



# Integrated analysis of tumor microenvironment features to establish a diagnostic model for papillary thyroid cancer using bulk and single-cell RNA sequencing technology

Yizeng Wang<sup>1</sup> · Wenbin Song<sup>1</sup> · Yingxi Li<sup>2</sup> · Zhaoyi Liu<sup>1</sup> · Ke Zhao<sup>1</sup> · Lanning Jia<sup>1</sup> · Xiaoning Wang<sup>1</sup> · Ruoyu Jiang<sup>1</sup> · Yao Tian<sup>1</sup> · Xianghui He<sup>1</sup>

Received: 8 July 2023 / Accepted: 12 September 2023 / Published online: 21 September 2023

© The Author(s) 2023

## Abstract

**Background** Characterizing tumor microenvironment using single-cell RNA sequencing has been a promising strategy for cancer diagnosis and treatment. However, a few studies have focused on diagnosing papillary thyroid cancer (PTC) through this technology. Therefore, our study explored tumor microenvironment (TME) features and identified potential biomarkers to establish a diagnostic model for papillary thyroid cancer.

**Methods** The cell types were identified using the markers from the CellMarker database and published research. The CellChat package was conducted to analyze the cell–cell interaction. The SCEVAN package was used to identify malignant thyroid cells. The SCP package was used to perform multiple single-cell downstream analyses, such as GSEA analysis, enrichment analysis, pseudotime trajectory analysis, and differential expression analysis. The diagnostic model of PTC was estimated using the calibration curves, receiver operating characteristic curves, and decision curve analysis. RT-qPCR was performed to validate the expression of candidate genes in human papillary thyroid samples.

**Results** Eight cell types were identified in the scRNA-seq dataset by published cell markers. Extensive cell–cell interactions like FN1/ITGB1 existed in PTC tissues. We identified 26 critical genes related to PTC progression. Further, eight subgroups of PTC tumor cells were identified and exhibited high heterogeneity. The MDK/LRP1, MDK/ALK, GAS6/MERTK, and GAS6/AXL were identified as potential ligand-receptor pairs involved in the interactions between fibroblasts/endothelial cells and tumor cells. Eventually, the diagnostic model constructed by TRPC5, TENM1, NELL2, DMD, SLC35F3, and AUTS2 showed a good efficiency for distinguishing the PTC and normal tissues.

**Conclusions** Our study comprehensively characterized the tumor microenvironment in papillary thyroid cancer. Through combined analysis with bulk RNA-seq, six potential diagnostic biomarkers were identified and validated. The diagnostic model we constructed was a promising tool for PTC diagnosis. Our findings provide new insights into the heterogeneity of thyroid cancer and the theoretical basis for diagnosing thyroid cancer.

**Keywords** Papillary thyroid cancer · Biomarker · Diagnostic model · Single-cell RNA sequencing · Tumor microenvironment

## Abbreviations

THCA	Thyroid cancer	CAFs	Cancer-associated fibroblasts
PTC	Papillary thyroid carcinoma	GO	Gene Ontology
TCGA	The Cancer Genome Atlas	KEGG	Kyoto Encyclopedia of Genes and Genomes
PCA	Principal component analysis	GSEA	Gene set enrichment analysis
GEO	Gene Expression Omnibus	NT	Normal tissue
ScRNA	Single-cell RNA	TME	Tumor microenvironment

Yizeng Wang, Wenbin Song, and Yingxi Li have contributed equally to this work.

