

## Hyper-expression and hypomethylation of TM4SF1 are associated with lymph node metastases in papillary thyroid carcinoma patients

Kun WANG<sup>1,2</sup>, Juan ZHENG<sup>3</sup>, Zhen WU<sup>4,5</sup>, Ju-Gao FANG<sup>4</sup>, Jun-Yu ZHAO<sup>6</sup>, Jin-Ming YAO<sup>6</sup>, Jian-Jun DONG<sup>7</sup>, Jie BAI<sup>2</sup>, Lin LIAO<sup>1,6,\*</sup>

<sup>1</sup>Department of Endocrinology and Metabology, Shandong Qianfoshan Hospital, Cheeloo College of Medicine, Shandong University, Jinan, Shandong, China; <sup>2</sup>Department of Endocrinology and Metabology, Liaocheng People's Hospital, Liaocheng, Shandong, China; <sup>3</sup>Joint Laboratory for Translational Medicine Research, Liaocheng People's Hospital, Liaocheng, Shandong, China; <sup>4</sup>Department of Otorhinolaryngology Head and Neck Surgery, Beijing Tongren Hospital, Capital Medical University, Beijing, China; <sup>5</sup>Department of Breast and Thyroid Surgery, Liaocheng People's Hospital, Liaocheng, Shandong, China; <sup>6</sup>Department of Endocrinology and Metabology, The First Affiliated Hospital of Shandong First Medical University and Shandong Provincial Qianfoshan Hospital, Jinan, Shandong, China; <sup>7</sup>Department of Endocrinology, Qilu Hospital of Shandong University, Jinan, Shandong, China

\*Correspondence: liaolin@sdu.edu.cn

Received August 19, 2021 / Accepted December 6, 2021

Lymph node metastases (LNM) are an indicator for recurrence in papillary thyroid carcinoma (PTC) patients. However, prophylactic neck dissection (ND) cannot improve survival or recurrence rate because of increased surgical complications and occult LNM. Biomarkers are needed for the prediction of high-risk of LNM to avoid unnecessary operation and reduce the missed malignant lymph nodules. GEO database was searched for the differentially expressed genes (DEGs) between PTC patients with LNM (N1) and those without LNM (N0), transcriptional and methylation data of DEGs in THCA were examined from databases. The expression and methylation of TM4SF1 in fresh and paraffin tissues of PTC patients were examined by qRT-PCR, IHC, and MSP. TM4SF1 was the only one significantly associated with LNM. UALCAN revealed that TM4SF1 was overexpressed and hypomethylated in LNM patients. MEXPRESS presented a negative correlation between gene expression and promoter methylation of TM4SF1. DEGs were enriched in multiple pathways and the Extracellular Matrix (ECM)-receptor interaction pathway showed the greatest correlation with LNM. IHC and qRT-PCR of tissues demonstrated that the expression of TM4SF1 in the N1 group was 4.5-fold higher than that in the N0 group ( $p < 0.05$ ). MSP exhibited that the positive rate of aberrant promoter methylation of TM4SF1 was 8.38% in N1 and 66.7% in N0 group ( $p < 0.05$ ). Hyper-expression and hypomethylation of TM4SF1 are associated with lymph node metastases in PTC patients.

*Key words:* papillary thyroid carcinoma, lymph node metastases, TM4SF1, ECM-receptor interaction, methylation

With a rapidly rising incidence in recent years, thyroid cancer (THCA) has been the most common endocrine malignancy [1–3]. Papillary thyroid carcinoma (PTC) accounts for approximately 90% of all the pathological types in diagnosed THCA cases [4]. Lymph node metastases (LNM) occur subclinically in a majority of PTC patients [5–7] and are an indicator for PTC recurrence [8, 9]. In addition, the presence of LNM is associated with compromised survival for PTC patients younger than 45 years of age [10]. However, the role of prophylactic lymph node dissection in the treatment of PTC is controversial because of its excessive complications and missed diagnosis rate [11]. Therefore, new and better markers are needed for the prediction of the high risk of LNM to avoid unnecessary operation and reduce the missed diagnosis of malignant lymph nodules.

Transmembrane-4-L-six-family-1 (TM4SF1), an L6 family member, is a 22 kDa four-transmembrane-domain protein [12–15]. TM4SF1 is overexpressed in the cell membrane and intracellular vesicles of lung, breast, colon, ovary, kidney, prostate cancer, and other epithelial malignant tumors, and hypoexpressed in normal vascular endothelial cells [16, 17]. TM4SF1 has been shown to be associated with the growth, motility, invasion, and metastasis of tumor cells [18].

Although TM4SF1 has been verified closely associated with metastasis in multiple tumors, whether TM4SF1 is related to LNM of PTC patients has not been reported. Our research was aimed to identify whether TM4SF1 was the candidate driver gene in LNM of PTC through the analyses of multiple gene expression array datasets and verify its expression and methylation in clinical tissues to explore possible molecular mechanisms.

## Patients and methods

**Data collection and microarray data.** GEO database [19] (<https://www.ncbi.nlm.nih.gov/geo/>) was searched for public studies prior to June 15, 2020. The following keywords were used: “papillary thyroid cancer” (Research keywords), “Homo sapiens” (organism), “array expression profile” (research type). The inclusion criteria of the study were as follows: 1) patients diagnosed with PTC with LNM tissue samples and PTC without LNM tissue samples; 2) gene expression profiling of mRNA; and 3) sufficient information to perform the analysis. After a systematic review, two gene expression profiles (GSE60542 and GSE129562) were collected for analysis. The GSE60542 dataset contained 19 PTC tissue samples with LNM (N1) and 14 PTC tissue samples without LNM (N0); the GSE129562 dataset included 5 PTC tissue samples with LNM (N1) and 3 PTC tissue samples without LNM (N0); the platforms used for gene profiling of GSE60542 and GSE129562 were GPL15207, GPL10558 (Affymetrix), respectively.

**Data analyses of differentially expressed genes.** GEO2R ([www.ncbi.nlm.nih.gov/geo/geo2r](http://www.ncbi.nlm.nih.gov/geo/geo2r)) [19], an interactive web tool that compares two or more groups of samples under the same experimental conditions in a GEO dataset, was used to identify common differentially expressed genes (DEGs) in the two microarray datasets. The genes that satisfied the inclusion criteria of  $p < 0.05$  and  $|\log FC| \geq 1$  were identified as DEGs. Funrich software [20] (<http://www.funrich.org/>) was used to identify common DEGs in the two datasets and to plot Venn diagrams.

**Expression and methylation of DEGs in THCA.** UALCAN (<http://ualcan.path.uab.edu>) [21], an easy-to-use interactive portal for in-depth analysis of TCGA gene expression data. UALCAN uses TCGA level 3 RNA-seq and clinical data from 31 cancer types. The mRNA expression and DNA promoter methylation of TM4SF1 in THCA patients with or without LNM were examined using the UALCAN database online.

MEXPRESS, an online tool for the integration of gene expression, DNA methylation, and clinical data from The Cancer Genome Atlas (TCGA) [22]. The correlation between the expression and its promoter methylation of TM4SF1 was explored from MEXPRESS.

**Pathway enrichment analyses of TM4SF1.** KEGG, an encyclopedia of genes and genomes, is available at <http://www.kegg.jp/> or <http://www.genome.jp/kegg/> [23]. Molecular-level functions are stored in the KO (KEGG Orthology) database, where each KO is defined as a functional ortholog of genes and proteins. Genomes (KEGG) analyses were all analyzed via R (version 3.6.3, AT&T Bell Laboratories, New York, NY, USA) software package “cluster Profiler”. And then, we constructed two bubble charts (based on p-value) representing the top 15 enrichment biological functions and pathways by using R software package “ggplot2”.

**Correlation analyses.** GEPIA (Gene Expression Profiling Interactive Analysis), a web-based tool that provides fast and

customizable functions based on TCGA and GTEx database, is available at <http://gepia.cancer-pku.cn/> [24]. In this study, we analyzed the correlations between TM4SF1 and pathway-related proteins using GEPIA.

