Original Article

Development and validation of a novel 14-gene signature for predicting lymph node metastasis in papillary thyroid carcinoma

Yuwei Ling^, Luyao Jia, Kaifu Li, Lina Zhang, Yajun Wang, Hua Kang

Center for Thyroid and Breast Surgery, Department of General Surgery, Xuanwu Hospital, Capital Medical University, Beijing, China

Contributions: (I) Conception and design: H Kang; (II) Administrative support: H Kang; (III) Provision of study materials or patients: K Li, Y Wang, H Kang; (IV) Collection and assembly of data: Y Ling, L Jia; (V) Data analysis and interpretation: Y Ling; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Hua Kang. Center for Thyroid and Breast Surgery, Department of General Surgery, Xuanwu Hospital, Capital Medical University, Beijing 100053, China. Email: kanghua@xwh.ccmu.edu.cn.

Background: There is still no reasonably accurate method of preoperatively predicting central lymph node metastasis (LNM), and it is essential to develop an effective evaluation model for predicting LNM in papillary thyroid carcinoma (PTC) patients.

Methods: PTC samples were collected from The Cancer Genome Atlas database. Candidate genes were identified as continuously upregulated or downregulated genes in the process of N0 to N1a and N1a to N1b. The least absolute shrinkage and selection operator (LASSO) regression analysis was used to construct the predictive model for LNM. Multivariate logistic regression analysis was performed to screen the potential factors related to LNM, and a nomogram was established. The risk score of the gene signature model for predicting disease-free survival (DFS) was evaluated by Kaplan-Meier analysis.

Results: A 14-gene signature was developed by LASSO regression for predicting LNM based on 69 differential expression genes (DEGs) that were continuously upregulated or downregulated in the progress of PTC. The receiver operating characteristic (ROC) curves of the 14-gene signature predicting LNM, central LNM and lateral LNM were generated. The area under the ROC (AUC) values were 0.806 [95% confidence interval (CI): 0.7608–0.8815], 0.755 (95% CI: 0.6839–0.8263) and 0.821 (95% CI: 0.7608–0.8815). The nomogram's C-index value, including the 14-gene signature and other potential risk factors, was 0.786 (95% CI: 0.7296–0.8425), and the calibration exhibited fairly good consistency with the perfect prediction. Based on the 14-gene risk score, high-risk PTC patients had a worse DFS.

Conclusions: A novel 14-gene signature was developed for predicting LNM in PTC patients. The risk score also correlated with DFS in PTC patients.

Keywords: Disease-free survival (DFS); lymph node metastasis (LNM); least absolute shrinkage and selection operator regression (LASSO regression); nomogram; papillary thyroid carcinoma (PTC)

Submitted May 31, 2021. Accepted for publication Aug 02, 2021.

doi: 10.21037/gs-21-361

View this article at: https://dx.doi.org/10.21037/gs-21-361

^ ORCID: 0000-0001-8487-3738.

Introduction

The global age-standardized incidence rate of thyroid cancer increased by 20% from 1990 to 2013 (1), but with increased attention to screening and management, the mortality rate is steady or declining (2). Approximately 95% of thyroid malignancies are known as differentiated thyroid carcinomas (DTCs), including papillary thyroid carcinomas (PTCs) and follicular thyroid carcinomas (FTCs) with a favorable 5-year overall survival (3). However, lymph node metastasis (LNM) accounts for 20–90% of DTC patients, especially in PTC (4,5). Despite this, the 2015 American Thyroid Association management guidelines for DTC did not recommend routine prophylactic central lymph nodes dissection (CLND) for T1 or T2, noninvasive, and clinically node-negative (cN0) patients (6). Unfortunately, it is not easy to clinically evaluate lymph node status in thyroid carcinoma, especially those in the central compartment. Preoperative ultrasound examination has fairly poor sensitivity in assessing LNM in the central compartment (7). In addition, although CT has a significant advantage in assessing deeper lymph nodes, it cannot evaluate lymph node micro-metastases or lymph nodes with a maximum diameter <5 mm. Therefore it is necessary to develop an effective preoperative evaluation model to avoid metastatic lymph nodes being missed because they will eventually lead to recurrence and reoperation (8). The higher incidence of surgical complications during reoperation significantly affects the patient's quality of life (9).

With high-throughput sequencing and bioinformatics technology development, potential biomarkers of LNM in PTC patients have been identified (10-14). However, the optimal biomarkers still need to be identified. It is generally believed that lateral LNM indicates a more advanced stage of PTC. A previously published study even proposed that LNM occurs in a stepwise fashion in PTC patients (15). The lymph nodes in the central compartment are firstly involved, followed by the ipsilateral lateral lymph nodes, and finally the contralateral lateral lymph nodes and mediastinal compartment. Accordingly, in our present study, we aimed to screen differential expression genes (DEGs) with the same variation trend from stage N0 to N1a and from stage N1a to N1b to discover the hub genes associated with the progression of LNM in PTC patients. We present the following article in accordance with the TRIPOD reporting checklist (available at https://dx.doi.org/10.21037/gs-21-361).

Methods

Datasets

The transcriptome data (FPKM) of 568 thyroid carcinoma samples in The Cancer Genome Atlas (TCGA) database were collected. The patients' clinical characteristics and survival data of all the PTC samples were obtained from UCSC Xena (https://xena.ucsc.edu; University of California, Santa Cruz). Clinical characteristics data including age at initial pathologic diagnosis, sex, number of lymph nodes examined, primary neoplasm focus type, primary thyroid gland neoplasm location anatomic site, pathologic TNM stage (which is defined following the AJCC 7th edition), radiation therapy status, and disease-free survival (DFS). The clinical data were re-evaluated according to the original pathologic reports. Cases of unknown lymph node status and non-PTC samples were excluded. Samples were divided into a training set (70%) and an internal validation set (30%) according to the status of cervical LNM. R software (version 4.0.3) was used for data collection and processing. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

DEGs analysis

To identify the DEGs associated with LNM in PTC patients, R package "limma" (version 3.46.0) was used to obtain the DEGs among the N0, N1a, and N1b PTC samples. A false discovery rate (FDR) <0.05 was the cut-off criterion for DEGs. Finally, the candidate genes were identified as continuously upregulated or downregulated genes in the process of N0 to N1a and N1a to N1b.