Extended author information available on the last page of the article

## Introduction

Thyroid cancer (THCA) is one of the most common endocrine tumors globally (Xia et al. 2022; Siegel et al. 2023). Papillary thyroid cancer (PTC)—a pathological type of THCA—accounts for approximately 90% of THCA patients (Fagin and Wells 2016; Chen et al. 2023). Ultrasound is currently the standard technique for evaluating the characteristics of thyroid nodules. Existing risk stratification systems classify the risk of malignant tumors according to the ultrasonic findings of thyroid nodules (Kobaly et al. 2022). Although 10–15% of nodules eventually prove malignant, 5% of patients have distant metastases at diagnosis (Alexander and Cibas 2022). About 15% of patients treated with surgery relapse during postoperative follow-up, leading to a poor prognosis (Ito et al. 2018; Alexander and Cibas 2022; Fallahi et al. 2022). Even after postoperative radioactive iodine therapy, some patients develop refractory PTC. With a better understanding of the molecular progression of THCA, several drugs have been developed for inhibiting oncogenic kinases or signaling kinases (RET/PTC, *BRAF* (V600E)), like those related to vascular endothelial growth factor receptor platelet-derived growth factor receptor and platelet-derived growth factor receptor (Fallahi et al. 2022). *BRAF* (V600E) mutations have been confirmed to be associated with poor prognosis of PTC patients (Costa et al. 2008; Song et al. 2021). Unfortunately, PTC patients without *BRAF* mutation could not benefit from these newly developed drugs. The effect of *BRAF* inhibitors in non-*BRAF* mutated cancers has reportedly been controversial (Agianian and Gavathiotis 2018). Therefore, future research must delve into the molecular heterogeneity of PTC to find novel diagnostic biomarkers and provide patients with individualized treatment.

With the development of sequencing technology, researchers are able to observe the cellular level changes in tumor tissue through single-cell RNA sequencing (scRNA-seq) technology, which brings new strategies for tumor diagnosis and treatment (Hwang et al. 2018). The tumor microenvironment, a daedal ecosystem comprise stromal, epithelial, and immune cells (Xiao and Yu 2021). Different immune infiltrating state in the tumor microenvironment is related to distinct prognostic outcomes in tumor patients. For instance, the presence of tumor-associated fibroblast is correlated with worse outcomes in gastric, bladder, and breast cancers (Bartoschek et al. 2018; Chen et al. 2020; Li et al. 2022). New markers and therapeutic targets have been developed by further exploring the tumor microenvironment. Sui et al. (2023) revealed the role of the *CCL18/PITPNM3* ligand-receptor pair in the interaction between tumor cells and macrophages by single-cell data of esophageal squamous cell carcinoma, and the

receptor could be used as a potential therapeutic target for esophageal cancer. Ma et al. (2020) found the heterogeneity of prostate tumor cells, and *HPN* could be used as an early-stage diagnostic marker for prostate cancer by analyzing single-cell data of prostate cancer. Therefore, a deep understanding of the tumor microenvironment of PTC helps to elucidate the underlying mechanisms of tumor progression and occurrence, as well as to search for new biomarkers and potential therapeutic targets.

The present study extensively characterized the tumor microenvironment of PTC and established a diagnostic model-based six genes. Our findings revealed extensive *FN1/ITGB1* communication between the T/NK and other cells in PTC compared to normal thyroid tissue. Furthermore, we found dynamic changes in the evolution of thyroid cells into tumor cells and a high degree of heterogeneity among tumor cells. *MDK/ALK/ALP* and *GAS/MERTK/AXL* interactions between tumor cells and endothelial cells/fibroblasts may be potential therapeutic targets for PTC. Finally, we established a diagnostic model with good diagnostic efficacy for PTC by combining bulk RNA-seq. Our findings provided novel insights into the diagnosis of PTC.

## Materials and methods

### Data collection and data preprocessing

A total of 665 samples were included in the present study: 7 scRNA-seq samples (1 normal thyroid sample and 3 pairs of bilateral PTC) from the Gene Expression Omnibus (GEO) cohort (GSE191288, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE191288>); 564 RNA-seq data (505 PTC and 59 normal samples) from The Cancer Genome Atlas (TCGA) cohort; 94 RNA-seq data (49 PTC and 45 normal samples) from GEO cohort. (GSE33630, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE33630>). The gene expression profiles of the TCGA dataset were downloaded using the TCGAblinks package (v2.26.0), then converted to TPM format and standardized with log<sub>2</sub>.

