morbidity? *Surgery* 145, 514–518. doi:10.1016/j.surg.2009.01.013
- Cai, Z., Xu, D., Zhang, Q., Zhang, J., Ngai, S. M., and Shao, J. (2015). Classification of lung cancer using ensemble-based feature selection and machine learning methods. *Mol. Biosyst.* 11, 791–800. doi:10.1039/c4mb00659c
- Chen, Y. C., Ke, W. C., and Chiu, H. W. (2014). Risk classification of cancer survival using ANN with gene expression data from multiple laboratories. *Comput. Biol. Med.* 48, 1–7. doi:10.1016/j.compbiomed.2014.02.006
- Detarya, M., Sawanyawisuth, K., Aphivatanasiri, C., Chuangchaiya, S., Saranaruk, P., Sukprasert, L., et al. (2020). The O-GalNAcylating enzyme GALNT5 mediates carcinogenesis and progression of cholangiocarcinoma via activation of AKT/ERK signaling. *Glycobiology* 30, 312–324. doi:10.1093/glycob/cwz098
- Duan, Y., Xie, E., Liu, C., Sun, J., and Deng, J. (2022). Establishment of a combined diagnostic model of abdominal aortic aneurysm with random forest and artificial neural network. *Biomed. Res. Int.* 2022, 7173972. doi:10.1155/2022/7173972
- Gong, Y., Yao, X., Yu, L., Wei, P., Han, Z., Fang, J., et al. (2022). Ultrasound grayscale ratio: A reliable parameter for differentiating between papillary thyroid microcarcinoma and micronodular goiter. *BMC Endocr. Disord.* 22, 75. doi:10.1186/s12902-022-00994-9
- Guo, M., Sun, Y., Ding, J., Li, Y., Yang, S., Zhao, Y., et al. (2021). Circular RNA profiling reveals a potential role of hsa_circ_IPCEF1 in papillary thyroid carcinoma. *Mol. Med. Rep.* 24, 603. doi:10.3892/mmr.2021.12241
- Huang, F., Wang, H., Xiao, J., Shao, C., Zhou, Y., Cong, W., et al. (2021). SLC34A2 up-regulation and SLC4A4 down-regulation correlates with invasion, metastasis, and the MAPK signaling pathway in papillary thyroid carcinomas. *J. Cancer* 12, 5439–5453. doi:10.7150/jca.56730
- Jin, L., Zheng, D., Bhandari, A., Chen, D., Xia, E., Guan, Y., et al. (2021). PSD3 is an oncogene that promotes proliferation, migration, invasion, and G1/S transition while inhibits apoptotic in papillary thyroid cancer. *J. Cancer* 12, 5413–5422. doi:10.7150/jca.60885
- Karger, S., Krause, K., Gutknecht, M., Schierle, K., Graf, D., Steinert, F., et al. (2012). ADM3, TFF3 and LGALS3 are discriminative molecular markers in fine-needle aspiration biopsies of benign and malignant thyroid tumours. *Br. J. Cancer* 106, 562–568. doi:10.1038/bjc.2011.578
- Krause, K., Eszlinger, M., Gimm, O., Karger, S., Engelhardt, C., Dralle, H., et al. (2008). TFF3-based candidate gene discrimination of benign and malignant thyroid tumors in a region with borderline iodine deficiency. *J. Clin. Endocrinol. Metab.* 93, 1390–1393. doi:10.1210/jc.2006-1255
- Kursa, M. B. (2014). Robustness of Random Forest-based gene selection methods. *BMC Bioinforma.* 15, 8. doi:10.1186/1471-2105-15-8
- Li, H., Guan, H., Guo, Y., Liang, W., Liu, L., He, X., et al. (2018). CITED1 promotes proliferation of papillary thyroid cancer cells via the regulation of p21 and p27. *Cell Biosci.* 8, 57. doi:10.1186/s13578-018-0256-9
- Li, H., Lai, L., and Shen, J. (2020). Development of a susceptibility gene based novel predictive model for the diagnosis of ulcerative colitis using random forest and artificial neural network. *Aging (Albany NY)* 12, 20471–20482. doi:10.18632/aging.103861
- Li, X., Yang, S., Zhao, C., Yang, J., Li, C., Shen, W., et al. (2021). CircHACE1 functions as a competitive endogenous RNA to curb differentiated thyroid cancer progression by upregulating Tfcp2L1 through adsorbing miR-346. *Endocr. J.* 68, 1011–1025. doi:10.1507/endocrj.EJ20-0806
- Ling, Y., Jia, L., Li, K., Zhang, L., Wang, Y., and Kang, H. (2021). Development and validation of a novel 14-gene signature for predicting lymph node metastasis in papillary thyroid carcinoma. *Gland. Surg.* 10, 2644–2655. doi:10.21037/gs-21-361