Least absolute shrinkage and selection operator (LASSO) regression analysis

To discover the potential biomarkers for LNM in PTC patients, the LASSO regression analysis was conducted because of the multicollinearity of DEGs. 10-fold cross-validation was applied in the LASSO regression analysis to determine the optimal penalty parameter (λ), and dimension reduction was performed on the DEGs to reduce interference or redundant genes to select primary predictive factors to build a relatively refined gene signature. The R package "glmnet" (version 4.1-1) was used to perform the LASSO regression analysis. The risk score of the gene signature model was calculated according

to the corresponding coefficients of each gene by the R package "rms" (version 6.0-1). The risk score = $\sum$ [β(i) × Exp(i)], where Exp(i) and β(i) are the relative abundances and the LASSO regression coefficient of the feature in the established gene signature. The receiver operating characteristic (ROC) curves for predicting LNM, central LNM and lateral LNM by the gene signature risk score was generated by the R package "pROC" (version 1.17.0.1), and the area under the ROC curve (AUC) was calculated. Eventually, the prediction value of the gene signature was further verified in the internal validation set.

Multivariate analysis and nomogram construction

In order to discover the potential indicators of LNM in PTC, multivariate logistic regression analysis was performed using the R Stats package, including potential clinical features and the risk score calculated based on the gene signature. The R package "rms" (version 6.1-0) was applied to construct the nomogram for predicting LNM. The length of the line corresponding to each factor in the nomogram reflects the contribution of each factor to LNM in PTC. The risk score was calculated by the R package "nomogramFormula" (version 1.2.0.0). The prediction value of LNM by the nomogram was examined by drawing the calibration curves. This scoring system's prediction and calibration performance was evaluated using the Hosmer-Lemeshow goodness-of-fit test using the R Package "ResourceSelection" (version 0.3-5).

Survival analysis

Kaplan-Meier survival curves were generated to explore the predictive value of the gene signature in DFS of PTC patients. The patients were divided into high-risk and low-risk groups according to the optimal cut-off of the gene signature risk score automatically calculated by the R package "survminer" (version 0.4.8). All possible cut-off values between the lower and upper quartiles were computed, and the best performing threshold was used as the cut-off value.

Statistical analysis

The student's *t*-test estimated the statistical significance of continuous variables. LASSO regression analysis determined the candidate genes for predicting LNM of PTC patients. After the gene signature model was established, multivariate logistic regression analysis was conducted using R software to explore the value of the risk score based on gene signature and other clinical features for predicting LNM. The ROC curves were generated based on the R package "pROC" (version 1.17.0.1) to verify the model's validity. The highest sum sensitivity + specificity threshold is calculated by the R package "pROC" and plotted in the ROC curve. The log-rank test determined the significant differences of the Kaplan-Meier survival curves. All statistical analyses were performed by R software 4.0.3. A P value of <0.05 was considered statistically significant.

Results

Clinical characteristics of PTC patients in TCGA Database

In total, there were 443 samples with data on the N stage, comprising 226 samples in stage N0 (51.01%), 87 samples in stage N1a (19.64%), 73 samples in stage N1b (16.48%), and 57 samples without further stratification as N1a or N1b (12.87%). According to the lymph node status, the 443 samples were divided into a training set (N=311) and an internal validation set (N=132). The baseline clinical characteristics are presented in *Table 1*. There was no significant difference in the status of LNM between the training and validation sets.

Identification of DEGs associated with LNM in PTC patients

We performed a stepwise analysis to identify the candidate DEGs associated with central and lateral LNM in PTC patients. A total of 7,833 DEGs significantly altered in N1a versus N0 samples were identified, consisting of 3,853 upregulated genes and 3,980 downregulated genes (*Figure 1A*). A total of 770 DEGs differentially expressed in N1b versus N1a samples were identified, comprising 342 upregulated genes and 428 downregulated genes in N1b samples (*Figure 1B*). Eventually, 50 continuously downregulated DEGs and 19 continuously upregulated DEGs in the process of stage N0 to N1a and N1a to N1b were selected as candidate genes associated with LNM (Table S1).

Development and validation of gene signature for predicting LNM in PTC patients

In order to screen the potential candidate genes for

Table 1 Clinical characteristics of papillary thyroid carcinoma patients in TCGA database [n (%)]

Variables	Total (N=443)	Training set (N=311)	Validation set (N=132)	Statistics	P value
Age*	47 [35, 58]	47 [36, 58]	46 [33, 58]	–0.694	0.487
Sex				0.131	0.718
Female	324 (73.1)	229 (73.6)	95 (72.0)		
Male	119 (26.9)	82 (26.4)	37 (28.0)		
Number of lymph node examined*	5 [2, 16]	5 [2, 15]	7 [3, 19.5]	–1.142	0.254
T stage				2.609	0.625
T1	131 (29.6)	98 (31.5)	33 (25.0)		
T2	139 (31.4)	94 (30.2)	45 (34.1)		
T3	150 (33.8)	102 (32.8)	48 (36.4)		
T4	22 (5.0)	16 (5.2)	6 (4.5)		
TX	1 (0.2)	1 (0.3)	0		
N stage				1.261	0.738
N0	226 (51.0)	159 (51.1)	67 (50.8)		
N1	57 (12.9)	37 (11.9)	20 (15.1)		
N1a	87 (19.6)	61 (19.6)	26 (19.7)		
N1b	73 (16.5)	54 (17.4)	19 (14.4)		
Multifocality				0.125	0.940
Unifocal	233 (52.6)	165 (53.1)	68 (51.5)		
Multifocal	201 (45.4)	140 (45.0)	61 (46.2)		
Unknown	9 (2.0)	6 (1.9)	3 (2.3)		
Tumor side				7.069	0.029
Unilateral	357 (80.6)	249 (80.1)	108 (81.8)		
Bilateral	81 (18.3)	61 (19.6)	20 (15.2)		
Unknown	5 (1.1)	1 (0.3)	4 (3.0)		
Radiation therapy				3.449	0.178
No	160 (36.1)	115 (37.0)	45 (34.1)		
Yes	266 (60.1)	181 (58.2)	85 (64.4)		
Unknown	17 (3.8)	15 (4.8)	2 (1.5)		

*, age and number of lymph node examined are abnormally distributed continuous variables and represented by the median and upper and lower quartiles. The statistical significance was estimated by Mann-Whitney U test.