**Immunohistochemical analysis (IHC).** The expression of TM4SF1 was verified in tumor tissues of 24 PTC patients obtained from Liaocheng People’s Hospital. Of these patients, 12 PTC cases had LNM (N1) and the other 12 PTC cases did not have LNM (N0). The 4  $\mu\text{m}$  thick FFPE tissue was dewaxed in xylene and then rehydrated by ethanol-water gradient. After 15 min of high temperature and pressure treatment in the antigen repair solution, the sections were incubated in 3% hydrogen peroxide for 10 min to inactivate the endogenous peroxidase activity. The sections were then incubated overnight with TM4SF1 (1:200, ab113504, Abcam, Cambridge, USA) antibody at 4°C. After washing with PBS, slide with combined goat anti-rabbit secondary antibody (pv9001, 1:400, Zhongshan, Beijing, China) and peroxidase enzyme was incubated at 37°C for 30 min. Peroxidase staining was detected with DAB Peroxidase Substrate (Vector Lab, SK-4013, USA). Then, the slides were counterstained with hematoxylin. The study was approved by the Research Ethics Committee of Liaocheng People’s Hospital. Informed consent was obtained from all participating patients.

**RNA extraction and quantitative real-time polymerase chain reaction (qRT-PCR).** Twenty-four cases of fresh tissue of PTC patients who underwent thyroidectomy were enrolled from February 2020 to June 2020 at Beijing Tongren Hospital. Of these patients, 12 PTC cases had LNM (N1) and the other 12 PTC cases did not have LNM (N0). According to the manufacturer’s instructions, RNA was isolated from tissues using the QIAamp RNA Mini kit (Qiagen, Hilden, Germany). Then, RNA was reverse transcribed into cDNA with Invitrogen and the cDNA was used as a template for real-time PCR (RT-PCR) detection. The expression level of TM4SF1 was detected by qRT-PCR. The Applied Biosystems Prism 7900 sequence detection system with SYBR Green was used for the next analysis of qRT-PCR (PE Applied Biosystems, [www.appliedbiosciences.com](http://www.appliedbiosciences.com)). The amplification conditions of all primers were the same: 50°C amplification for 2 min, 95°C amplification for 10 min, then 40 cycles of 95°C amplification for 30 s, 60°C amplification for 30 s, then 72°C amplification for 30 s. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as the internal control standard. The experiment was carried out independently in triplicate, and the number of qPCR cycles was converted into gene quantity (NG) by the accepted formula. The relative expression level between the two treatments was calculated by  $2^{-(\Delta ct^{\text{sample}} - \Delta ct^{\text{control}})}$ . The primers sequences: TM4SF1 F: TTTGGCATCGTAGGAGGTG, R: GCCACAG-CAGTCATCCTGTT; GAPDH F: TCGACAGTCAGCCG-CATCTTCTTT, R: ACCAAATCCGTTGACTCCGACCTT.

**DNA extraction, modification, methylation-specific PCR (MSP).** Genomic DNA was isolated from samples by using the QIAamp DNA Mini kit (Qiagen, Hilden,

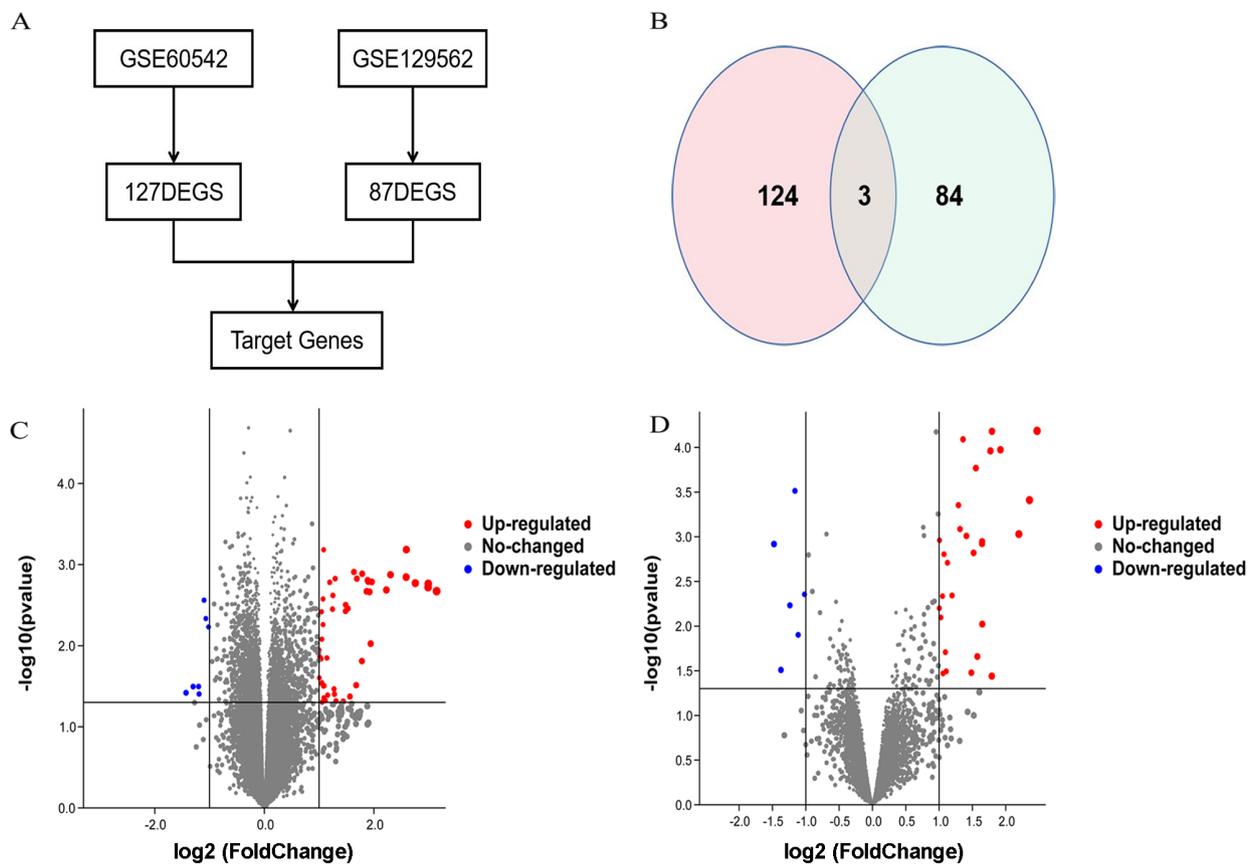
Germany). Bisulfite modification of DNA was performed using ZYMO EZ DNA Methylation Kit (ZYMO Research, D5005, Los Angeles, USA), bisulfite modification was set as follows: 98 °C for 10 min, 64 °C for 2.5 h, and 4 °C for 20 h. MSP was performed using TianGen DNA MSP kit (TianGen, EM101, Beijing, China). The MSP amplification was set as follows: One cycle at 95 °C for 5 min; 35 cycles at 94 °C for 20 s, 30 s at 60 °C, and 20 s at 72 °C; one cycle at 72 °C for 5 min. The PCR products were analyzed by electrophoresis on a 3% agarose gel and by ethidium bromide staining. The methylation-specific PCR primers sequences: TM4SF1-M-F TAGGATGGATTTTAGTGTTTAGCGTT; TM4SF1-M-R AACATTACTTTTCCTAACAACAACACTC; TM4SF1-U-FTTGTAGGATGGATTTTAGTGTTTAGTGT; TM4SF1-U-R AACATTACTTTTCCTAACAACAACACTC

**Statistical analysis.** All statistical analyses were performed using SPSS 25.0 software (IBM's Statistical Product and Service Solutions). The results were analyzed by statistics method independent t-test,  $\chi^2$ -test, and Fisher exact test. A p-value <0.05 represents statistical difference.

## Results

**Three significantly DEGs between N0 and N1.** Two datasets (GSE60542 and GSE129562), including 24 N1 samples and 17 N0 samples, were identified from the GEO database. GEO2R, the online analysis tool of the GEO database, was employed for the differentially expressed genes. Among all the differential genes, the ones with p-value <0.05 and  $|\log_2FC| \geq 1$  were regarded as candidate DEGs. The study flow chart of selection was shown in Figure 1A. There were 116 upregulated and 11 downregulated DEGs in GSE60542, 82 upregulated and 5 downregulated DEGs in GSE129562. Finally, we got 3 significantly differentially expressed genes between N1 and N0 (Figure 1B) by overlapping the DEGs of the two GEO datasets. The three target genes were TM4SF1, SLPI, and RASD1.