The “Seurat” R package (v4.3.0) was used to conduct quality control procedures and downstream bioinformatics analyses for scRNA-seq datasets. We used the following criteria to obtain high-quality cells and filter out low-quality cells: the proportion of mitochondrial genes counts  $\leq 15\%$ ; the proportion of erythrocyte genes counts  $\leq 3\%$ ; UMIs  $\geq 500$ ; genes detected per cell  $\geq 200$ . The Doublet-Finder (McGinnis et al. 2019) package was used to remove the doublets in each sample. After removing the doublets, 28,587 cells were included in further research. Subsequently, we normalized the scRNA-seq expression through the glmGamPoi (Ahlmann-Eltze and Huber 2021) package. The top 3000 highly variable genes were identified using the

SCTransform method and then used to calculate principal components. The TSNE and UMAP methods in Seurat were performed for cell clustering. The Clustree package was utilized to find the optimal cluster resolution. The markers used for cell identity were obtained from published research and the CellMarker database (Zhang et al. 2019; Pu et al. 2021).

### Cell–cell interaction analysis

The CellChat (Jin et al. 2021) package was used to evaluate the difference in cell–cell interaction between the normal and THCA samples. The two data sets were normalized separately, the high-variable genes were identified, and PCA analysis was performed in each of the two gene sets using the high-variable genes. Anchors were identified with the findintegrationanchors() function, and the IntegrateData() function was used to combine two data sets. Additionally, the cell–cell interaction analysis among 16 cell types was also performed according to standard procedures.  $P$  value  $< 0.05$  was considered statistically significant.

### CNV analysis

SCEVAN—an R package—can infer non-malignant and malignant cells in the tumor microenvironment by calculating the raw count matrix of scRNA data. The thyrocyte cell expression matrix of six tumor samples (N1L, N1R, N2L, N2R, N3L, and N3R) were extracted and performed copy number variation (CNV) analysis to determine malignant thyrocyte cells with the SCEVAN package.

### Single-cell downstream analysis

The SCP package (<https://github.com/zhanghao-njmu/SCP>) provides a comprehensive set of tools for single-cell data downstream analysis. The present study used SCP to perform multiple single-cell downstream analyses, such as pseudo-time trajectory analysis, enrichment analysis, GSEA analysis, and differential expression analysis. In detail, the trajectory analysis was performed by RunSlingshot() function in SCP to infer the evolution of thyrocytes into tumor cells. The tumor cells were extracted and regrouped into eight groups. Enrichment analysis between tumor groups was performed using the RunGSEA() function in SCP, and an adjusted  $P$  value  $< 0.05$  was considered statistically significant. Furthermore, RunDEtest() function in SCP was used to identify the over-expression genes in eight distinct groups of tumor cells. Genes with  $\log_2FC > 1$  and  $P$  value  $< 0.05$  were considered over-expressed genes. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed to annotate the biological processes over-expressed genes are involved in.

### Clinical relevance

Differential expression analysis between thyrocytes and tumor cells was performed using the RunDEtest() function in SCP. We obtained 22 markers with  $\log_2FC > 1$  and  $P$  value  $< 0.05$ . A total of 1186 genes up-regulated in PTC tissue were calculated by edgeR package. In detail, differential methods were set as ANOVA, and genes with  $\log_2FC > 1$  and  $P$  value  $< 0.05$  were considered up-regulated genes. The intersection of the aforementioned genes led to the identification of eleven genes. Boruta and LASSO method were utilized to identify the most important genes.

Logistic regression was used to construct a diagnostic model with important genes to better predict the thyrocyte tissue type. Receiver operating characteristic (ROC) curves were utilized to assess the discriminative performance. The calibration curves were applied to estimate the predictive accuracy of the model using the bootstrap method with 1000 re-samplings. To better explore model fitting, Hosmer–Lemeshow (HL) tests were performed, and  $P$  value  $> 0.05$  was recognized as a good model fitting. The clinical applicability of the diagnostic model was estimated by the decision curve analysis (DCA). The TCGA dataset was used as training cohort, and GEO dataset was used as testing cohort.

### RT-qPCR

Total RNA was extracted using TRIzol Reagent (Invitrogen, Carlsbad, CA, USA) from frozen THCA tissue and corresponding control tissue resected surgically according to the manufacturer's recommendations. RNA quality and concentration were measured using the NanoDrop 2000 spectrophotometer (Thermo Scientific, USA). Real-time polymerase chain reaction (PCR) reactions were performed using SYBR Green PCR Master Mix (TransGen Biotech, Beijing, China) on the 7500 Real-Time PCR System (Applied Biosystems, Waltham, MA, USA). The specific sequences of primers are listed in Additional file 1: Table S1.