predicting LNM in the training set, LASSO regression analysis was performed because of the multicollinearity among the 69 DEGs (*Figure 1C*). Finally a 14-gene signature model was constructed (*FAM240C*, *C12orf60*, *ZNF79*, *INKA2*, *ZNF544*, *KIAA0319L*, *ZNF618*, *APMAP*, *ATP6V1B2*, *BRIX1*, *DNAJC21*, *BAZ1A*, *PI15*, *ZMYND8*; *Figure 1D*,*1E*). The risk scores based on the 14-gene signature were calculated. The expression pattern of the 14 candidate genes is shown in *Figure 2*; the risk score gradually increased with the severity of LNM in PTC

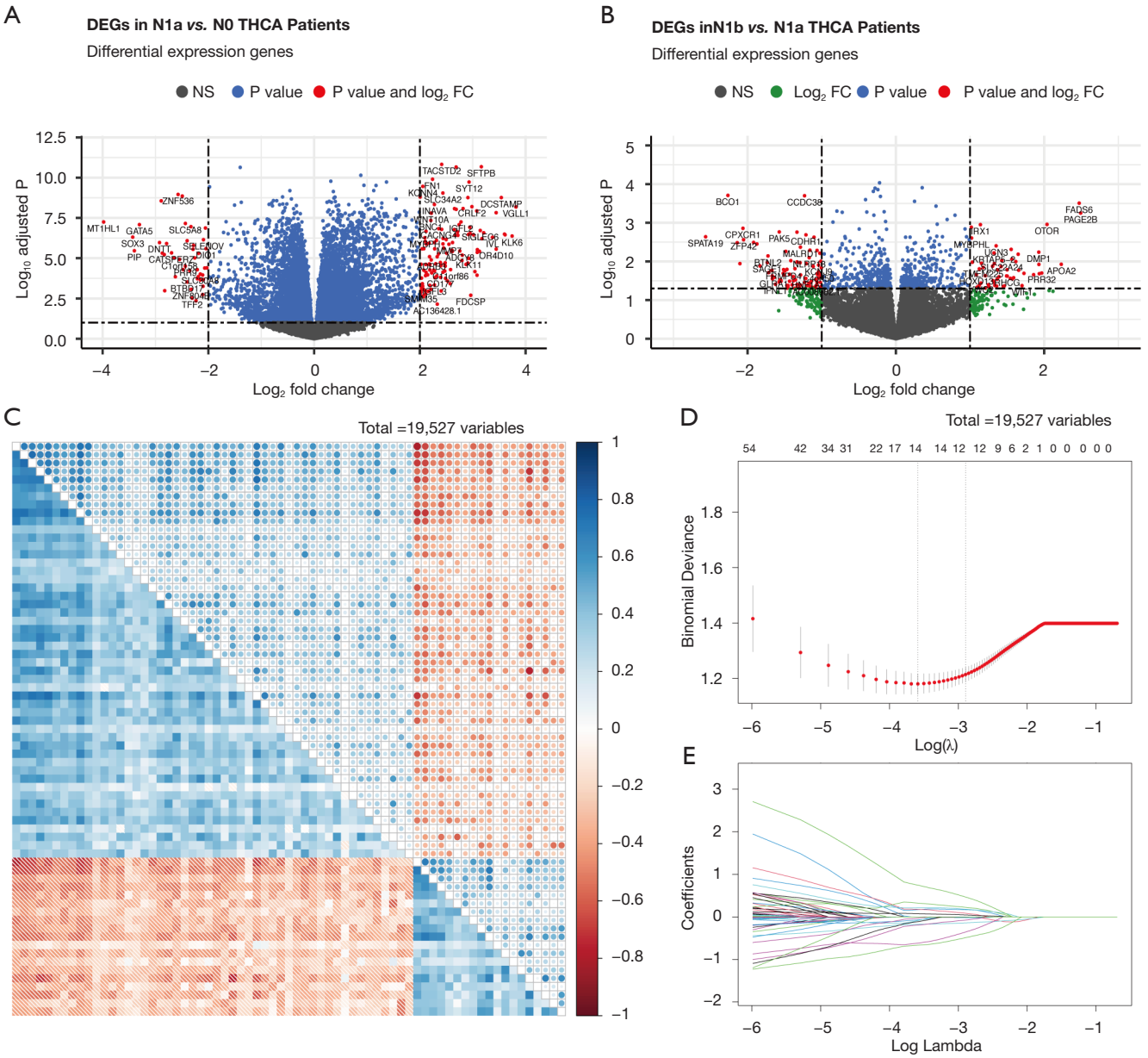

Figure 1 Identification of candidate genes associated with lymph node metastasis in papillary thyroid carcinoma (PTC). (A) Differential expression genes (DEGs) in PTC patients staged as N1a versus N0. (B) DEGs in PTC patients staged as N1b versus N1a. (C) Correlation coefficient matrix of 69 DEGs, which shows multicollinearity among them. (D,E) Determination of the 14-gene signature by LASSO regulation analysis.

patients. The ROC curves of the 14-gene signature predicting LNM, central LNM and lateral LNM were generated. The AUC values were 0.806 [95% confidence interval (CI): 0.7608–0.8815, *Figure 3A*], 0.755 (95% CI: 0.6839–0.8263, *Figure 3B*) and 0.821 (95% CI: 0.7608–0.8815, *Figure 3C*). The predictive value was verified in the internal validation set and the AUC reached 0.733 (95% CI: 0.6478–0.8181, *Figure 3D*), 0.661 (95% CI: 0.5441–0.7785, *Figure 3E*) and 0.786 (95% CI: 0.662–0.909, *Figure 3F*). These results illustrated that the 14-gene signature had a favorable predictive value, especially in predicting lateral LNM.