**TM4SF1 was overexpressed in THCA patients with LNM.** The mRNA expressions of TM4SF1, RASD1, and SPLI were analyzed in 58 cases of thyroid carcinoma with LNM (N1), 230 cases of thyroid carcinoma without LNM,



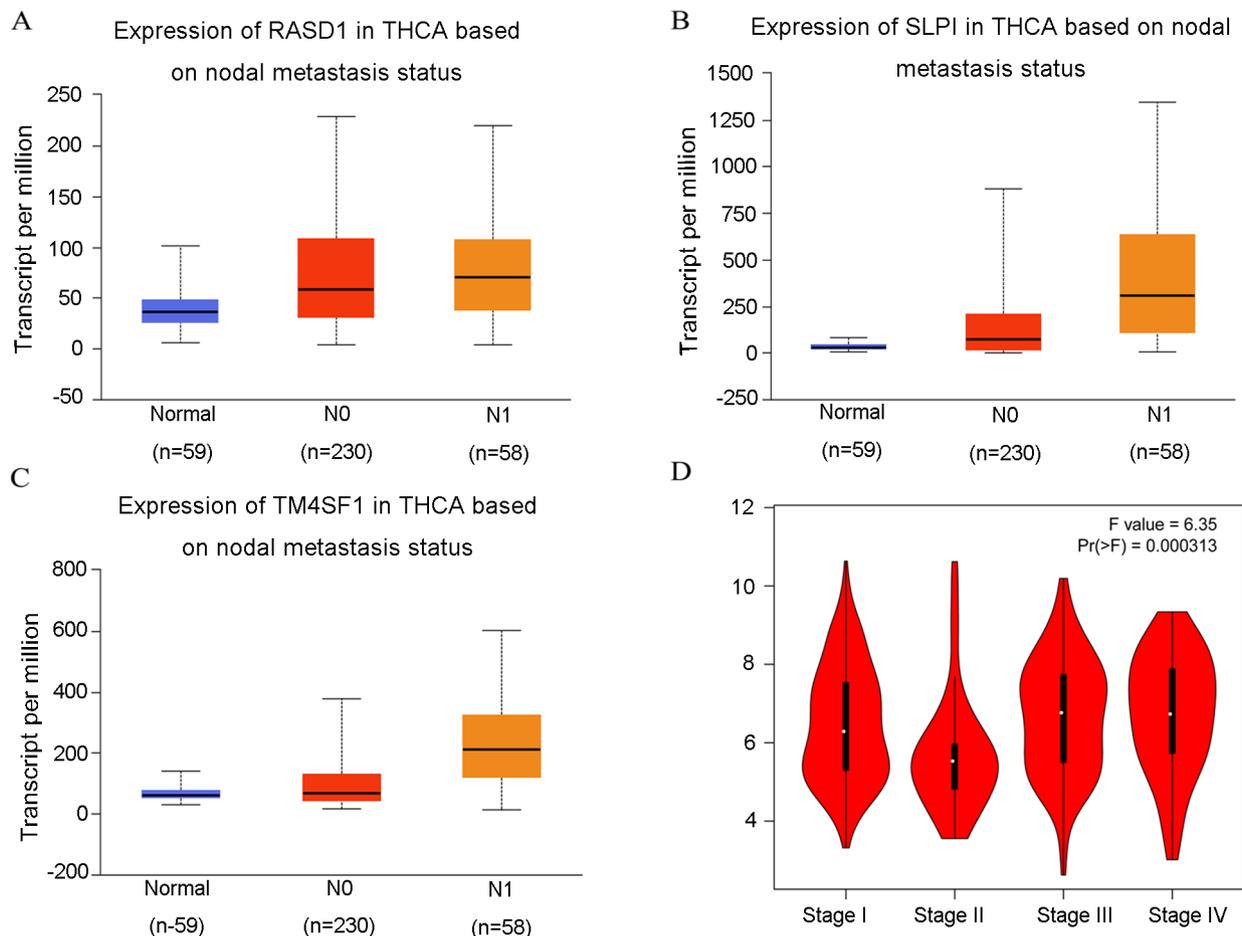
**Figure 1.** Integrated analysis of GEO datasets identified DEGs in N1 and N0. A) Flow chart of the analysis. GSE60542 and GSE129562 as downloaded in order to identify DEGs between N1 and N0. B) Venn diagrams of common DEGs combined with two datasets (GSE60542 and GSE129562). Each circle represents a dataset, and the overlap of the circles amounts to the overlap of the datasets. C, D) Volcano plot for differentially expressed genes of GSE60542 and GSE129562. The x-axis shows the fold-change in gene expression between different samples, and the y-axis shows the statistical significance of the differences. Significantly up- and down-regulated genes are filtered ( $|\log_2 \text{Fold Change}| > 1$ , p-value adj < 0.05) and highlighted in red and blue dots, respectively.

and 59 cases of normal thyroid carcinoma tissues using the UALCAN database. The results suggested that the expression of TM4SF1 was the only significantly upregulated in N1 tumor tissues than that in N0 and normal tissues (Figure 2A). The expressions of the other two genes showed no significant difference between N1 and N0 groups (Figures 2B, 2C). Then, the expression of TM4SF1 in different stages of THCA was examined by GEPIA depending on TCGA database. The Violin Plot suggested that TM4SF1 was highly expressed in a large majority of samples in stage IV.

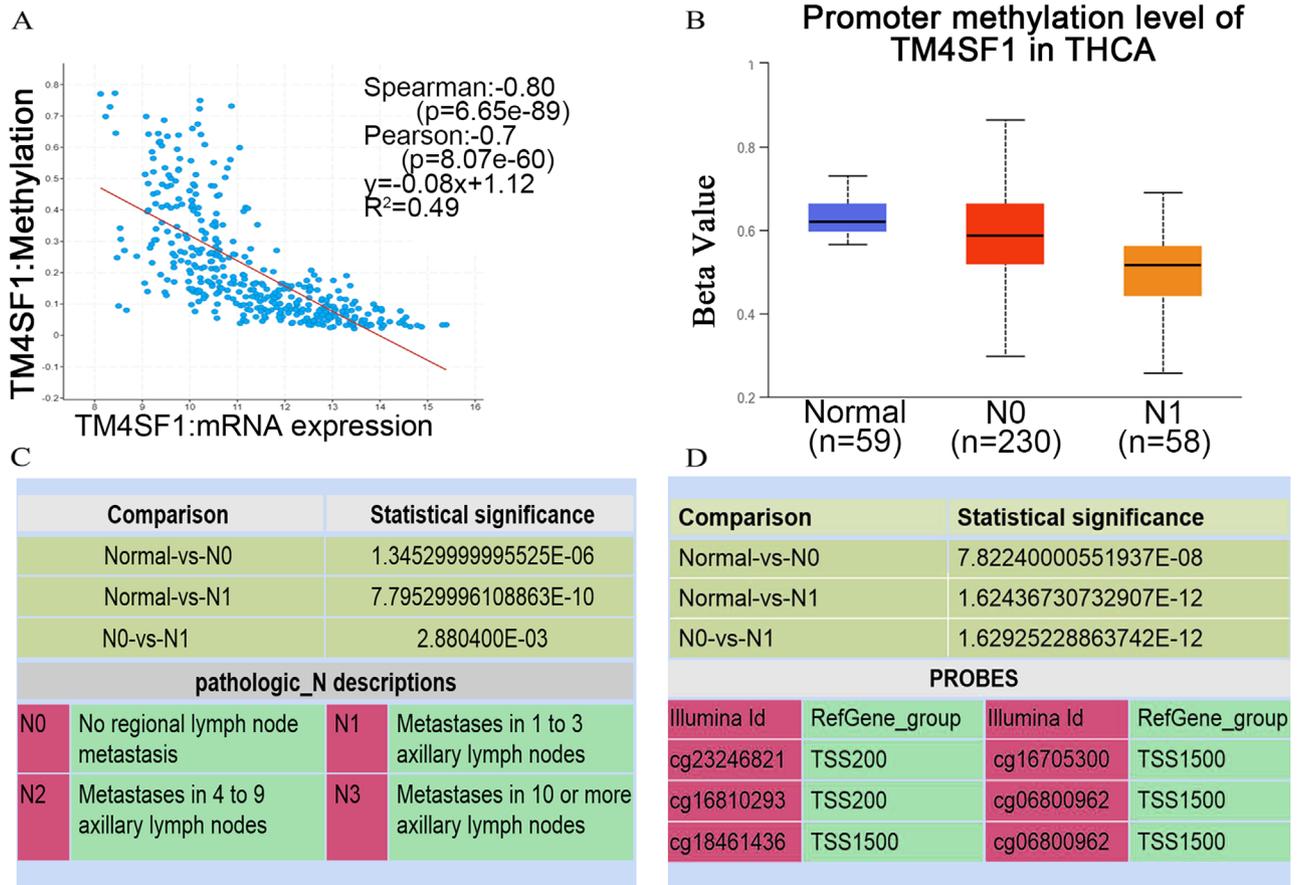
**TM4SF1 was hypomethylated in THCA patients with LNM.** The Spearman's correlation between the expression and methylation of TM4SF1 was determined by analyses of cBioportal (TCGA, Figure 3A). The expression of mRNA of TM4SF1 was significantly negatively correlated with its DNA methylation (Spearman:  $-0.80$ ,  $p=6.65e-89$ ). The DNA promoter methylations of TM4SF1 were also analyzed between N1 and N0 patients of THCA by UALCAN.