### Statistical analyses

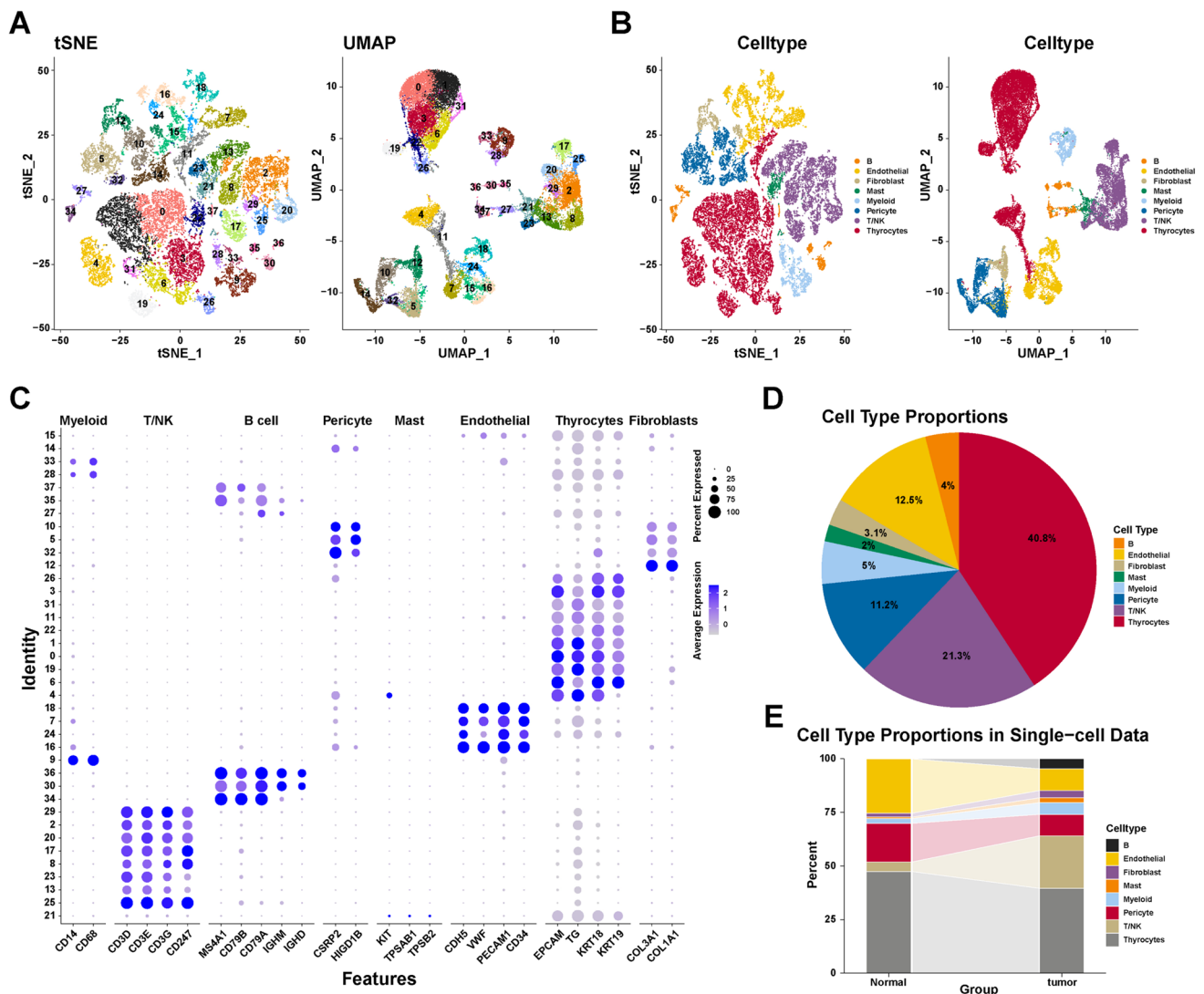
The RT-qPCR results were statistically analyzed using Prism 8 (Graph pad Software, CA) and are presented as mean  $\pm$  standard deviation (SD) for at least three individual experiments. The statistical significance of differences was determined with the unpaired, two-tailed student  $t$  test, and  $P < 0.05$  was considered statistically significant. Other statistical analyses in this study were performed using R studio software 4.2.2

