



Prioritizing Susceptible Genes for Thyroid Cancer Based on Gene Interaction Network

Lin-kun Zhong^{1†}, Chang-lian Xie^{2†}, Shan Jiang^{3†}, Xing-yan Deng⁴, Xiao-xiong Gan⁵, Jian-hua Feng⁵, Wen-song Cai⁵, Chi-zhuai Liu¹, Fei Shen⁵, Jian-hang Miao^{1*} and Bo Xu^{5*}

¹ Department of General Surgery, Zhongshan City People's Hospital, Zhongshan, China, ² Intensive Care Unit, Zhongshan Hospital of Traditional Chinese Medicine Affiliated to Guangzhou University of Chinese Medicine, Zhongshan, China, ³ Reproductive Medicine Center, Boai Hospital of Zhongshan, Zhongshan, China, ⁴ Department of Thyrovascular Surgery, Maoming People's Hospital, Maoming, China, ⁵ Department of Thyroid Surgery, Guangzhou First People's Hospital, School of Medicine, South China University of Technology, Guangzhou, China

OPEN ACCESS

Edited by:

Lei Deng,
Central South University, China

Reviewed by:

Shihua Zhang,
Wuhan University of Science
and Technology, China
Junwei Han,
Harbin Medical University, China

*Correspondence:

Jian-hang Miao
lymjh@yeah.net
Bo Xu
eyboxu@scut.edu.cn

[†]These authors share first authorship

Specialty section:

This article was submitted to
Molecular and Cellular Pathology,
a section of the journal
Frontiers in Cell and Developmental
Biology

Received: 12 July 2021

Accepted: 02 August 2021

Published: 23 August 2021

Citation:

Zhong L-k, Xie C-l, Jiang S,
Deng X-y, Gan X-x, Feng J-h,
Cai W-s, Liu C-z, Shen F, Miao J-h
and Xu B (2021) Prioritizing
Susceptible Genes for Thyroid Cancer
Based on Gene Interaction Network.
Front. Cell Dev. Biol. 9:740267.
doi: 10.3389/fcell.2021.740267

Thyroid cancer ranks second in the incidence rate of endocrine malignant cancer. Thyroid cancer is usually asymptomatic at the initial stage, which makes patients easily miss the early treatment time. Combining genetic testing with imaging can greatly improve the diagnostic efficiency of thyroid cancer. Researchers have discovered many genes related to thyroid cancer. However, the effects of these genes on thyroid cancer are different. We hypothesize that there is a stronger interaction between the core genes that cause thyroid cancer. Based on this hypothesis, we constructed an interaction network of thyroid cancer-related genes. We traversed the network through random walks, and sorted thyroid cancer-related genes through ADNN which is fusion of Adaboost and deep neural network (DNN). In addition, we discovered more thyroid cancer-related genes by ADNN. In order to verify the accuracy of ADNN, we conducted a fivefold cross-validation. ADNN achieved AUC of 0.85 and AUPR of 0.81, which are more accurate than other methods.

Keywords: thyroid cancer, gene interaction, genetic testing, Adaboost, deep neural network

INTRODUCTION

Thyroid carcinoma (TC) is the most common malignant tumor of endocrine system, accounting for 2.5% of all human cancers, accounting for 90% (Maniakas et al., 2020; Sahu and Pattanayak, 2020; Xia et al., 2020) are cell-derived thyroid malignancies are derived from the follicular cells, including follicular thyroid carcinoma (FTC), papillary thyroid carcinoma (PTC), poorly differentiated thyroid carcinoma (PDTC), and anaplastic thyroid carcinoma (ATC) (Dralle et al., 2015). PTC and FTC with low malignancy are classified as differentiated thyroid carcinoma (DTC), accounting for about 90% of all thyroid cancers (Zanella et al., 2021). And the majority of deaths from thyroid carcinoma was caused by ATC (Molinaro et al., 2017). Medullary thyroid carcinoma (MTC) originates from parafollicular (c) cells, accounts for 2–4% of all thyroid carcinoma (Ceolin et al., 2019; Chen et al., 2020). About 25% of MTC cases are caused by germline genetic mutations, that is, familial medullary thyroid carcinoma (FMTC), while 75% are sporadic cases. Hereditary cases can occur alone, it can also be used as a part of multiple endocrine neoplasia type 2 (men2) syndrome (Vijayan et al., 2021).