Data from probes cg23246821, cg16810293, cg18461436, cg16705300, cg06800962 in Infinium Human Methylation 450K chip were used for the promoter methylation data by UALCAN. The results demonstrated that the DNA promoter methylations of TM4SF1 in N1 patients were the lowest among normal, N0, and N1 patients of THCA (N1 vs. N0,  $p<0.01$ ; N0 vs. normal,  $p<0.01$ , Figure 3).

**The expression of TM4SF1 was negatively correlated with promoter methylation.** Then, the correlation between methylation of promoter probes and expression of TM4SF1 was analyzed by the MEXPRESS tool. As shown in Figure 4A, this tool permitted us to analyze the methylation of TM4SF1 tested with 7 probes distributed in different regions of the gene (the localization of each probe was represented in the figure and the ones localized in the promoter region were highlighted in the red box, the ID of promoter probes were cg08124030, cg23246821, cg16810293, and cg06800962). All the promoter regions analyses presented a negative correla-



**Figure 2.** The expression of DEGs in THCA based on nodal metastasis. A) The expression of TM4SF1 in THCA based on nodal metastasis (N1 vs. N0,  $p<0.01$ ; N0 vs. normal,  $p<0.01$ ; N1 vs. normal,  $p<0.01$ ). B) The expression of RASD1 in THCA based on nodal metastasis (N1 vs. N0,  $p=0.3$ ; N0 vs. normal,  $p<0.01$ , N1 vs. normal,  $p<0.01$ ). C) The expression of SPLI in THCA based on nodal metastasis (N1 vs. N0,  $p=0.054$ ; N0 vs. normal,  $p<0.01$ ; N1 vs normal,  $p<0.01$ ). D) The expression of TM4SF1 in different stages by GEPIA.



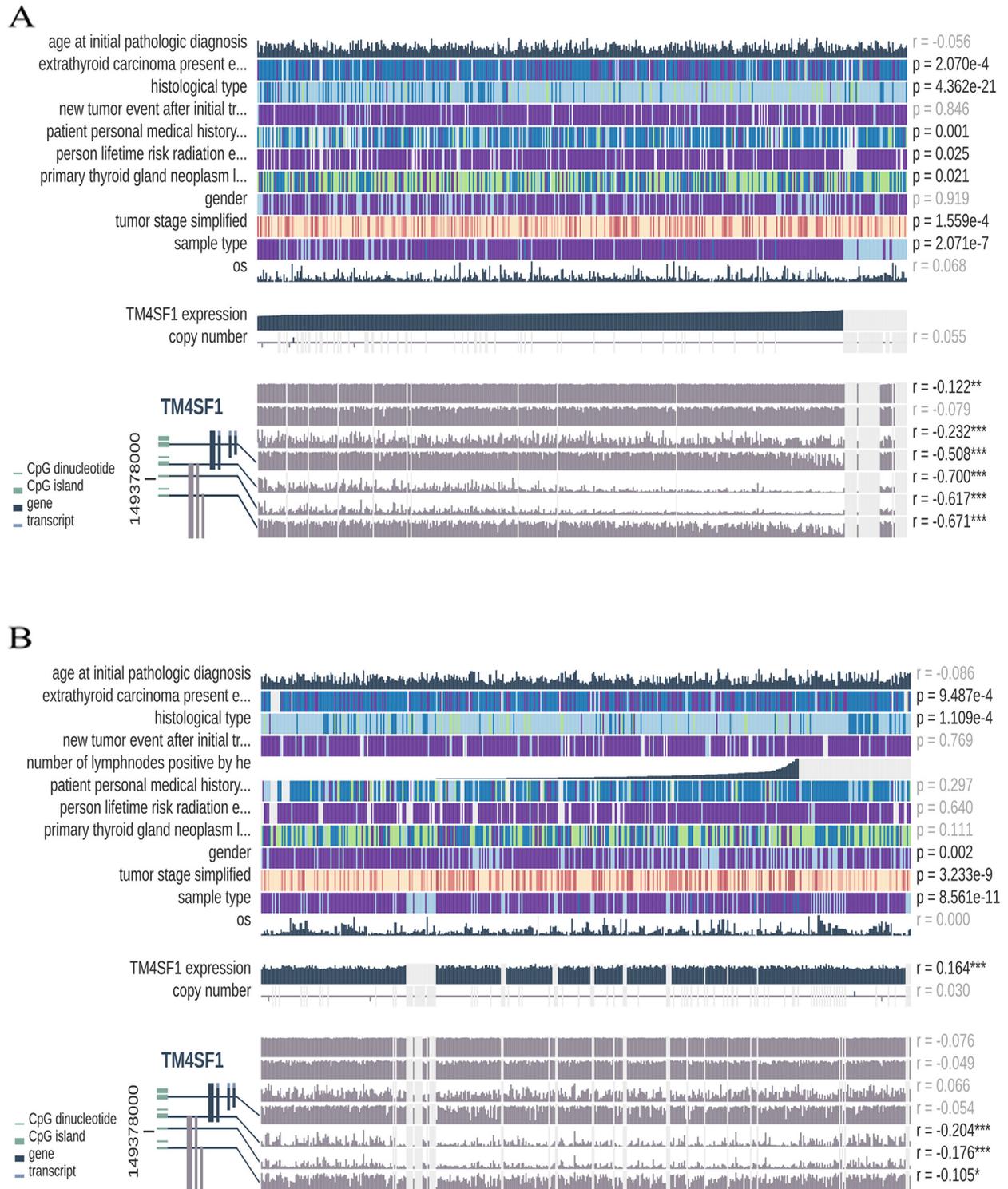
**Figure 3.** The methylation of TM4SF1 in THCA. **A)** The correlation between the expression and methylation of TM4SF1. Spearman:  $-0.80$ ,  $p=6.65e-89$ , Pearson:  $-0.70$ ,  $p=8.07e-60$ ,  $y=-0.08x + 1.12$ ,  $R^2=0.49$ . **B-D)** The promoter methylation of TM4SF1 in THCA based on nodal metastasis (N1 vs. N0,  $p<0.01$ ; N0 vs. normal,  $p<0.01$ ).

tion with respect to the TM4SF1 gene expression (Pearson's correlation coefficients for each probe are indicated on the right). The MEXPRESS tool also allowed us to visualize TM4SF1 expression and methylation status according to the number of lymph nodes positive classification. As shown in Figure 4B, the expression of TM4SF1 was negatively correlated with promoter methylation, which further demonstrated that the expression of TM4SF1 was downregulated with promoter probes methylation in THCA patients.