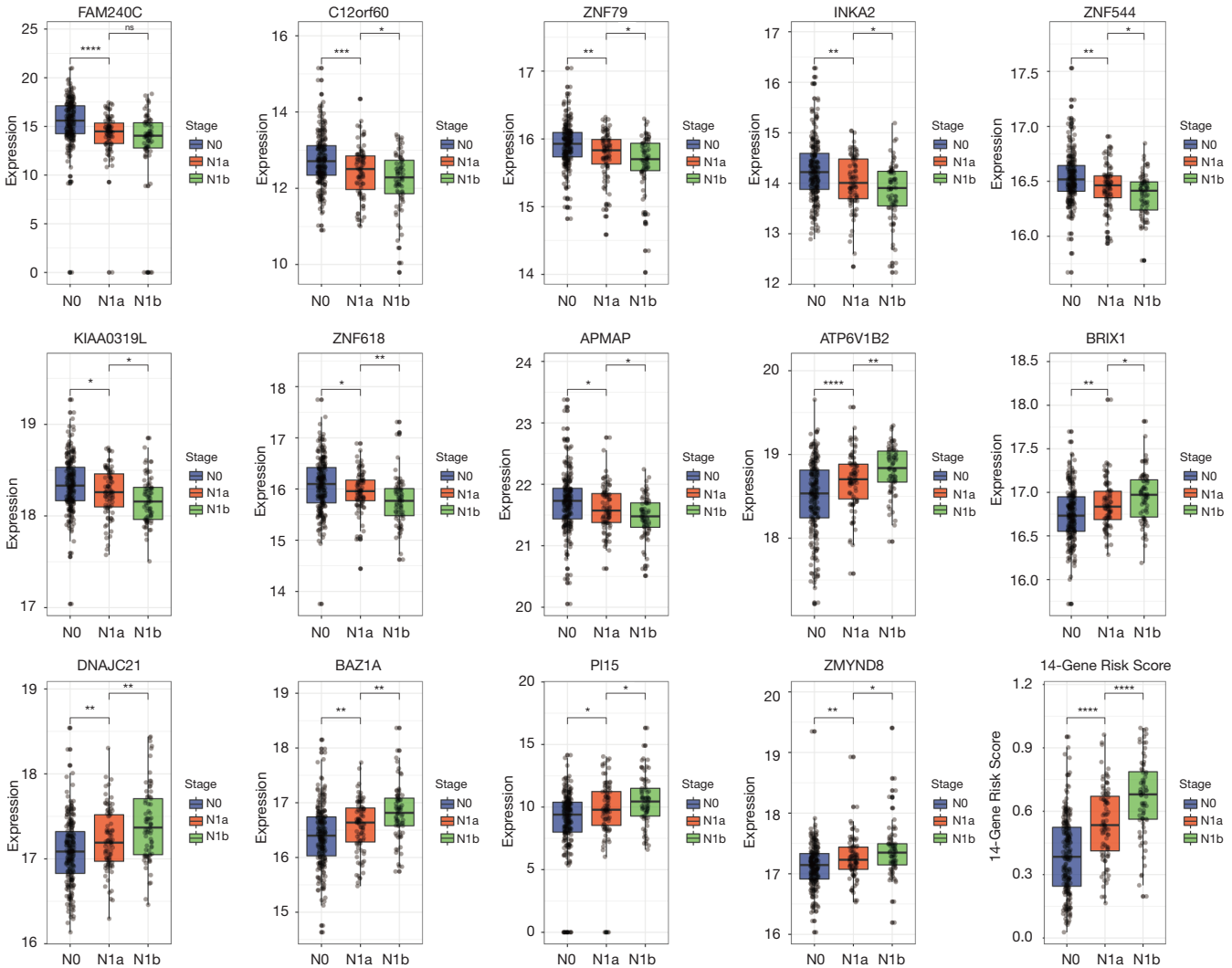

Figure 2 Expression patterns of 14 candidate genes in the gene signature model and the 14-gene risk score distribution in different N stages. *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.

Multivariate logistic regression analysis for LNM in PTC patients

In order to further explore the predictive value of the 14-gene signature, multivariate logistic regression analysis was conducted, including 14-gene signature and other potential risk factors of LNM in PTC patients. According to the optimal cut-off calculated by the ROC curve (0.559, *Figure 3A*), patients in the training set were divided into a low-risk group (14-gene signature risk score <0.559) and a high-risk group (14-gene signature risk score ≥0.559). The results showed that age [odds ratio (OR) =0.980, 95% CI: 0.962–0.997, P=0.026], T stage (T3-T4, OR =1.825, 95% CI: 1.034–3.228, P=0.038) and the 14-gene risk score (high risk, OR =8.150, 95% CI: 4.656–14.745, P<0.001) were potential predictors of LNM in PTC patients (*Table 2*).

Development and validation of a nomogram for predicting LNM

A nomogram for predicting LNM, including the 14-gene signature and other potential risk factors, was established (*Figure 4A*). To evaluate the predictive value of the nomogram for LNM in PTC patients, firstly, the calibration curve was generated by 1,000 times resample using the bootstrap method. The calibration