According to statistics, the incidence rate of thyroid malignancies in recent decades is almost entirely due to the improvement of diagnostic accuracy and over diagnosis of PTC tumors, while the incidence rate of FTC, ATC, and MTC remains relatively stable (Xing, 2019; Zhang et al., 2021). The degree of differentiation of PTC is relatively high, and the corresponding degree of malignancy is relatively low, but it is not equivalent to the low risk of PTC. There are generally no obvious symptoms in the early stage of PTC (Tsukatani et al., 2020), but once clinical symptoms appear, such as hoarseness, Tracheal compression, etc., usually have entered the local advanced stage, and the best time for treatment has been missed at this time, and the metastasis of cervical lymph nodes and the invasion of local muscles, nerves and other tissues can often be seen during surgical treatment, resulting in postoperative complications. The treatment effect is not satisfactory.

Molecular markers are an effective tool for diagnosis, especially for thyroid nodules whose Fine needle aspiration cytology (FNAC) is uncertain (Sanguedolce et al., 2015). Gene mutation and chromosome rearrangement are important genetic changes in the occurrence and development of thyroid cancer. The molecular pathogenesis of most thyroid cancer involves mitogen activated protein kinase (MAPK) and phosphatidylinositide 3-kinases/protein kinase B (Sui et al., 2014), PI3K/Akt signaling pathway (Petrulea et al., 2015) is out of balance. BRAF and RAS point mutations, RET/PCT and Pax8/PPAR γ Rearrangement can activate MAPK pathway, and mainly occurs in DTC. BRAF mutation and RET/PCT rearrangement are common in PTC, while Ras mutation and Pax8/PPAR γ Rearrangement is a common molecular change in FTC. Pik3/Akt pathway is mainly activated by Ras, TP53 and TERTP mutations. TP53 and TERTP mutations are rare in well-differentiated thyroid cancer (Penna et al., 2016), and the mutation frequency is high in ATC and PDTC, which may be related to tumor invasion. Therefore, FNAC is difficult to determine the benign and malignant thyroid nodules, which can be combined with relevant molecular detection to help diagnosis, so as to improve the diagnostic accuracy of thyroid cancer.

According to the data released by the Cancer Genome Atlas (TCGA) in 2014 (Tomczak et al., 2015), 402 patients with thyroid cancer were analyzed. Compared with other cancers, the frequency of gene somatic mutations in thyroid cancer is relatively low. The frequency of BRAF V600 was 58.5%, which was the highest mutation site in thyroid cancer. In addition, the high-frequency mutation gene also includes three RAS gene family members, such as NRAS and KRAS, which are known tumor related genes, with a mutation frequency of 12.9% in European and American populations. In addition, some new thyroid cancer driving genes eif1ax, ppm1d, and CHEK2 were identified, and some of them also had gene fusion (Agrawal et al., 2014). TCGA research, with large sample size and various analysis methods, not only found a large number of somatic mutations, but also copy number variation and gene fusion information, which has great reference value and clinical significance. BRAF gene is used to assist in the diagnosis of benign and malignant thyroid nodules (Salvatore et al., 2006), which greatly reduces the misdiagnosis rate in clinical diagnosis and improves the

accuracy of preoperative diagnosis of patients with papillary thyroid cancer. BRAF gene is used for clinical diagnosis of papillary thyroid cancer (Trovisco et al., 2004). It can also be used as an important factor for clinical prediction of post-operative recurrence and guidance of medication, so as to facilitate the formulation of individualized and precise diagnosis, treatment and follow-up plans. In recent years, deep learning methods have been widely used in the diagnosis thyroid cancer. Lee JH et al. developed a deep learning-based computer-aided diagnosis (CAD) system could accurately classify cervical lymph node metastasis (LNM) on CT images in patients with thyroid cancer (Lee et al., 2019). Li X et al. indicated that compared with a group of skilled radiologists, deep convolutional neural network (DCNN) models that showed similar sensitivity and improved the diagnostic accuracy of thyroid cancer on sonographic images (Li et al., 2019).