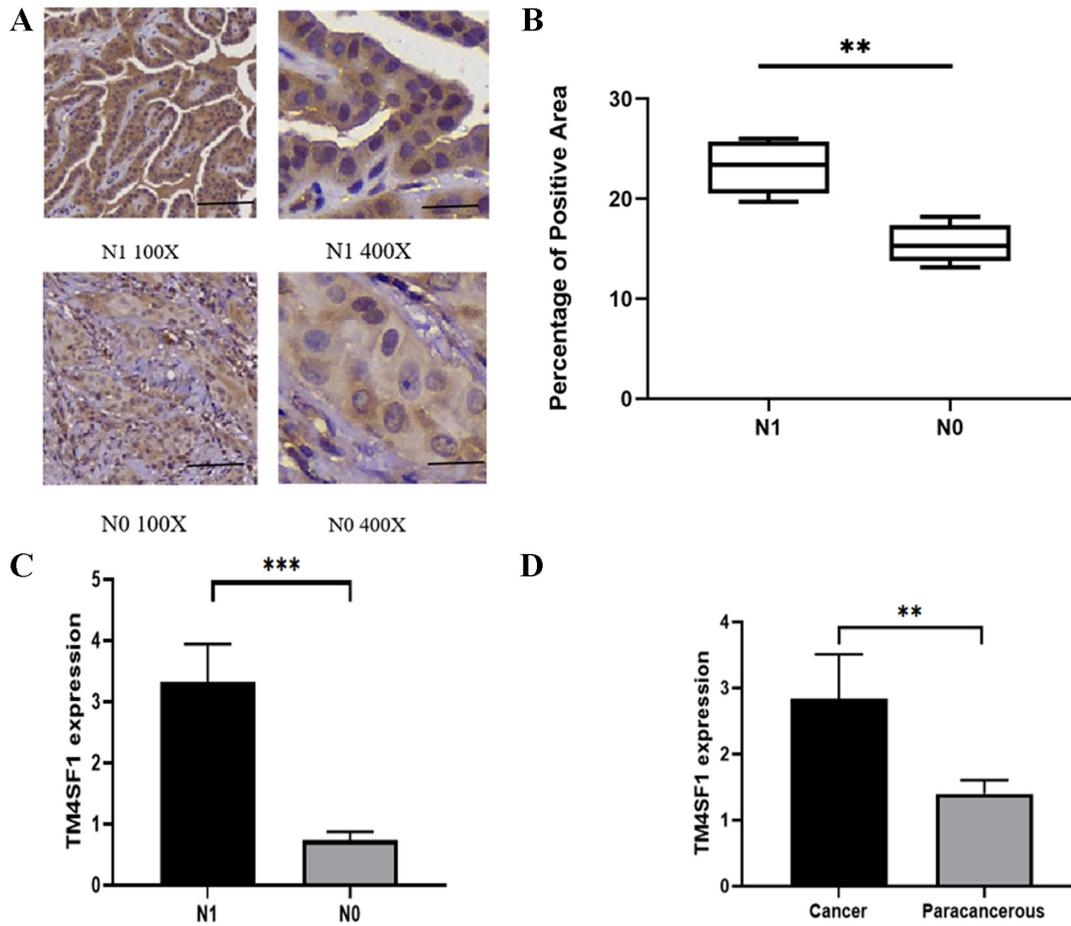
**IHC and qRT-PCR validation of TM4SF1 in tissues of PTC.** TM4SF1 protein expressions of N1 and N0 in thyroid tissues were tested with IHC. The results demonstrated that TM4SF1 was mainly localized in the cytoplasm (Figure 5A). Then, IHC quantification of positive areas was performed using ImageJ analysis software and IHC\_Profiler. Quantitative image analysis for IHC staining was expressed as a positive area percentage across five different fields for each section. The average gray value (staining intensity) and percentage of the positive area (staining area) of positive cells were used as IHC measurement indicators. Four scores were assigned: high-positive (3+), positive (2+), low positive

(1+), and negative (0). High-positive (3+) and positive (2+) areas were defined as positive areas and the percentage of the positive area was calculated. The positive area of the N1 section (23.18%) was significantly higher than that of N0 (15.52%; Figure 5B,  $p<0.01$ ), which indicates that the expressions of TM4SF1 were higher in the N1 tissue. The higher expression of TM4SF1 in the N1 tissue was confirmed by qPCR (Figure 5C). The expressions of TM4SF1 in N1 were 4.5-fold higher than that of N0 ( $p<0.001$ ) and 2-fold higher in PTC cancer compared with paracancerous tissues ( $p<0.01$ ).

**DNA promoter of TM4SF1 was significantly hypomethylated in PTC patients with LNM.** The methylation rates of DNA promoter of TM4SF1 were 16.7% in N1 and 66.7% in N0 tissues as shown in Table 1 ( $p<0.05$ ). Interestingly, we found that the methylation status of all the tissues was MM (complete methylated) and UM (partial methylated, Figures 6A, 6B), which means that none of them was UU (unmethylated). It implied that the LNM of PTC patients might be correlated with the methylation of the DNA promoter of TM4SF1.



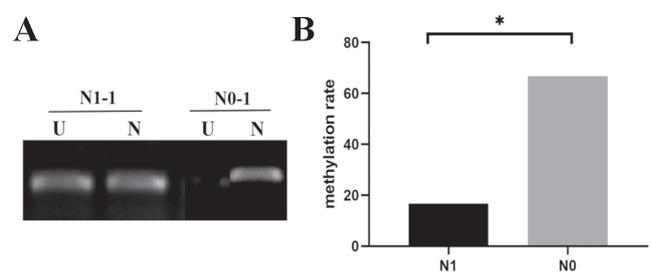
**Figure 4. Correlation between expression and methylation of TM4SF1 in THCA using MEXPRESS. A) TM4SF1 expression and methylation status according to expression classification. B) TM4SF1 expression and methylation status according to a number of lymph nodes positive classification. At the right-hand, the Pearson's correlation coefficient  $r$  and the  $p$  values for Wilcoxon rank-sum test were shown. The expression of TM4SF1 was symbolized as the dark blue line at the top of the plot. The samples were ranked according to their expressions of TM4SF1, the highest expressions were on the right and the lowest on the left. The gray lines represented the Infinium 450 k probes that were linked to TM4SF1. The heights of the gray lines indicated the beta values for the probes. The probes localized in the promoter region of the gene were highlighted in the red box. (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ).**



**Figure 5.** High expression of TM4SF1 in N1 tissues. **A)** TM4SF1 expression in thyroid cancer tissues of N1 and N0 by IHC. Magnification,  $\times 100$  (left panels) and  $\times 400$  (right panels). **B)** Quantification of positive-staining for TM4SF1 in thyroid cancer tissues of N1 and N0 groups. TM4SF1 was significantly higher in N1 than N0 samples.  $**p<0.01$ . **C)** TM4SF1 mRNA expressions were detected by real-time PCR in N1 and N0 tissues. The expressions of TM4SF1 in N1 were 4.5-fold higher than that of N0 ( $p<0.001$ )  $***p<0.001$ . **D)** TM4SF1 mRNA expression was detected by real-time PCR in PTC cancer tissues and paracancerous tissues. The expressions of TM4SF1 in PTC cancer were 2-fold higher than paracancerous tissues.  $**p<0.01$ .

**ECM-receptor interaction pathway might be related to LNM.** The DEGs (127 of GSE60542 and 87 of GSE129562) were subjected to pathway enrichment analyses via “cluster-Profiler” R software package respectively and the top 15 enriched pathways (based on p-value) had been plotted using the “ggplot2” R software package (Figures 7A, 7B). Among the pathways, the ECM-receptor interaction pathway is significant in both GSE60542 and GSE129562 (Supplementary Figure S1). These data demonstrated that the ECM-receptor interaction pathway might be related to LNM.

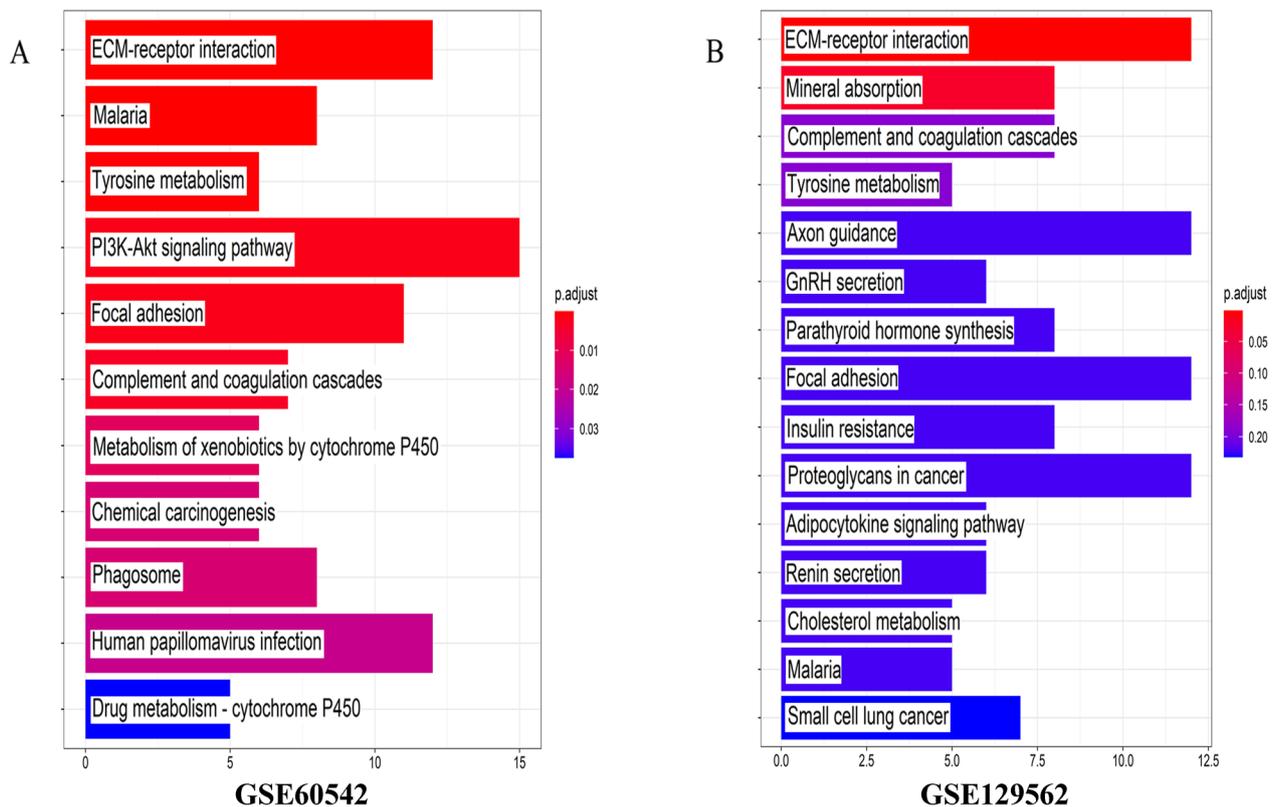
**Correlation between TM4SF1 and the ECM pathway-related proteins.** The extracellular matrix (ECM) is composed of a myriad of fibrous proteins, proteoglycans, and matricellular-associated proteins [25]. Cells sense the mechanical and biochemical properties of the ECM through specialized transmembrane receptors such as DDR1, FN1, COL1A1, COL1A2, MMP2, MMP9, CDH2, and CDH1 to affect the development of invasive tumors and their metastatic dissemination [26]. The Spearman’s correlations