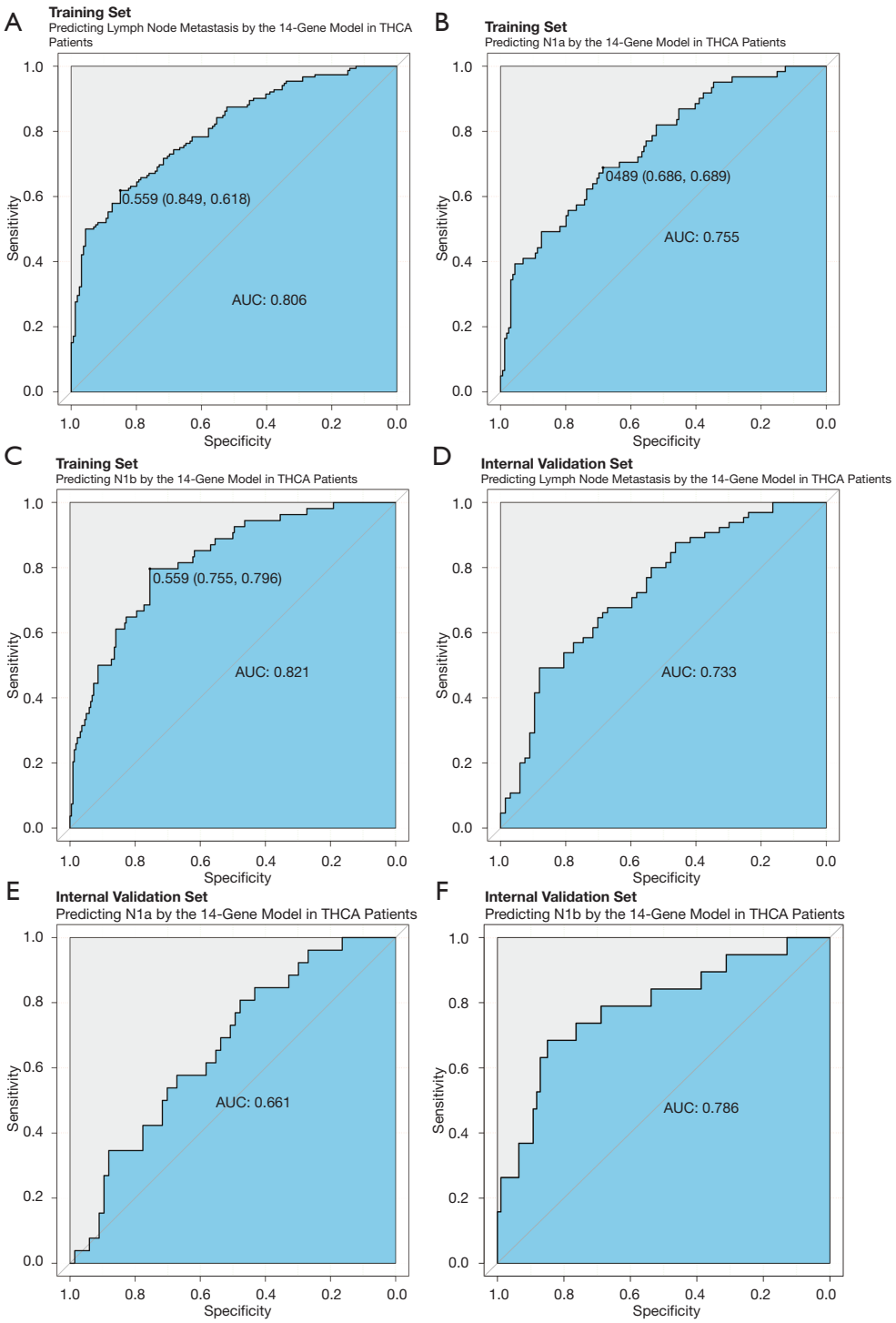

Figure 3 Predictive value of 14-gene signature for lymph node metastasis in papillary thyroid carcinoma (PTC) patients. (A) Receiver operating characteristic (ROC) curve of 14-gene signature for predicting lymph node metastasis in the training set. (B) ROC curve of 14-gene signature for predicting lymph node metastasis in the internal validation set. (C) ROC curve of 14-gene signature for predicting central lymph node metastasis in the training set. (D) ROC curve of 14-gene signature for predicting lateral lymph node metastasis in training set. (E) ROC curve of 14-gene signature for predicting central lymph node metastasis in internal validation set. (F) ROC curve of 14-gene signature for predicting lateral lymph node metastasis in internal validation set.

Table 2 Multivariate analysis for predicting lymph node metastasis in papillary thyroid carcinoma patients

Variables	B	SE	OR	95% CI		P value
				Lower	Upper	
Age	–2.233	0.452	0.980	0.962	0.997	0.026
Sex						
Female			1.000			
Male	1.192	0.313	1.452	0.787	2.693	0.233
T Stage						
T1-T2			1.000			
T3-T4	2.078	0.290	1.825	1.034	3.228	0.038
Multifocality						
Unifocal			1.000			
Multifocal	0.066	0.314	1.021	0.548	1.886	0.948
Tumor side						
Unilateral			1.000			
Bilateral	1.571	0.401	1.878	0.862	4.172	0.116
14-gene risk score						
Low risk			1.000			
High risk	7.156	0.293	8.150	4.656	14.745	<0.001

curve exhibited fairly good consistency with the perfect prediction (*Figure 4B*). The Hosmer-Lemeshow goodness-of-fit test showed good consistency between the true state of LNM and the predicted value based on the nomogram (Chi-square =4.8085, P=0.7778). The C-index value of the nomogram model was 0.786 (95% CI: 0.7296–0.8425). The risk scores were calculated for each sample, and the ROC curves for predicting LNM by the nomogram were generated (*Figure 4C*). The discrimination and calibration of the nomogram were further verified in the internal validation set (*Figure 4D*). The Hosmer-Lemeshow goodness-of-fit test also showed fairly good consistency in the validation set (Chi-Square =7.6795, P=0.4654). The risk scores were further calculated for each sample in the validation set, and the ROC curve for predicting LNM had an AUC value of 0.712 (95% CI: 0.6192–0.8057, *Figure 4E*).

Predictive value of the 14-gene signature for DFS

LNM is often blamed for local recurrence of thyroid cancer, so we performed survival analysis. The samples were divided into high-risk and low-risk groups based on the 14-gene signature. The Kaplan-Meier curve revealed that patients in the high-risk group had unfavorable DFS in both the training and internal validation set (*Figure 5*).

Discussion

Although in recent years there have been several attempts at developing an optimal method for clinically evaluating the lymph nodes status of PTC patients (16,17), no particularly accurate method has emerged to preoperatively predict central LNM, especially in cN0 patients (18). With high-throughput sequencing and bioinformatics technology development, several biomarkers of LNM have been identified in previously published literatures. Wang *et al.* (10) identified 752 upregulated and 309 downregulated DEGs in thyroid cancer compared to normal tissue. Zhang *et al.* (11) discovered that BCL2 and hsa-miR-181a-5p are potential biomarkers associated with PTC, based on GEO database analysis. Liu *et al.* (12) identified 358 DEGs related to thyroid carcinoma, including 135 upregulated and 224 downregulated genes, and eventually filtered out five hub