Although multiple TC-related genes have been found by collecting samples and implementing gene differential expression analysis (Zhao et al., 2021b), people are still unclear about the pathogenesis and early diagnosis of thyroid cancer. With the increasing computational power and omics data, machine learning methods can identify disease-related molecules on a large scale to reveal the pathogenesis (Zhao et al., 2020c), disease occurrence process (Zhao et al., 2020a) and clinical medication guidance (Tianyi et al., 2020). Most of the calculation methods are based on similarity and interaction (Zhao et al., 2020b, 2021a). In this article, we propose hypotheses: there is a stronger interaction between the core genes that cause thyroid cancer. There is a close relationship between the pathogenic genes of thyroid cancer, but the interaction between the genes only related to thyroid cancer and these genes is not that close. Based on this hypotheses, we constructed a gene interaction network and used Random Walk (RW) to traverse this network. Then, Adaboost and deep neural network (DNN) was fused to identify TC-related genes.

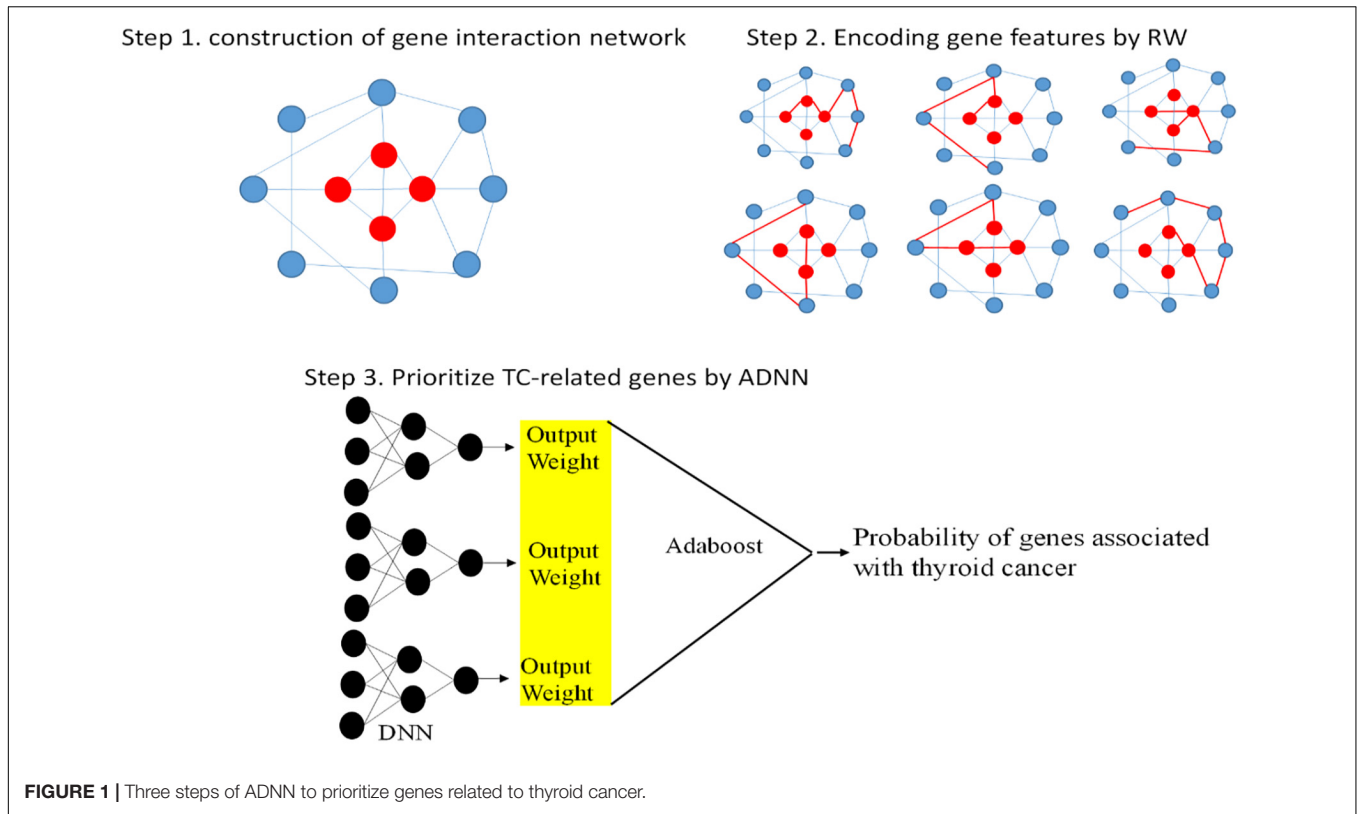
MATERIALS AND METHODS

There are three steps to implement ADNN. First, TC-related genes are obtained from DisGeNET (Piñero et al., 2015). Then, we collected genes which can interaction with TC-related genes to construct gene interaction network. The red points represent TC-related genes and blue points represent other genes. The second step is to use RW to traverse this network. The features of genes can be encoded by this step. The last step is to fuse Adaboost with DNN to prioritize TC-related genes. The whole process of ADNN is shown as **Figure 1**.

Construction of Gene Interaction Network

First, we obtained TC-related genes from DisGeNET. According to DisGeNET, there are 147 genes related to TC. Using String database (Szklarczyk et al., 2016), we draw these genes interaction network as **Figure 2**.

As we can see in **Figure 2**, the interaction between some core genes in the center of the network is very close. Although there are still some marginal genes that cannot interact with other



genes, most of the genes have close relationship with other genes. We collected genes which can interaction with 147 TC-related genes to construct a whole gene interaction network.