**Figure 6.** DNA promoter of TM4SF1 was hypomethylated in PTC patients with LNM. **A)** The methylation rate of TM4SF1 of N1 and N0.  $*p<0.05$ . **B)** Electrophoretic map of promoter methylation. Abbreviations: M-methylated; U-unmethylated; MM-complete methylated; UU-unmethylated; UM-partial methylated.

**Table 1.** Promoter methylation frequency of TM4SF1 in PTC.

	UM	MM	Rate	p-value
N1	10	2	16.7%	<0.05
N0	5	7	66.7%	



**Figure 7.** Pathway enrichment analyses of DEGs. **A)** Pathway enrichment analyses of DEGs of GSE60542. **B)** Pathway enrichment analyses of DEGs of GSE129562. Y-axis: name of the signaling pathway or function; Notes: X-axis: number of genes assigned to a pathway or function; Color-enriched p-value; red bubble indicates a greater significance level.

between TM4SF1 and ECM-associated proteins in thyroid cancer were determined by analyses of the GEPIA (TCGA, Figures 8A–8H). TM4SF1 had significant correlations with all ECM-associated proteins ( $p < 0.01$ ) and among them, FN1, COL1A1, and MMP2 were the top three proteins.

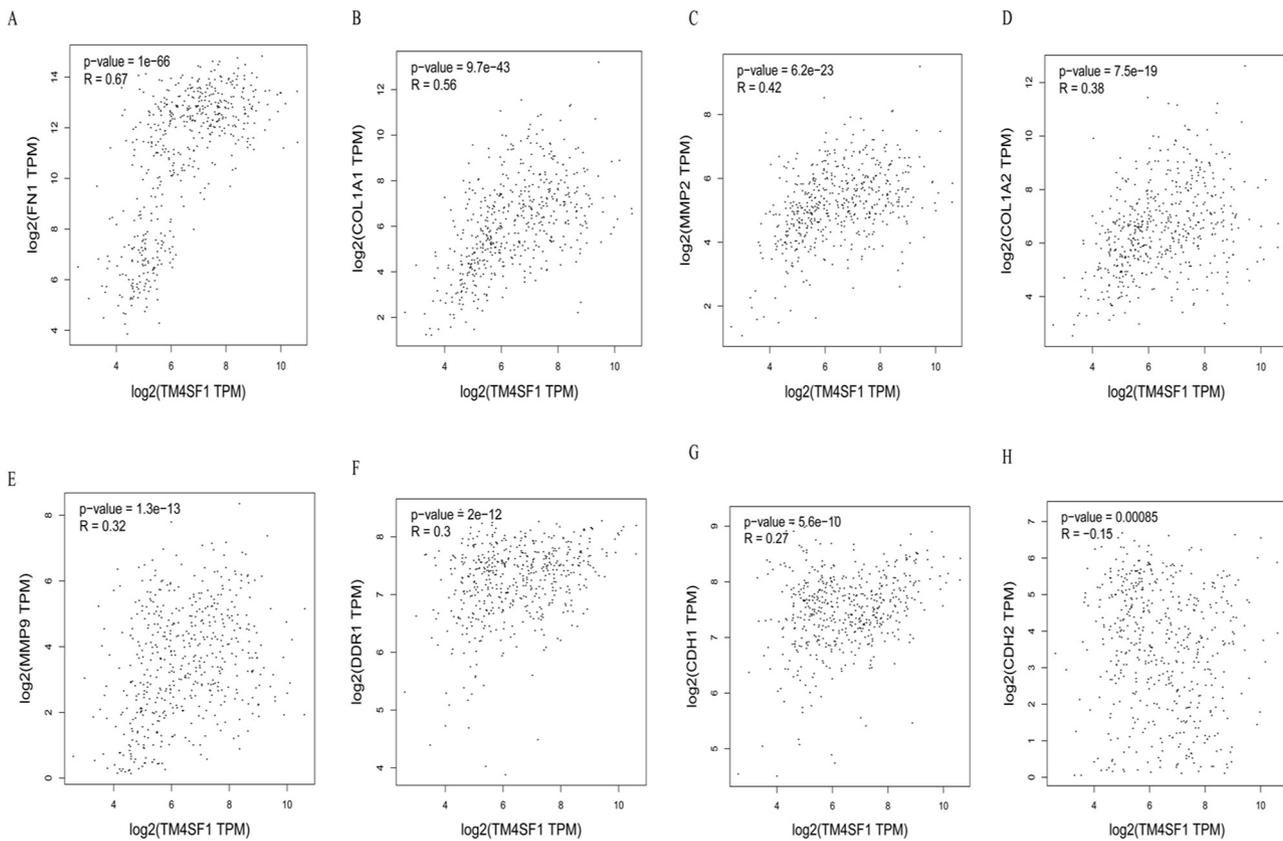
**Relationship between TM4SF1 and clinical characteristics.** We also explored the relationship between TM4SF1 and clinical characteristics. Pearson's correlation analyses of all participants demonstrated that increasing age was positively associated with the expression of TM4SF1 ( $r = 0.694$ ,  $p = 0.012$ ). Meanwhile, there was no significant correlation between sex and the expression of TM4SF1 ( $p = 0.955$ ). The receiver operating characteristic (ROC) curve analysis was carried out using SPSS 25.0 (SPSS, Chicago, IL, USA) to determine the best discrimination point of patients with and without LNM. Considering 1.13 as the normalized cutoff value for TM4SF1, the area under curve (AUC) curve was calculated to be 0.965, with a sensitivity of 0.889 and a specificity of 0.875.

## Discussion

Papillary thyroid cancer (PTC) is the most common well-differentiated cancer of the thyroid, which has a good overall prognosis with a 10-year survival rate  $> 90\%$ . However, once

local recurrence and distant metastasis occur, it will seriously affect the quality of life and prognosis. Lymph node metastases (LNM) are closely related to recurrence and distant metastasis. Thus, in an extensive study, prophylactic neck dissection (ND) cannot improve the survival or recurrence rate of PTC [7]. One of the primary reasons is that excessive ND is associated with increased surgical complications, such as hypoparathyroidism, RLN injury, bleeding, cranial nerve XI palsy, and chylorrhea [7]. What's worse is that about 54.1% of the LNM patients are occult central neck LNM, which means the true metastatic lymph nodes may not have been clearance and a second operation may be performed with more complications [11]. A preoperative diagnosis of PTC by cytology or molecular is necessary to assess for LNM. Once a diagnosis of locally metastatic was performed, the operative plan would be changed [7].