- Nakagawa, Y., Nishikimi, T., Kuwahara, K., Fujishima, A., Oka, S., Tsutamoto, T., et al. (2017). MiR30-GALNT1/2 axis-mediated glycosylation contributes to the increased secretion of inactive human pro-hormone for brain natriuretic peptide (proBNP) from failing hearts. *J. Am. Heart Assoc.* 6, e003601. doi:10.1161/JAHA.116.003601
- Prete, A., Borges de Souza, P., Censi, S., Muzza, M., Nucci, N., and Sponziello, M. (2020). Update on fundamental mechanisms of thyroid cancer. *Front. Endocrinol.* 11, 102. doi:10.3389/fendo.2020.00102

- Seib, C. D., and Sosa, J. A. (2019). Evolving understanding of the epidemiology of thyroid cancer. *Endocrinol. Metab. Clin. North Am.* 48, 23–35. doi:10.1016/j.ecl.2018.10.002
- Sutherland, R., Tsang, V., Clifton-Bligh, R. J., and Gild, M. L. (2021). Papillary thyroid microcarcinoma: Is active surveillance always enough? *Clin. Endocrinol.* 95, 811–817. doi:10.1111/cen.14529
- Tabl, A. A., Alkhateeb, A., ElMaraghy, W., Rueda, L., and Ngom, A. (2019). A machine learning approach for identifying gene biomarkers guiding the treatment of breast cancer. *Front. Genet.* 10, 256. doi:10.3389/fgene.2019.00256
- Thomassen, M., Tan, Q., and Kruse, T. A. (2009). Gene expression meta-analysis identifies chromosomal regions and candidate genes involved in breast cancer metastasis. *Breast Cancer Res. Treat.* 113, 239–249. doi:10.1007/s10549-008-9927-2
- Toraih, E. A., Hussein, M. H., Zerfaoui, M., Attia, A. S., Ellythy, A. M., Mostafa, A., et al. (2021). Site-specific metastasis and survival in papillary thyroid cancer: The importance of brain and multi-organ disease. *Cancers (Basel)* 13, 1625. doi:10.3390/cancers13071625
- Van Calster, B., Wynants, L., and Mamdani, M. (2019). Machine learning in medicine. *N. Engl. J. Med.* 380, 2588–2589. doi:10.1056/NEJMc1906060
- van den Boom, J., Wolter, M., Blaschke, B., Knobbe, C. B., and Reifenberger, G. (2006). Identification of novel genes associated with astrocytoma progression using suppression subtractive hybridization and real-time reverse transcription-polymerase chain reaction. *Int. J. Cancer* 119, 2330–2338. doi:10.1002/ijc.22108
- Walker, C. J., Mrozek, K., Ozer, H. G., Nicolet, D., Kohlschmidt, J., Papaioannou, D., et al. (2021). Gene expression signature predicts relapse in adult patients with cytogenetically normal acute myeloid leukemia. *Blood Adv.* 5, 1474–1482. doi:10.1182/bloodadvances.2020003727
- Wang, D., Li, J. R., Zhang, Y. H., Chen, L., Huang, T., and Cai, Y. D. (2018). Identification of differentially expressed genes between original breast cancer and xenograft using machine learning algorithms. *Genes (Basel)* 9, E155. doi:10.3390/genes9030155
- Wang, J., Lei, M., and Xu, Z. (2021). Aberrant expression of PROS1 correlates with human papillary thyroid cancer progression. *PeerJ* 9, e11813. doi:10.7717/peerj.11813
- Wang, T. S., and Sosa, J. A. (2018). Thyroid surgery for differentiated thyroid cancer - recent advances and future directions. *Nat. Rev. Endocrinol.* 14, 670–683. doi:10.1038/s41574-018-0080-7
- Wang, Y., Huang, H., Hu, F., Li, J., Zhang, L., and Pang, H. (2019). CITED1 contributes to the progression of papillary thyroid carcinoma via the Wnt/ β -catenin signaling pathway. *Oncotargets Ther.* 12, 6769–6777. doi:10.2147/OTT.S215025
- Wang, Y., Wang, C., Fu, Z., Zhang, S., and Chen, J. (2021). miR-30b-5p inhibits proliferation, invasion, and migration of papillary thyroid cancer by targeting GALNT7 via the EGFR/PI3K/AKT pathway. *Cancer Cell Int.* 21, 618. doi:10.1186/s12935-021-02323-x
- Xie, N. N., Wang, F. F., Zhou, J., Liu, C., and Qu, F. (2020). Establishment and analysis of a combined diagnostic model of polycystic ovary syndrome with random forest and artificial neural network. *Biomed. Res. Int.* 2020, 2613091. doi:10.1155/2020/2613091
- Yu, Y., Wang, S., Zhang, X., Xu, S., Li, Y., Liu, Q., et al. (2021). Clinical implications of TPO and AOX1 in pediatric papillary thyroid carcinoma. *Transl. Pediatr.* 10, 723–732. doi:10.21037/tp-20-301