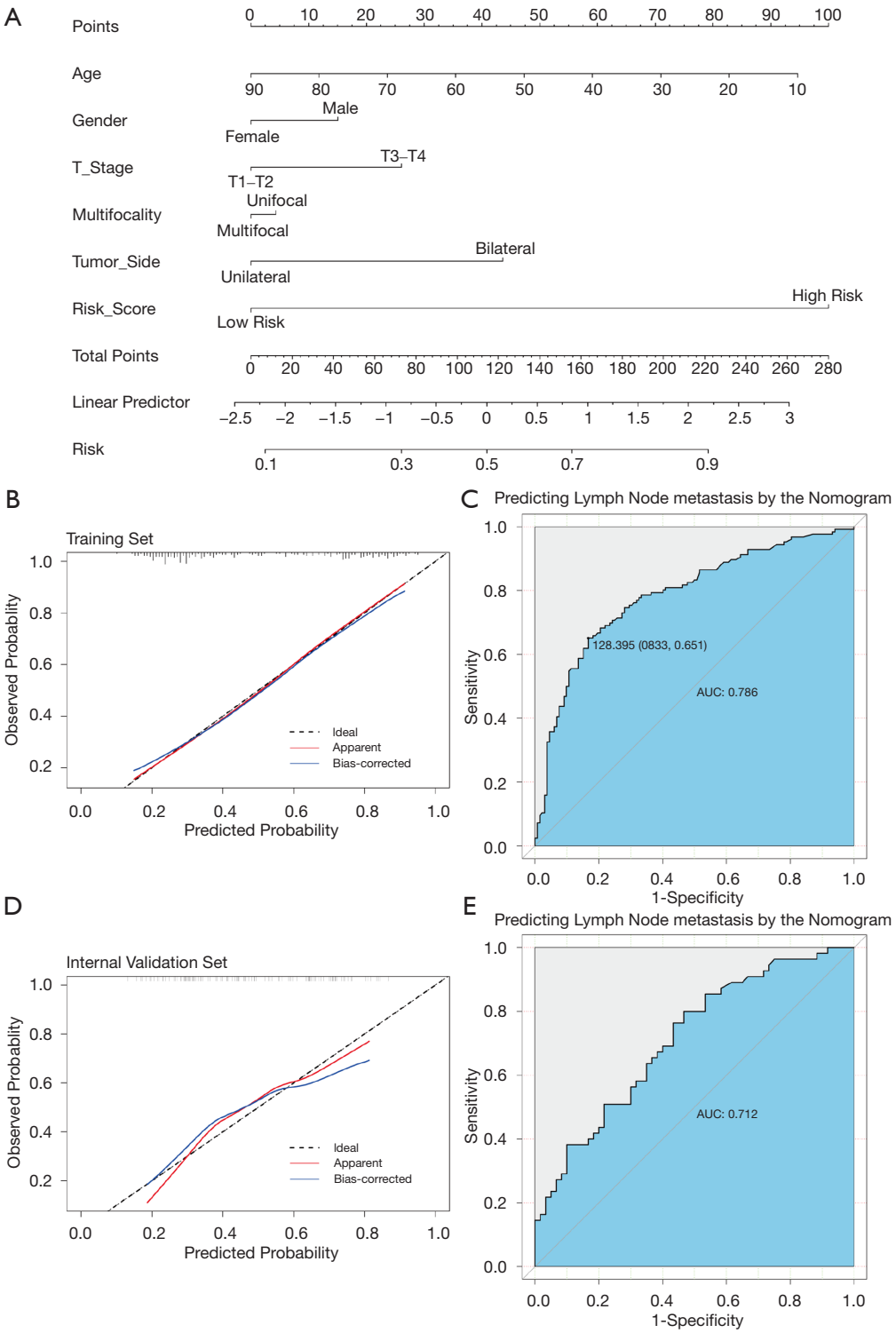

Figure 4 Establishment and validation of nomogram for predicting lymph node metastasis. (A) Nomogram for predicting lymph node metastasis in papillary thyroid carcinoma samples. (B) Calibration curve for the nomogram in the training set, which shows excellent goodness-of-fit. (C) Receiver operating characteristic (ROC) curve for predicting lymph node metastasis by the nomogram in the training set. (D) Calibration curve for nomogram in the internal validation set. (E) ROC curve for predicting lymph node metastasis by the nomogram in the internal validation set.

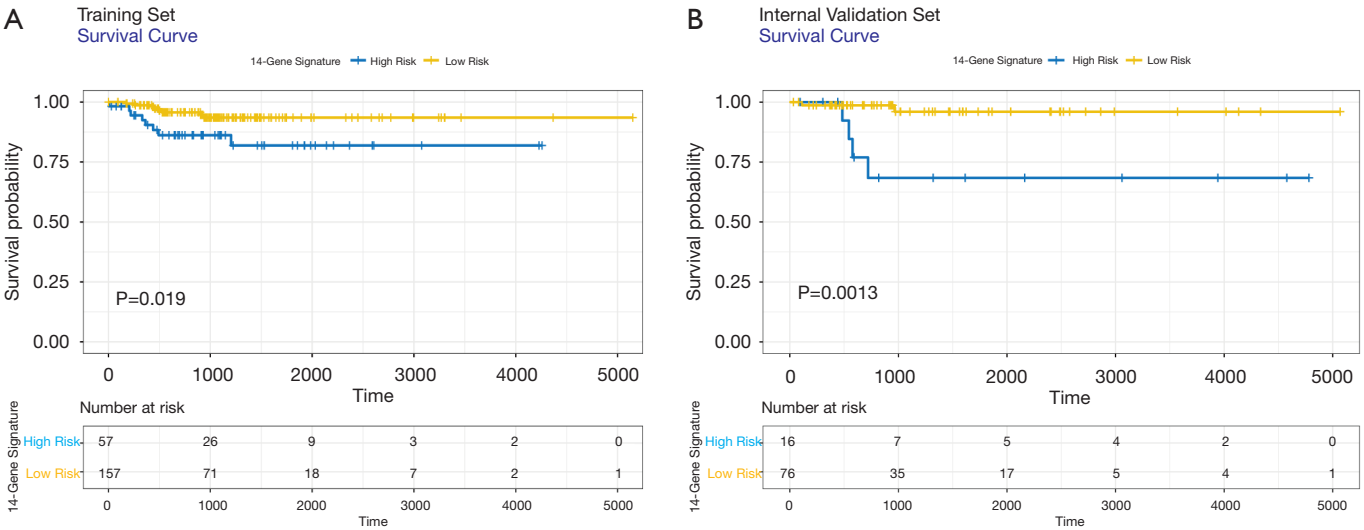

Figure 5 Kaplan-Meier analysis of disease-free survival (DFS) in papillary thyroid carcinoma (PTC) patients in different risk groups according to the 14-gene signature. (A) Kaplan-Meier analysis of DFS in PTC patients in different risk groups according to the 14-gene signature in the training set. (B) Kaplan-Meier analysis of DFS in PTC patients in different risk groups according to the 14-gene signature in the internal validation set.

genes: *LPAR5*, *NMU*, *FN1*, *NPY1R* and *CXCL12*. Shen *et al.* (13) proposed that the DEGs between the tumor and normal samples were mainly associated with extracellular matrix–receptor interaction, p53 signaling pathway, and transforming growth factor-β (TGF-β) signaling pathway. The DEGs related to thyroid carcinoma LNM have been identified by Ruiz *et al.* (14), and a 25-gene panel has been constructed to differentiate N0 and N1 papillary thyroid cancer samples. Song *et al.* (19) revealed that mesenteric estrogen-dependent adipogenesis is a predictor of LNM in PTC. In our present study, a stepwise screening based on the severity of LNM in PTC was performed, which is different from the previous studies. We finally isolated 69 DEGs that were continuously upregulated or downregulated from N0 to N1a and from N1a to N1b. Based on the LASSO regression analysis, a novel 14-gene signature was constructed for predicting LNM in PTC patients.