In our study, by searching the GEO database, two datasets and 3 DEGs were found. Combined with the data of cBioPortal, UALCAN, and GEPIA, TM4SF1 was identified as the candidate driver gene of LNM in papillary thyroid cancer, which had been verified in fresh and paraffin tissues by IHC and PCR. Consistently, both Simpson *et al.* and Cao *et al.* confirmed that the expression of TM4SF1 was positively correlated with LNM and poor prognosis in breast cancer



**Figure 8. Correlation between TM4SF1 and the ECM pathway-related proteins. A) FN1,  $r=0.67$ , B) COLA1,  $r=0.56$ , C) MMP2,  $r=0.42$ , D) COLA2,  $r=0.38$ , E) MMP9,  $r=0.32$ , F) DDR1,  $r=0.3$ , G) CDH1,  $r=0.27$ , H) CDH2,  $r=-0.15$ .**

and bladder cancer [27–29]. However, the specific role of TM4SF1 in the control of LNM, which further leads to poor prognosis has not yet been clearly elucidated.

Multiple evidences suggested that there was a strong connection between DNA methylation and cancer [30]. In recent years, more and more studies have found that DNA promoter methylation plays an important role in the development and metastasis of thyroid cancer [31, 32]. It usually silences the gene when it occurs in the promoter region [31]. Interestingly, another study also found that a large range of genes was hypomethylated and hence overexpressed in human cancer [32]. These hypermethylated or hypomethylated genes have important established metabolic and cellular functions. In our study, the data from the MEXPRESS database demonstrated that the TM4SF1 expression is downregulated through promoter probes methylation in THCA patients. Next, we verified that the promoter methylation of TM4SF1 in the N1 tissues was lower than that in the N0 tissues, which further revealed that low promoter methylation might be the mechanism of the high expression of TM4SF1.

The development of invasive tumors and their metastatic dissemination involves a series of discrete biological steps, each of which associates with distinct changes in ECM

composition, posttranslational modifications, organization, and biomechanics [25]. Cells bind to the ECM via transmembrane receptors including integrins, DDRs, and syndecans. From KEGG Orthology and GEPIA, we found that the ECM-receptor interaction pathway might be related to LNM and TM4SF1 had significant correlations with all ECM-associated proteins ( $p < 0.01$ ). Among all these proteins above, FN1, COLA1, and MMP2 were the top three and we have reasons to believe that TM4SF1 might reduce the adhesion between cells by affecting ECM-related proteins such as FN1 and furthermore leads to tumor spread and metastasis.

Correlation analysis between TM4SF1 and age indicated that the old age might increase the risk of lymph node metastases in patients with PTC, which again confirms the importance of early prediction of lymph node metastasis and the urgent need for a biomarker. Nevertheless, although the ROC curve analysis indicated that the sensitivity and specificity are relatively high, the current conclusion might be limited because of the premise of a small sample size. Consequently, in future work, we aim to collect more samples to confirm our findings, based on which we will construct a prognostic risk stratification model to help clinicians screen patients with LNM.

Above all, we put forward a hypothesis that high expression of TM4SF1 was caused by aberrant gene promoter methylation and eventually lead to LNM of PTC patients via the ECM-receptor interaction pathway. Next, we will verify the mechanism of LNM induced by TM4SF1 at cell and animal level.

In conclusion, hyperexpression and hypomethylation of TM4SF1 are associated with LNM in PTC patients and TM4SF1 might be a potential new biomarker for both preoperative evaluation and postoperative follow-up of the high-risk LNM of PTC patients.

**Supplementary information** is available in the online version of the paper.

## References

- [1] HUNDAHL SA, FLEMING ID, FREMGEN AM, MENCK HR. A National Cancer Data Base report on 53,856 cases of thyroid carcinoma treated in the U.S., 1985–1995 [see comments]. *Cancer* 1998; 83: 2638–2648. [https://doi.org/10.1002/\(sici\)1097-0142\(19981215\)83:12<2638::aid-ncr31>3.0.co;2-1](https://doi.org/10.1002/(sici)1097-0142(19981215)83:12<2638::aid-ncr31>3.0.co;2-1)
- [2] KONTUREK A, BARCZYŃSKI M, STOPA M, NOWAK W. Trends in Prevalence of Thyroid Cancer Over Three Decades: A Retrospective Cohort Study of 17,526 Surgical Patients. *World J Surg* 2016; 40: 538–544. <https://doi.org/10.1007/s00268-015-3322-z>
- [3] DAVIES L, WELCH HG. Increasing incidence of thyroid cancer in the United States, 1973–2002. *JAMA* 2006; 295: 2164–2167. <https://doi.org/10.1001/jama.295.18.2164>
- [4] ODA H, MIYAUCHI A, ITO Y, YOSHIOKA K, NAKAYAMA A et al. Incidences of Unfavorable Events in the Management of Low-Risk Papillary Microcarcinoma of the Thyroid by Active Surveillance Versus Immediate Surgery. *Thyroid* 2016; 26: 150–155. <https://doi.org/10.1089/thy.2015.0313>
- [5] HEI H, SONG Y, QIN J. A nomogram predicting contralateral central neck lymph node metastasis for papillary thyroid carcinoma. *J Surg Oncol* 2016; 114: 703–707. <https://doi.org/10.1002/jso.24403>
- [6] HEI H, SONG Y, QIN J. Individual prediction of lateral neck metastasis risk in patients with unifocal papillary thyroid carcinoma. *Eur J Surg Oncol* 2019; 45: 1039–1045. <https://doi.org/10.1016/j.ejso.2019.02.016>
- [7] PATEL K N, YIP L, LUBITZ CC, GRUBBS EG, MILLER BS et al. Executive Summary of the American Association of Endocrine Surgeons Guidelines for the Definitive Surgical Management of Thyroid Disease in Adults. *Ann Surg* 2020; 271: 399–410. <https://doi.org/10.1097/SLA.0000000000003735>
- [8] LEBoulLEUX S, RUBINO C, BAUDIN E, CAILLOU B, HARTL DM et al. Prognostic factors for persistent or recurrent disease of papillary thyroid carcinoma with neck lymph node metastases and/or tumor extension beyond the thyroid capsule at initial diagnosis. *J Clin Endocrinol Metab* 2005; 90: 5723–5729. <https://doi.org/10.1210/jc.2005-0285>
- [9] PATRON V, HITIER M, BEDFERT C, LE CLECH G, JÉ-GOUX F. Occult lymph node metastases increase locoregional recurrence in differentiated thyroid carcinoma. *Ann Otol Rhinol Laryngol* 2012; 121: 283–290. <https://doi.org/10.1177/000348941212100501>
- [10] ADAM MA, PURA J, GOFFREDO P, DINAN MA, REED SD et al. Presence and Number of Lymph Node Metastases Are Associated With Compromised Survival for Patients Younger Than Age 45 Years With Papillary Thyroid Cancer. *J Clin Oncol* 2015; 33: 2370–2375. <https://doi.org/10.1200/JCO.2014.59.8391>
- [11] KOO BS, CHOI EC, YOON YH, KIM DH, KIM EH et al. Predictive factors for ipsilateral or contralateral central lymph node metastasis in unilateral papillary thyroid carcinoma. *Ann Surg* 2009; 249: 840–844. <https://doi.org/10.1097/SLA.0b013e3181a40919>
- [12] EDWARDS CP, FARR AG, MARKEN JS, NELSON A, BAJORATH J et al. Cloning of the murine counterpart of the tumor-associated antigen H-L6: epitope mapping of the human and murine L6 antigens. *Biochemistry* 1995; 34: 12653–12660. <https://doi.org/10.1021/bi00039a022>
- [13] KANEKO R, TSUJI N, KAMAGATA C, ENDOH T, NAKAMURA M et al. Amount of expression of the tumor-associated antigen L6 gene and transmembrane 4 superfamily member 5 gene in gastric cancers and gastric mucosa. *Am J Gastroenterol* 2001; 96: 3457–3458. <https://doi.org/10.1111/j.1572-0241.2001.05355.x>
- [14] MARKEN JS, SCHIEVEN GL, HELLSTRÖM I, HELLSTRÖM KE, ARUFFO A. Cloning and expression of the tumor-associated antigen L6. *Proc Natl Acad Sci U S A* 1992; 89: 3503–3507. <https://doi.org/10.1073/pnas.89.8.3503>
- [15] WRIGHT MD, NI J, RUDY GB. The L6 membrane proteins—a new four-transmembrane superfamily. *Protein Sci* 2000; 9: 1594–1600. <https://doi.org/10.1110/ps.9.8.1594>
- [16] KAO YR, SHIH JY, WEN WC, KO YP, CHEN BM et al. Tumor-associated antigen L6 and the invasion of human lung cancer cells. *Clin Cancer Res* 2003; 9: 2807–2816.
- [17] SHIH SC, ZUKAUSKASA, LI D, LIU G, ANG LH et al. The L6 protein TM4SF1 is critical for endothelial cell function and tumor angiogenesis. *Cancer Res* 2009; 69: 3272–3277. <https://doi.org/10.1158/0008-5472.CAN-08-4886>
- [18] LIN CI, MERLEY A, SCIUTO TE, LI D, DVORAK AM et al. TM4SF1: a new vascular therapeutic target in cancer. *Angiogenesis* 2014; 17: 897–907. <https://doi.org/10.1007/s10456-014-9437-2>
- [19] CLOUGH E, BARRETT T. The Gene Expression Omnibus Database. *Methods Mol Biol* 2016; 1418: 93–110. [https://doi.org/10.1007/978-1-4939-3578-9\\_5](https://doi.org/10.1007/978-1-4939-3578-9_5)
- [20] PATHAN M, KEERTHIKUMAR S, CHISANGA D, ALESSANDRO R, ANG CS et al. A novel community driven software for functional enrichment analysis of extracellular vesicles data. *J Extracell Vesicles* 2017; 6: 1321455. <https://doi.org/10.1080/20013078.2017.1321455>
- [21] CHANDRASHEKAR DS, BASHEL B, BALASUBRAMANYA SAH, CREIGHTON CJ, PONCE-RODRIGUEZ I et al. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* 2017; 19: 649–658. <https://doi.org/10.1016/j.neo.2017.05.002>