## Results

### Identification of cell types

A total of 28,587 cells were obtained after quality control procedures through Seurat. All cells in the scRNA-seq dataset were clustered into 38 clusters using PCA with a resolution of 1. As shown in Fig. 1A, two-dimensionality reduction methods—TSNE and UMAP—revealed that different cell clusters were separated. We further annotated the cell types with cell markers from the Cell Marker database and published research. All cells were

further annotated into eight cell types: T/NK cells, B cells, endothelial cells, myeloid cells, fibroblasts, mast cells, pericytes, and thyrocytes (Fig. 1B). The used cell markers are shown in Fig. 1C. For example, *CD3D*, *CD3E*, *CD3G*, and *CD247* are markers of T/NK cells, and nine-cell clusters (2, 8, 13, 17, 20, 21, 23, 25, and 29) are identified as T/NK cells. These results indicated that the used markers could clearly distinguish different cell types. Furthermore, Fig. 1D shows the proportion of different cell types: thyrocytes (40.8%), T/NK cells (21.3%), pericytes (11.2%), myeloid cells (5%), endothelial cells (12.5%), fibroblasts (3.1%), B cells (4%), and mast cells (2%). Figure 1E depicts the proportion of cell types in normal and tumor



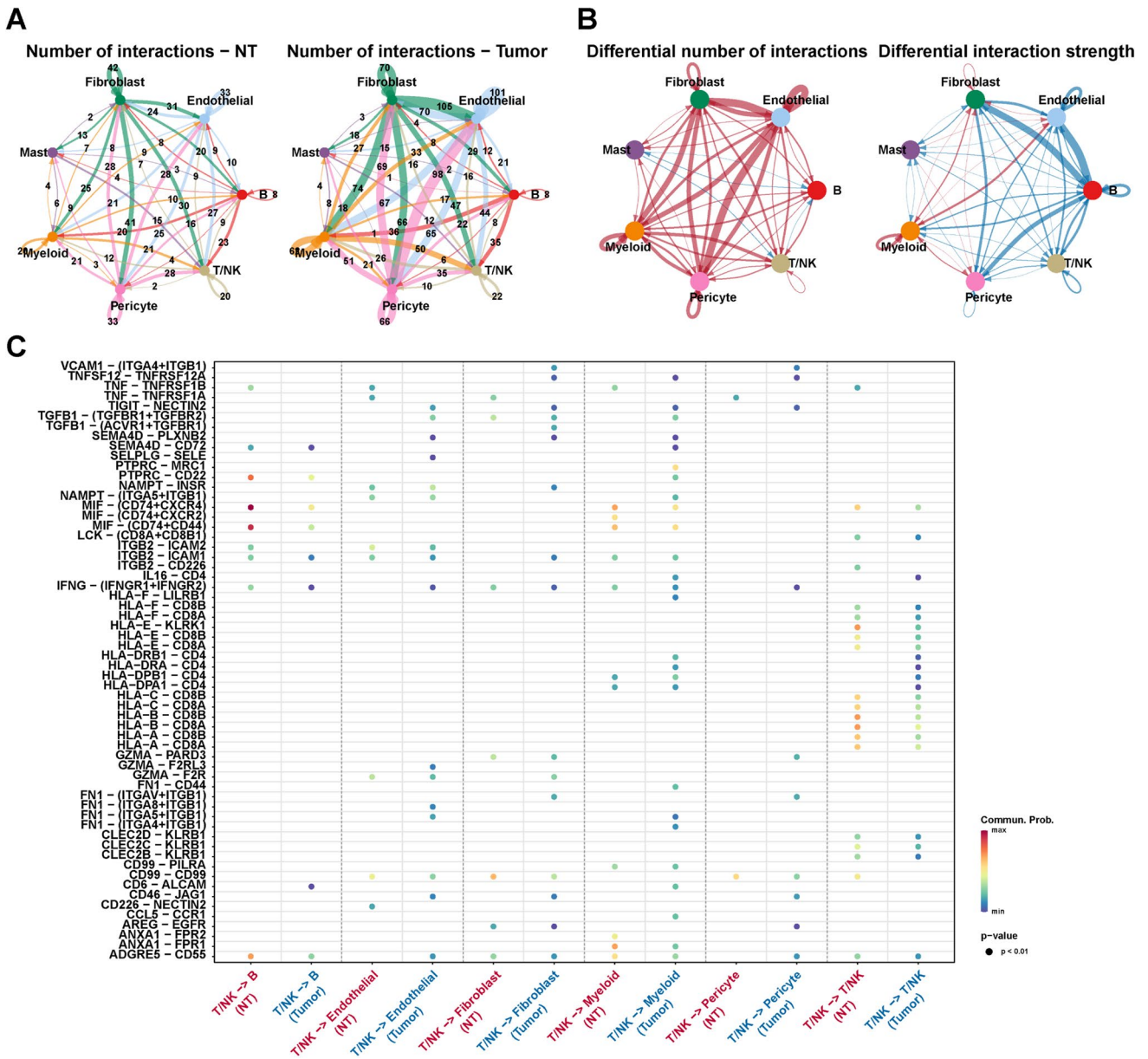
**Fig. 1** Distinct cell types in PTC were identified through single-cell sequencing. The cell clusters (**A**) and cell types (**B**) in PTC tissue demonstrated using the uniform manifold approximation and projection (UMAP) and t-distributed stochastic neighbor embedding (TSNE) plots according to their featured gene expression profiles. **C**