Encoding Gene Features by RW

Since the gene interaction network we construct is a two-dimensional graph, when we walk through the network in a probabilistic manner based on gene interaction, when the node information of the current gene is known, the historical gene node traversal information and the future gene node The traversal path is irrelevant. Therefore, we can regard the genetic coding method based on random walk as a Markov chain. In each step of the Markov chain, the probability distribution of gene interaction can change from one state to another, or maintain the current state. The change of state is called transition, and the probability associated with different states is called transition probability.

If A is the adjacency matrix of the gene interaction network, we can normalize A as:

$$P = D^{-1}A \tag{1}$$

D is a diagonal matrix and the degree matrix of the gene interaction network:

$$D(i, i) = \sum A(i, j) \tag{2}$$

P is a random walk matrix, the transition probability of each node is 1, and P is the probability matrix associated with TC and all genes.

A random walk matrix corresponds to a Markov chain, and the probability distribution of TC-related genes changes as the state in the Markov chain changes. Starting from any state, the probability of going to the next state is as follows:

$$P_{t+1} = A_t P \tag{3}$$

This process continues, and the relationship between TC and genes is constantly changing. After a period of time, it reaches a state of equilibrium. The equilibrium state is also called steady state, which means that the probability distribution of the association between TC and genes no longer changes. The calculation method of steady state is as follows:

$$\pi = D(i, j) / \sum_i \sum_j A(i, j) \tag{4}$$

When $\pi P = \pi$, the entire system reaches a steady state. This steady state is the final calculated association between TC and gene.

Prioritize TC-Related Genes by ADNN

DNN neural network layers can also be simply divided into three categories: input layer, hidden layer and output layer. Its layers are fully connected, that is, all neurons in the upper layer are connected to any neuron in the next layer. Its partial model is:

$$O = \sigma \left(\sum wx + b \right) \tag{5}$$

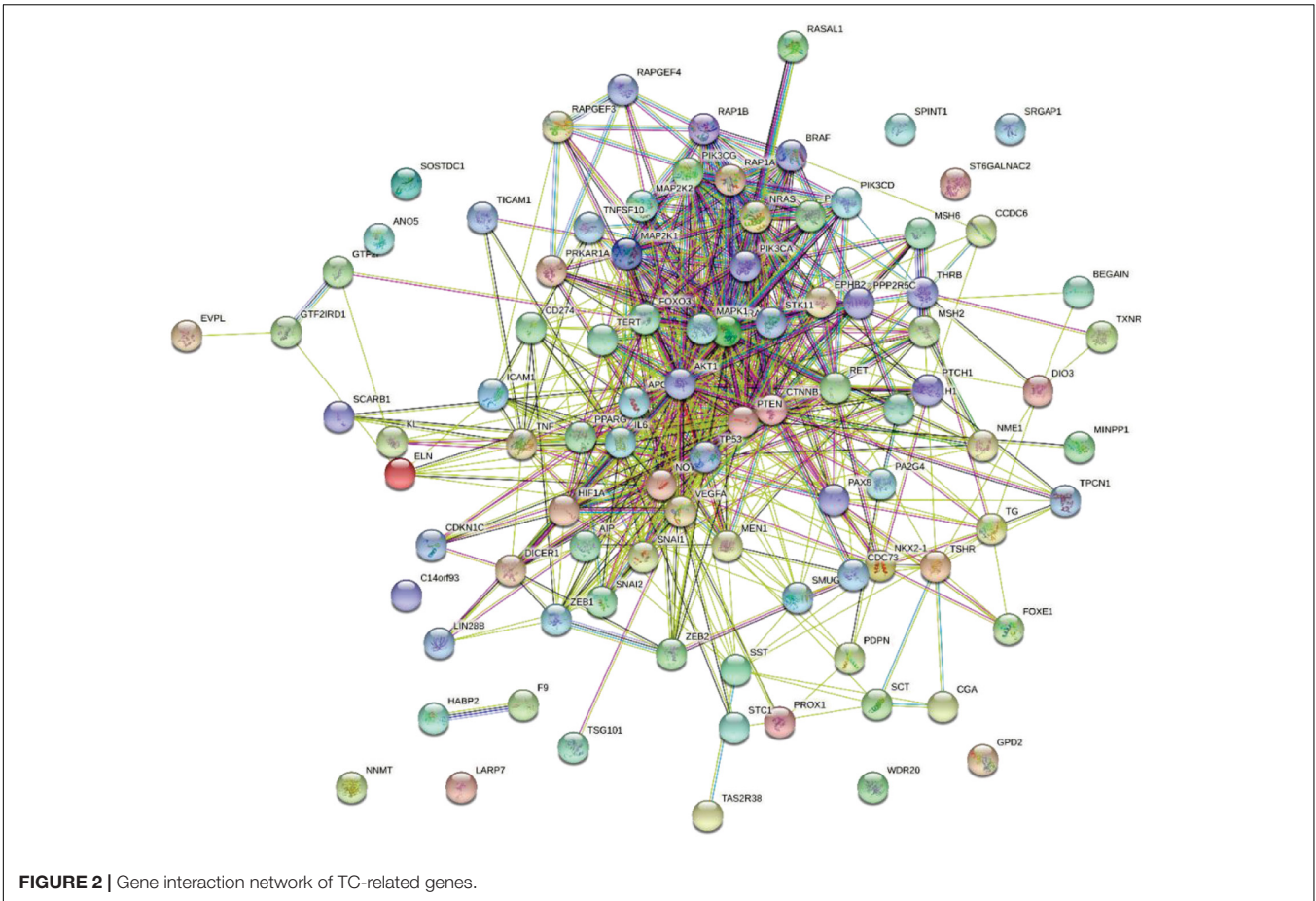


FIGURE 2 | Gene interaction network of TC-related genes.

O is output. $\sigma()$ is activation function. w is the coefficient of linear relationship, b is bias model parameters.

Using DNN network architecture to identify the interaction pattern between TC and gene, we need to define the objective function to measure the loss of model fitting.

$$J(w, b, x, y) = \frac{1}{2} \| a^L - y \|^2 \quad (6)$$

The process of training DNN is to minimize the loss function. The parameters of DNN model is shown in **Table 1**.

Due to the small sample set, DNN is used as a weak classifier. In order to make the model more accurate, we introduced AdaBoost.

First, set the initial weight of each sample to $1/N$. Then, training samples to get the first DNN model, test this DNN model, increase the weight of the unclassified correct samples and reduce the weight of the classified correct samples. At the same time, the weight of the DNN model is obtained. Repeating the above process, we can get multiple DNN models and corresponding weights, thereby obtaining the final strong classifier.

The error rate of each model can be calculated as following:

$$err_m = \sum_{i=1}^N w_{mi} |(G_m(x_i) \neq y_i)| \quad (7)$$

The weight of the model is:

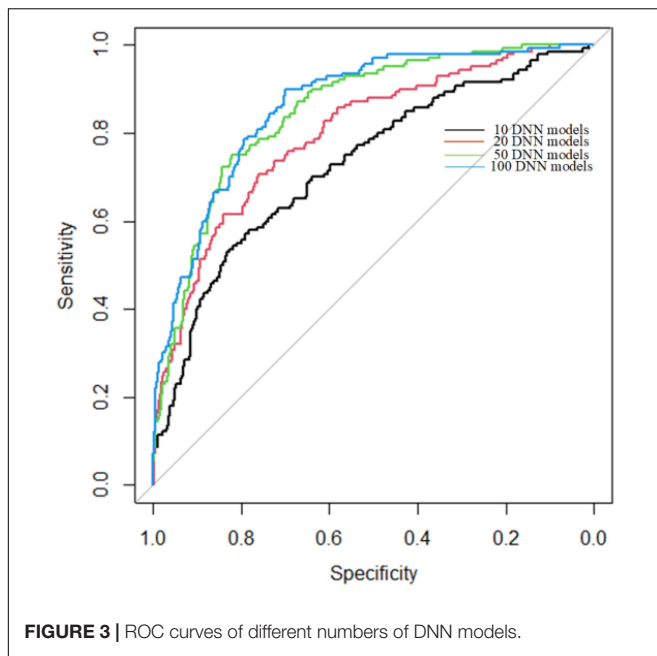
$$a_m = \frac{1}{2} \log \frac{1 - err_m}{err_m} \quad (8)$$

The final model is the summary of all DNN models:

$$G(x) = \arg \max_{m: G_m(x)=y} \sum a_m \quad (9)$$

TABLE 1 | The parameters of DNN model.

Structure	Parameters
Layer 1	Units: 256 Activation function: Tanh Dropout rate: 0.3
Layer 2	Units: 128 Activation function: Tanh Dropout rate: 0.2
Layer 3	Units: 2 Activation function: Tanh
Loss function	Binary cross entropy
Optimizer	RMSprop



RESULTS

Since we used DNN as a weak classifier and the number of DNN models is set by experience, we used 5-cross validation to find the best number of DNN models of ADNN. The process of 5-cross validation is to divide whole sample set into five groups. We used one group for testing and four groups for training each time. After repeating five times, each group has been tested once and trained four times. We use 10, 20, 50, and 100 DNN models to build ADNN, respectively. The experiment results are shown as **Figure 3**.

In **Figure 3**, the AUC are 0.73, 0.79, 0.85, 0.86 for 10, 20, 50, 100 DNN models, respectively. As shown in **Figure 3**, the AUC of 50 and 100 DNN models are very close. However, constructing 100 DNN models is twice time consuming than 50 DNN models. Therefore, we used 50 DNN models to build ADNN model.

In order to show the superiority of ADNN, we compared it with several other methods such as DNN, ASVM, ANB, Random Forest (RF). ASVM is the fusion of Adaboost and Support Vector Machine (SVM). ANB is the fusion of Adaboost and Naïve Bayes (NB).

The comparison results is listed in **Table 2**.

Compared ADNN with ASVM, we can find that DNN is more suitable than SVM in prioritizing susceptible genes for thyroid

TABLE 2 | Comparison of ADNN and other methods.

Method	AUC	AUPR
ADNN	0.85	0.81
DNN	0.69	0.65
ASVM	0.82	0.79
ANB	0.76	0.71
RF	0.78	0.76

cancer. Compared ADNN with DNN, we can find that Adaboost can significantly increase the accuracy of prioritizing susceptible genes for thyroid cancer.

CONCLUSION

Genetic factors are an important cause of thyroid cancer. Exploring the susceptibility genes of thyroid cancer is the key to understanding the pathogenesis and developing new treatment options. Collecting samples from patients and healthy individuals and analyzing differential gene expression is very costly and time-consuming. After years of research, researchers have found only 147 genes related to thyroid cancer. The role of these genes in thyroid cancer is unknown. In addition, there are more genes associated with thyroid cancer. To prioritize susceptible genes of thyroid cancer in large-scale, we proposed a novel method, named ADNN, to identify TC-related genes by gene interaction network. We constructed gene interaction network based on known TC-related genes and used RW to encode the features of genes. Then, we fused Adaboost with DNN to classify whether a gene is related to TC and obtain the probability of genes associated with TC. We get the best number of DNN models needed to construct ADNN through experiments. Finally, we compared ADNN with several other methods. Overall, we propose a precise and efficient method for prioritizing susceptible genes for thyroid cancer.

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.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

L-KZ, C-LX, SJ, J-HM, and BX participated in the study design. L-KZ, C-LX, SJ, X-XG, J-HF, W-SC, C-ZL, and FS analyzed the data. L-KZ, C-LX, SJ, and X-YD wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

Financial support came from the Zhongshan Medical Scientific Research Project (2021A020606) and the Guangzhou Science and Technology Plan Project (202102080170).