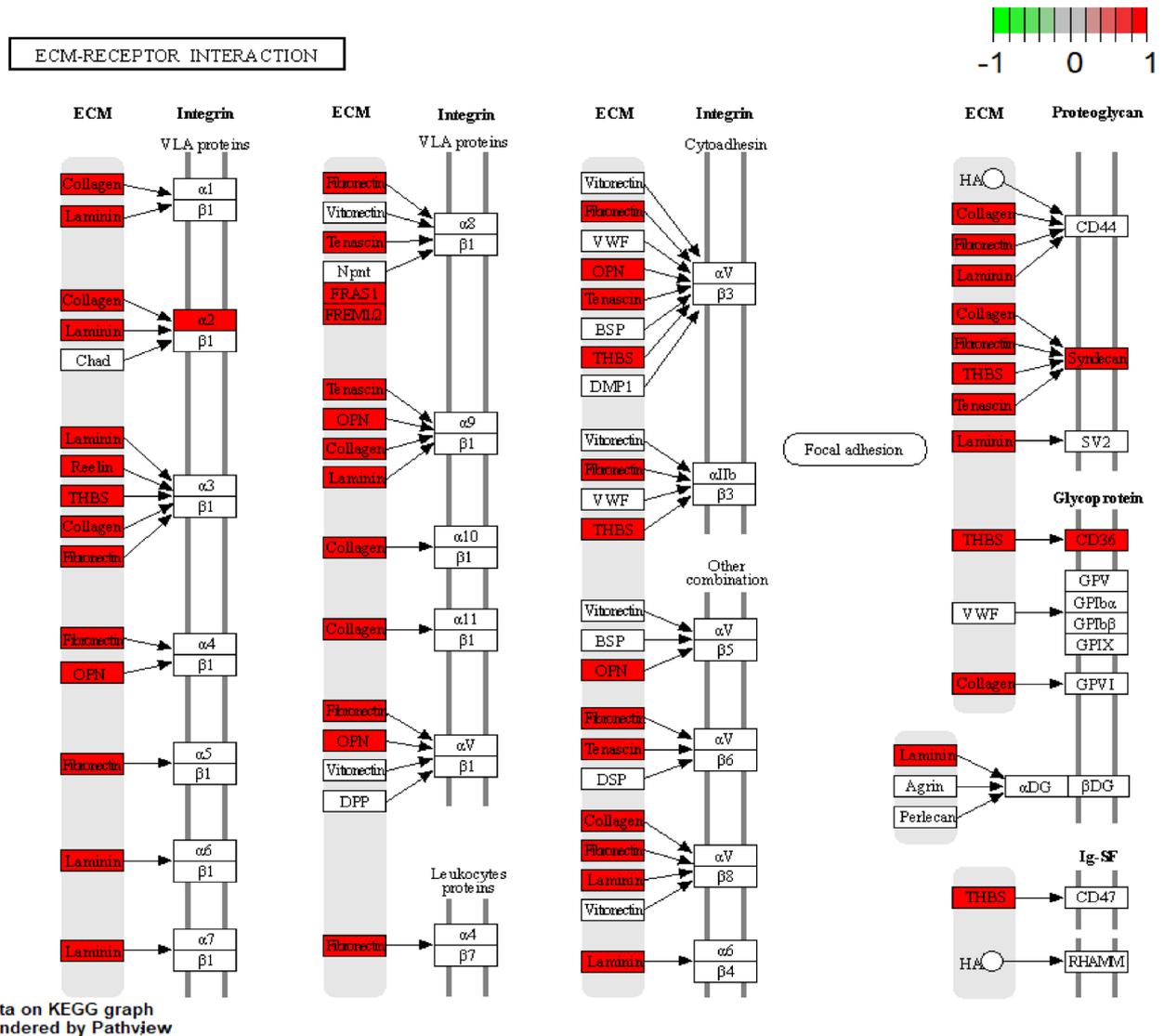
- [22] KOCH A, JESCHKE J, VAN CRIEKINGE W, VAN ENGE- LAND M, DE MEYER T. MEXPRESS update 2019. *Nucleic Acids Res* 2019; 47: W561–W565. <https://doi.org/10.1093/nar/gkz445>
- [23] KANEHISA M, FURUMICHI M, TANABE M, SATO Y, MORISHIMA K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017; 45: D353–D361. <https://doi.org/10.1093/nar/gkw1092>
- [24] TANG Z, LI C, KANG B, GAO G, LI C et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* 2017; 45: W98–W102. <https://doi.org/10.1093/nar/gkx247>
- [25] MOUW JK, OU G, WEAVER VM. Extracellular matrix assembly: a multiscale deconstruction. *Nat Rev Mol Cell Biol* 2014; 15: 771–785. <https://doi.org/10.1038/nrm3902>
- [26] KAI F, DRAIN AP, WEAVER VM. The Extracellular Matrix Modulates the Metastatic Journey. *Dev Cell* 2019; 49: 332–346. <https://doi.org/10.1016/j.devcel.2019.03.026>
- [27] CARON NR, CLARK OH. Papillary thyroid cancer. *Curr Treat Options Oncol* 2006; 7: 309–319. <https://doi.org/10.1007/s11864-006-0040-7>
- [28] FOLTZ G, RYU GY, YOON JG, NELSON T, FAHEY J et al. Genome-wide analysis of epigenetic silencing identifies BEX1 and BEX2 as candidate tumor suppressor genes in malignant glioma. *Cancer Res* 2006; 66: 6665–6674. <https://doi.org/10.1158/0008-5472.CAN-05-4453>
- [29] LEE CH, WONG TS, CHAN JY, LU SC, LIN P et al. Epigenetic regulation of the X-linked tumour suppressors BEX1 and LDOC1 in oral squamous cell carcinoma. *J Pathol* 2013; 230: 298–309. <https://doi.org/10.1002/path.4173>
- [30] JIN Z, LIU Y. DNA methylation in human diseases. *Genes Dis* 2018; 5: 1–8. <https://doi.org/10.1016/j.gendis.2018.01.002>
- [31] XING M. Molecular pathogenesis and mechanisms of thyroid cancer. *Nat Rev Cancer* 2013; 13: 184–199. <https://doi.org/10.1038/nrc3431>
- [32] HANSEN KD, TIMP W, BRAVO HC, SABUNCIYAN S, LANGMEAD B et al. Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* 2011; 43: 768–775. <https://doi.org/10.1038/ng.865>

[https://doi.org/10.4149/neo\\_2021\\_210819N1191](https://doi.org/10.4149/neo_2021_210819N1191)

# Hyper-expression and hypomethylation of TM4SF1 are associated with lymph node metastases in papillary thyroid carcinoma patients

Kun WANG<sup>1,2</sup>, Juan ZHENG<sup>3</sup>, Zhen WU<sup>4,5</sup>, Ju-Gao FANG<sup>4</sup>, Jun-Yu ZHAO<sup>6</sup>, Jin-Ming YAO<sup>6</sup>, Jian-Jun DONG<sup>7</sup>, Jie BAI<sup>2</sup>, Lin LIAO<sup>1,6,\*</sup>

## Supplementary Information



Supplementary Figure S1. ECM-receptor interaction pathway.