Dot plot displaying the expression level of marker genes for annotating the cell types. **D** The cell type portions of the whole scRNA-seq dataset **E** The cell type portions of normal and tumor samples in scRNA-seq dataset

samples. The proportion of non-epithelial cells in tumor samples is higher than that in normal samples, indicating a more complex microenvironment of tumor samples. These results revealed the characteristics of the data set and the differences in cell proportions between the samples. Taken together, we identified eight main cell types for further exploration.

### Communication network difference between the normal and thyroid cancer tissue

Cell communication analysis can infer cell–cell interactions to further explore the changes in the tumor microenvironment. To analyze the changes in cell–cell interactions between normal and PTC tissues, we performed the cell communication analysis among seven cell types, excluding thyrocytes, using CellChat. Figure 2A depicts the number



**Fig. 2** Cell–cell interaction differences in tumor and normal tissues. **A** Interaction net count plot of non-tumor (NT) and tumor tissues. The thicker the line represented, the more the number of interactions between the two cell types. Number represents the number of interactions. **B** Interaction net count and strength plot of tumor tissues compared with those of NT tissue. The thicker the line represented,

the more the number of interactions, and the stronger the interaction weights/strength between the two cell types. Red represents tumor tissues. Blue represents NT tissue. **C** Dot plot exhibited the differences in ligand-receptor pairs between the T/NK and other cells in NT and tumor tissues

of interactions among seven cell types between normal (NT) and tumor samples, with cell–cell interactions in tumor tissues more than that in normal tissue. Subsequently, we compared the number and intensity of interactions between tumors and normal tissue (Fig. 2B). Compared to the normal sample, the number and intensity of interactions among endothelial cells, myeloid cells, and fibroblasts were higher in tumor tissues (N1L, N1R, N2L, N2R, N3L, N3R), suggesting the crucial role of three cell types in the formation and development of the PTC tumor microenvironment.