REFERENCES

- Agrawal, N., Akbani, R., Aksoy, B. A., Ally, A., Arachchi, H., Asa, S. L., et al. (2014). Integrated genomic characterization of papillary thyroid carcinoma. *Cell* 159, 676–690. doi: 10.1016/j.cell.2014.09.050
- Ceolin, L., Duval, M., Benini, A. F., Ferreira, C. V., and Maia, A. L. (2019). Medullary thyroid carcinoma beyond surgery: advances, challenges, and perspectives. *Endocr. Relat. Cancer* 26, R499–R518. doi: 10.1530/erc-18-0574
- Chen, L., Zhao, K., Li, F., and He, X. (2020). Medullary thyroid carcinoma with elevated serum CEA and normal serum calcitonin after surgery: a case report and literature review. *Front. Oncol.* 10:526716. doi: 10.3389/fonc.2020.526716
- Dralle, H., Machens, A., Basa, J., Fatourechi, V., Franceschi, S., Hay, I. D., et al. (2015). Follicular cell-derived thyroid cancer. *Nat. Rev. Dis. Primers* 1:15077. doi: 10.1038/nrdp.2015.77
- Lee, J. H., Ha, E. J., and Kim, J. H. (2019). Application of deep learning to the diagnosis of cervical lymph node metastasis from thyroid cancer with CT. *Eur. Radiol.* 29, 5452–5457. doi: 10.1007/s00330-019-06098-8
- Li, X., Zhang, S., Zhang, Q., Wei, X., Pan, Y., Zhao, J., et al. (2019). Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. *Lancet. Oncol.* 20, 193–201. doi: 10.1016/s1470-2045(18)30762-9
- Maniakas, A., Dadu, R., Busaidy, N. L., Wang, J. R., Ferrarotto, R., Lu, C., et al. (2020). Evaluation of overall survival in patients with anaplastic thyroid carcinoma, 2000–2019. *JAMA Oncol.* 6, 1397–1404. doi: 10.1001/jamaoncol.2020.3362
- Molinaro, E., Romei, C., Biagini, A., Sabini, E., Agate, L., Mazzeo, S., et al. (2017). Anaplastic thyroid carcinoma: from clinicopathology to genetics and advanced therapies. *Nat. Rev. Endocrinol.* 13, 644–660. doi: 10.1038/nrendo.2017.76
- Penna, G. C., Vaisman, F., Vaisman, M., Sobrinho-Simões, M., and Soares, P. (2016). Molecular markers involved in tumorigenesis of thyroid carcinoma: focus on aggressive histotypes. *Cytogenet. Genome. Res.* 150, 194–207. doi: 10.1159/000456576
- Petrulea, M. S., Plantinga, T. S., Smit, J. W., Georgescu, C. E., and Netea-Maier, R. T. (2015). PI3K/Akt/mTOR: a promising therapeutic target for non-medullary thyroid carcinoma. *Cancer Treat. Rev.* 41, 707–713. doi: 10.1016/j.ctrv.2015.06.005
- Piñero, J., Queralt-Rosinach, N., Bravo, A., Deu-Pons, J., Bauer-Mehren, A., Baron, M., et al. (2015). DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database* 2015:bav028. doi: 10.1093/database/bav028
- Sahu, R., and Pattanayak, S. P. (2020). Strategic developments & future perspective on Gene therapy for breast cancer: role of mTOR and Brk/PTK6 as molecular targets. *Curr. Gene. Ther.* 20, 237–258. doi: 10.2174/1566523220999200731002408
- Salvatore, G., De Falco, V., Salerno, P., Nappi, T. C., Pepe, S., Troncone, G., et al. (2006). BRAF is a therapeutic target in aggressive thyroid carcinoma. *Clin. Cancer Res.* 12, 1623–1629. doi: 10.1158/1078-0432.ccr-05-2378
- Sanguedolce, F., Cormio, A., Bufo, P., Carrieri, G., and Cormio, L. (2015). Molecular markers in bladder cancer: Novel research frontiers. *Crit. Rev. Clin. Lab. Sci.* 52, 242–255. doi: 10.3109/10408363.2015.1033610
- Sui, T., Ma, L., Bai, X., Li, Q., and Xu, X. (2014). Resveratrol inhibits the phosphatidylinositol 3-kinase/protein kinase B/mammalian target of rapamycin signaling pathway in the human chronic myeloid leukemia K562 cell line. *Oncol. Lett.* 7, 2093–2098. doi: 10.3892/ol.2014.2014
- Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2016). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 45, D362–D368. doi: 10.1093/nar/gkw937
- Tianyi, Z., Yang, H., Valsdottir, L. R., Tianyi, Z., and Jiajie, P. (2020). Identifying drug–target interactions based on graph convolutional network and deep neural network. *Brief. Bioinform.* 2020:bbaa044. doi: 10.1093/bib/bbaa044
- Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp. Oncol.* 19, A68–A77. doi: 10.5114/wo.2014.47136
- Trovisco, V., Vieira de Castro, I., Soares, P., Máximo, V., Silva, P., Magalhaes, J., et al. (2004). BRAF mutations are associated with some histological types of papillary thyroid carcinoma. *J. Pathol.* 202, 247–251. doi: 10.1002/path.1511
- Tsukatan, T., Niwa, H., Komori, T., Yoneyama, T., Tsuji, H., Michigishi, T., et al. (2020). Superior mediastinal lymphadenopathy by silicosis mimicking metastasis of papillary thyroid carcinoma—Case report and literature review. *Auris Nasus Larynx* 47, 1054–1057. doi: 10.1016/j.anl.2019.12.003
- Vijayan, R., Nair, V., Menon, U., and Kumar, H. (2021). A rare RET mutation in an Indian pedigree with familial medullary thyroid carcinoma. *Indian J. Cancer* 58, 98–100. doi: 10.4103/ijc.IJC_639_19
- Xia, Y., Li, X. Q., and Sun, W. (2020). Applications of recombinant adenovirus-p53 gene therapy for cancers in the clinic in china. *Curr. Gene. Ther.* 20, 127–141. doi: 10.2174/1566523220999200731003206
- Xing, M. (2019). Genetic-guided risk assessment and management of thyroid cancer. *Endocrinol. Metab. Clin.* 48, 109–124. doi: 10.1016/j.ecl.2018.11.007
- Zanella, A. B., Scheffel, R. S., Weinert, L., Dora, J. M., and Maia, A. L. (2021). New insights into the management of differentiated thyroid carcinoma in children and adolescents. *Int. J. Oncol.* 58:13. doi: 10.3892/ijo.2021.5193
- Zhang, S., He, X., Zhang, R., and Deng, W. (2021). LncR2metasta: a manually curated database for experimentally supported lncRNAs during various cancer metastatic events. *Brief. Bioinform.* 22:bbaa178. doi: 10.1093/bib/bbaa178
- Zhao, T., Hu, Y., and Cheng, L. (2020a). Deep-DRM: a computational method for identifying disease-related metabolites based on graph deep learning approaches. *Brief. Bioinform.* 2020:bbaa212. doi: 10.1093/bib/bbaa212
- Zhao, T., Hu, Y., Peng, J., and Cheng, L. (2020b). DeepLGP: a novel deep learning method for prioritizing lncRNA target genes. *Bioinformatics* 36, 4466–4472. doi: 10.1093/bioinformatics/btaa428
- Zhao, T., Hu, Y., Zang, T., and Cheng, L. (2020c). MRTFB regulates the expression of NOMO1 in colon. *Proc. Nat. Acad. Sci.* 117, 7568–7569. doi: 10.1073/pnas.2000499117
- Zhao, T., Liu, J., Zeng, X., Wang, W., Li, S., Zang, T., et al. (2021a). Prediction and collection of protein–metabolite interactions. *Brief. Bioinform.* 2021:bbab014. doi: 10.1093/bib/bbab014
- Zhao, T., Lyu, S., Lu, G., Juan, L., Zeng, X., Wei, Z., et al. (2021b). SC2disease: a manually curated database of single-cell transcriptome for human diseases. *Nucleic Acids Res.* 49, D1413–D1419. doi: 10.1093/nar/gkaa838

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.

Copyright © 2021 Zhong, Xie, Jiang, Deng, Gan, Feng, Cai, Liu, Shen, Miao and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). 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.