In comparison with the 25-gene panel developed by Ruiz *et al.* (14), our 14-gene signature includes fewer genes and obtained fairly favorable AUC values, suggesting that it might be easier to apply in clinical practice. It is worth mentioning that due to the lack of validation of the 25-gene panel, its reliability is limited. In addition, our multivariant logistic regression analysis illustrated that the 14-gene signature was a potential indicator of LNM. For the nomogram we established, the length of the line corresponding to the 14-gene risk score also reflected the highest contribution to LNM compared with other potential risk factors in PTC patients.

During clinical practice, the risk of central or lateral LNM could be evaluated according to the optimal cut-off value determined by the ROC curve. The ROC curves showed that when the risk score was ≥0.489, patients might have a higher likelihood of central LNM. Also, when the risk score was ≥0.559, patients might be at a high risk of lateral LNM. Therefore, quantitative real-time polymerase chain reaction (qRT-PCR) could be conducted for tissues obtained from preoperative fine needle biopsy, and the risk score of the 14-gene signature could be calculated to guide surgical decision-making. Moreover, based on the 14-gene risk score, low-risk patients exhibited a lower risk of recurrence. High-risk patients require close monitoring and follow-up, and secondary surgery or radioactive iodine (RAI) therapy should be performed if necessary.

There are still some limitations to our research. Firstly, all the clinical and transcriptome data collected in our study were based on public TCGA datasets, so the model's accuracy should be further verified using samples collected from our clinical practice. Secondly, some potential factors could not show their significance due to the sample size, so further research with larger sample size is necessary.

Conclusions

We identified a novel 14-gene signature for predicting LNM in PTC patients, and the risk score also correlated with DFS in PTC patients. A larger number of clinical cases is necessary for further research to validate the accuracy of the 14-gene signature.

Acknowledgments

We acknowledge the TCGA database for providing their platforms and contributors for uploading their meaningful datasets. We thank the English language editors: K. Brown and J. Chapnick from AME Editing Service for revising the language of this article.
Funding: This work was supported by the Beijing Municipal Health System Academic Leaders of High-level Health Personnel Program, China (No. 2011-2-28).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://dx.doi.org/10.21037/gs-21-361

Peer Review File: Available at https://dx.doi.org/10.21037/gs-21-361

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://dx.doi.org/10.21037/gs-21-361). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Our present study was based upon open-source data obtained from The Cancer Genome Atlas (TCGA, https://www.cancer.gov/tcga), which belongs to a public database. The patients involved in the database have given ethical approval. Users can download relevant data for free for research and publish relevant articles.

References

1. Kim J, Gosnell JE, Roman SA. Geographic influences in the global rise of thyroid cancer. Nat Rev Endocrinol 2020;16:17-29.
2. La Vecchia C, Malvezzi M, Bosetti C, et al. Thyroid cancer mortality and incidence: a global overview. Int J Cancer 2015;136:2187-95.
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin 2020;70:7-30.
4. Luo X, Wang J, Xu M, et al. Risk model and risk stratification to preoperatively predict central lymph node metastasis in papillary thyroid carcinoma. Gland Surg 2020;9:300-10.
5. Ling Y, Zhang L, Li K, et al. Carbon nanoparticle-guided intraoperative lymph node biopsy predicts the status of lymph nodes posterior to right recurrent laryngeal nerve in cN0 papillary thyroid carcinoma. Gland Surg 2021;10:1554-63.
6. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid 2016;26:1-133.
7. Khokhar MT, Day KM, Sangal RB, et al. Preoperative High-Resolution Ultrasound for the Assessment of Malignant Central Compartment Lymph Nodes in Papillary Thyroid Cancer. Thyroid 2015;25:1351-4.
8. Brassard M, Borget I, Edet-Sanson A, et al. Long-term follow-up of patients with papillary and follicular thyroid cancer: a prospective study on 715 patients. J Clin Endocrinol Metab 2011;96:1352-9.
9. Ling Y, Zhao J, Zhao Y, et al. Role of intraoperative neuromonitoring of recurrent laryngeal nerve in thyroid and parathyroid surgery. J Int Med Res 2020;48:300060520952646.
10. Wang J, Chi S, Huang Z, et al. Comprehensive characterization of differentially expressed genes in thyroid cancer. Future Oncol 2017;13:2159-69.
11. Zhang C, Bo C, Guo L, et al. BCL2 and hsa-miR-181a-5p

are potential biomarkers associated with papillary thyroid cancer based on bioinformatics analysis. World J Surg Oncol 2019;17:221.
12. Liu L, He C, Zhou Q, et al. Identification of key genes and pathways of thyroid cancer by integrated bioinformatics analysis. J Cell Physiol 2019;234:23647-57.
13. Shen Y, Dong S, Liu J, et al. Identification of Potential Biomarkers for Thyroid Cancer Using Bioinformatics Strategy: A Study Based on GEO Datasets. Biomed Res Int 2020;2020:9710421.
14. Ruiz EML, Niu T, Zerfaoui M, et al. A novel gene panel for prediction of lymph-node metastasis and recurrence in patients with thyroid cancer. Surgery 2020;167:73-9.
15. Kim YS. Patterns and predictive factors of lateral lymph node metastasis in papillary thyroid microcarcinoma. Otolaryngol Head Neck Surg 2012;147:15-9.
16. Sun J, Jiang Q, Wang X, et al. Nomogram for Preoperative Estimation of Cervical Lymph Node Metastasis Risk in Papillary Thyroid Microcarcinoma. Front Endocrinol (Lausanne) 2021;12:613974.
17. Zhu J, Zheng J, Li L, et al. Application of Machine Learning Algorithms to Predict Central Lymph Node Metastasis in T1-T2, Non-invasive, and Clinically Node Negative Papillary Thyroid Carcinoma. Front Med (Lausanne) 2021;8:635771.
18. Rui ZY, Liu Y, Zheng W, et al. A retrospective study of the risk factors and the prognosis in patients with papillary thyroid carcinoma depending on the number of lymph node metastasis. Clin Exp Med 2021;21:277-86.
19. Song Y, Fu LJ, Li HT, et al. Evaluation of MEDAG gene expression in papillary thyroid microcarcinoma: associations with histological features, regional lymph node metastasis and prognosis. Sci Rep 2019;9:5800.