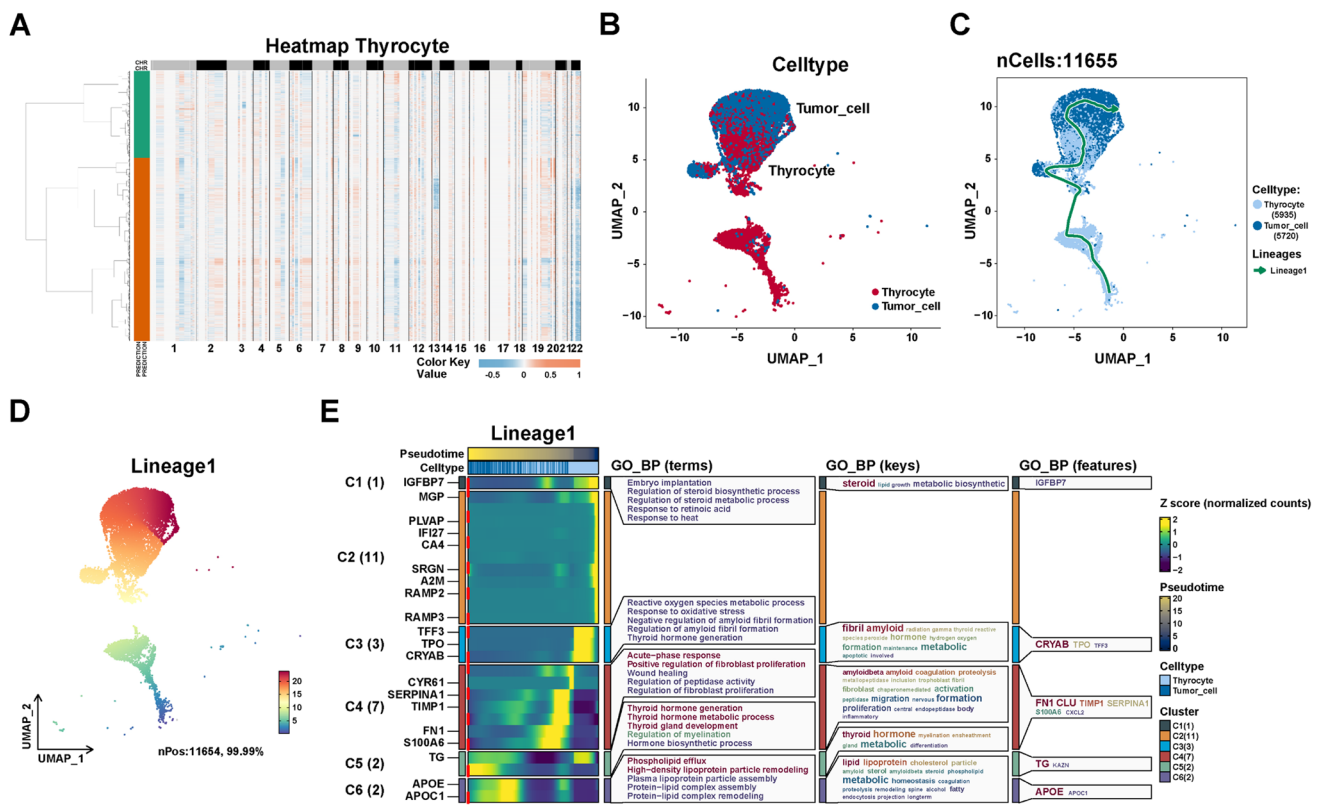
Considering the significance of T/NK cells in the TME, we further analyzed the interactions between T/NK and other cells, including themselves. For ligand–receptor pairs such as *FN1-ITGB1*, the probability of T/NK cells interacting with endothelial cells, fibroblasts, and myelocytes in PTC tissue was significantly higher than that in non-tumor tissue (Fig. 2C). This result suggested that *FN1/ITGB1* could play a critical role in the tumor microenvironment of PTC.

These findings reveal the potential role of extensive cellular communication in promoting tumor microenvironment formation in PTC tissue.

### The cell trajectory of PTC epithelial cells

As not all epithelial cells in tumor tissues are malignant, SCEVAN was used to infer the benign and malignant nature of epithelial cells in tumor tissues. SCEVAN was used to analyze copy number variation (CNV) on 9577 thyrocytes from tumor tissues (N1L, N1R, N2L, N2R, N3L, and N3R). As shown in Fig. 3A, 5720 tumor cells were identified with abnormal levels of CNV compared to thyrocytes. Subsequently, we extracted 11,655 thyrocytes and tumor cells in 7 samples (1 NT and 6 tumor samples) for further analysis (Fig. 3B).