(English Language Editors: K. Brown and J. Chapnick)

Cite this article as: Ling Y, Jia L, Li K, Zhang L, Wang Y, Kang H. Development and validation of a novel 14-gene signature for predicting lymph node metastasis in papillary thyroid carcinoma. Gland Surg 2021;10(9):2644-2655. doi: 10.21037/gs-21-361

Table S1 Continuously Upregulated or Downregulated Genes in the Process of N0 to N1a and N1a to N1b in PTC patients

Gene Symbol	Description	Trends
GALNT7	Polypeptide N-Acetylgalactosaminyltransferase 7	Up
ATP6V1B2	ATPase H+ Transporting V1 Subunit B2	Up
LRATD2	LRAT Domain Containing 2	Up
SLCO4A1	Solute Carrier Organic Anion Transporter Family Member 4A1	Up
SNX27	Sorting Nexin 27	Up
BRIX1	Biogenesis of Ribosomes BRX1	Up
DNAJC21	DnaJ Heat Shock Protein Family (Hsp40) Member C21	Up
OR5A1	Olfactory Receptor Family 5 Subfamily A Member 1	Up
BAZ1A	Bromodomain Adjacent to Zinc Finger Domain 1A	Up
TREM1	Triggering Receptor Expressed on Myeloid Cells 1	Up
PI15	Peptidase Inhibitor 15	Up
ANPEP	Alanyl Aminopeptidase, Membrane	Up
ZMYND8	Zinc Finger MYND-Type Containing 8	Up
NR2F1	Nuclear Receptor Subfamily 2 Group F Member 1	Up
FXYD5	FXYD Domain Containing Ion Transport Regulator 5	Up
PABPC4L	Poly(A) Binding Protein Cytoplasmic 4 Like	Up
RARA	Retinoic Acid Receptor Alpha	Up
CRYBG1	Crystallin Beta-Gamma Domain Containing 1	Up
ARHGAP17	Rho GTPase Activating Protein 17	Up
WFS1	Wolframin ER Transmembrane Glycoprotein	Down
CFAP46	Cilia and Flagella Associated Protein 46	Down
CLCNKB	Chloride Voltage-Gated Channel Kb	Down
PEG3	Paternally Expressed 3	Down
ATP2C2	ATPase Secretory Pathway $Ca2^{+}$ Transporting 2	Down
ALDH1A1	Aldehyde Dehydrogenase 1 Family Member A	Down
RILPL2	Rab Interacting Lysosomal Protein Like 2	Down
AKAP3	A-Kinase Anchoring Protein 3	Down
RILPL1	Rab Interacting Lysosomal Protein Like 1	Down
MAPK8IP1	Mitogen-Activated Protein Kinase 8 Interacting Protein 1	Down
FAM240C	Family with Sequence Similarity 240 Member C	Down
C12orf60	Chromosome 12 Open Reading Frame 60	Down
RWDD2A	RWD Domain Containing 2A	Down
ARHGEF33	Rho Guanine Nucleotide Exchange Factor 33	Down
ZNF79	Zinc Finger Protein 79	Down
INKA2	Inka Box Actin Regulator 2	Down
ZNF544	Zinc Finger Protein 544	Down
RDH13	Retinol Dehydrogenase 13	Down
CABLES1	Cdk5 and Abl Enzyme Substrate 1	Down
PRDM16	PR/SET Domain 16	Down
TM2D2	TM2 Domain Containing 2	Down
SDR42E1	Short Chain Dehydrogenase/Reductase Family 42E, Member 1	Down
HS6ST3	Heparan Sulfate 6-O-Sulfotransferase 3	Down
ESRRG	Estrogen Related Receptor Gamma	Down
ZBED9	Zinc Finger BED-Type Containing 9	Down
TRPV6	Transient Receptor Potential Cation Channel Subfamily V Member 6	Down
VPS37D	VPS37D Subunit Of ESCRT-I	Down
TUB	TUB Bipartite Transcription Factor	Down
MYCN	MYCN Proto-Oncogene, BHLH Transcription Factor	Down
KIAA0319L	KIAA0319 Like	Down
PLA2R1	Phospholipase A2 Receptor 1	Down
SLC5A7	Solute Carrier Family 5 Member 7	Down
NCOA5	Nuclear Receptor Coactivator 5	Down
BMP7	Bone Morphogenetic Protein 7	Down
NUPR2	Nuclear Protein 2, Transcriptional Regulator	Down
DDOST	Dolichyl-Diphosphooligosaccharide--Protein Glycosyltransferase Non-Catalytic Subunit	Down
PPP1R42	Protein Phosphatase 1 Regulatory Subunit 42	Down
NCS1	Neuronal Calcium Sensor 1	Down
ZDHHC22	Zinc Finger DHHC-Type Palmitoyltransferase 22	Down
SDHAF3	Succinate Dehydrogenase Complex Assembly Factor 3	Down
KCNJ13	Potassium Inwardly Rectifying Channel Subfamily J Member 13	Down
TMEM86A	Transmembrane Protein 86A	Down
AQP11	Aquaporin 11	Down
NRDE2	NRDE-2, Necessary for RNA Interference, Domain Containing	Down
ROPN1B	Rhophilin Associated Tail Protein 1B	Down
SYNGR1	Synaptogyrin 1	Down
METTL2A	Methyltransferase Like 2A	Down
ZNF618	Zinc Finger Protein 618	Down
APMAP	Adipocyte Plasma Membrane Associated Protein	Down
ZNF582	Zinc Finger Protein 582	Down

https://dx.doi.org/10.21037/gs-21-361