To investigate the dynamic development of epithelial cells in PTC microenvironment, we performed the cell



**Fig. 3** Reconstructing the pseudotime trajectory of tumor cells using thyrocytes and tumor cells and identifying genes varied during the trajectory. **A** Heatmap of CNV levels in thyrocytes. Green represents normal thyrocytes, and yellow represents malignant thyroid (tumor) cells in legend. Color key from deep blue to yellow indicates relative CNV levels from low to high. **B** Cell type assignment following UMAP-based visualization of expression differences for 11,655 single thyrocytes (normal and malignant thyrocytes) from 7 samples in the scRNA-seq dataset. **C** Cell trajectory of normal thyrocytes and tumor cells was generated using a slingshot in SCP. Lineages repre-

sented cell trajectory directions. **D** Pseudo-time is colored in a gradient from blue to red. The start of pseudo-time is indicated by blue, whereas the end of pseudo-time by red. **E** The differentially expressed genes (DEGs) with expression levels that changed the most over the pseudo-time trajectory were divided into six clusters based on their expression trend, and the representative processes of each cluster are shown. Color key from deep blue to yellow indicates relative expression levels of the DEGs from low to high. The numbers in parentheses after the cluster represent the trees of the gene

trajectory analysis with the slingshot method in the SCP package to infer the trajectory from thyrocytes to tumor cells. As shown in Figs. 3C, D, thyrocytes were shown at the beginning of the trajectory, whereas tumor cells were located in the end of trajectory. To further explore the critical genes and biological progress in PTC progression, we analyzed the dynamic changes of genes in the trajectory “lineage 1”. As shown in Fig. 3E, the most important genes involved in PTC progression were identified: *IGFBP7*, *MGP*, *PLVAP*, *IFI27*, *CA4*, *SRGN*, *A2M*, *RAMP3*, *RAMP4*, *TFF3*, *TPO*, *CRYAB*, *CYR61*, *SERPINA1*, *TIMP1*, *FN1*, *S100A6*, *TG*, *APOE*, and *APOC1*. Additionally, the biological progress, including dysregulation of steroid metabolic process, dysregulation of thyroid hormone metabolic process, and positive regulation of fibroblast proliferation, were potentially involved in PTC progression. These findings reveal the dynamics of the biological processes that occur on the trajectory from thyrocytes to tumor cells. Furthermore, these genes closely related to tumor cell development may serve as potential biomarkers for PTC progression.

### The heterogeneity between thyroid tumor cells

To further explore the heterogeneity among thyroid tumor cells, we re-clustered 5720 tumor cells using the same procedures and the SCTransform method. As shown in Fig. 4A, B, tumor cells were divided into eight groups with PCA. We performed gene set enrichment analysis (GSEA) based on each group compared to other groups. Compared with other groups, subgroup 0 of tumor cells was related to enzyme-linked receptor protein signaling pathway, organ growth, and DNA-templated transcription elongation (Fig. 4C). Subgroup 1 of tumor cells was mainly involved in tube morphogenesis, angiogenesis, blood vessel development, and morphogenesis, suggesting that this subgroup may be involved in angiogenesis in the tumor microenvironment. Subgroups 2 and 3 were associated with cytoplasmic translation, oxidative phosphorylation, and ribosome biogenesis compared with other groups (Fig. 4C). Similarly, Subgroups 4 and 5 were correlated with cell adhesion, cell motility, and cell migration (Fig. 4C), indicating the potential role of Subgroups 4 and 5 in tumor metastasis. In addition, Subgroups 6 and 7 were also involved in different biological processes (Fig. 4C). These findings indicate the characteristics of tumor subgroups at the level of the biological processes involved and the role of each subgroup in the TME.

We further analyzed the differential genes of each subpopulation compared to other subpopulations (Fig. 4D). The genes with  $\log_{2}FC > 1$  and  $P$  value  $< 0.05$  were considered markers of each group (Table S2). As shown in Fig. 4E, the up-regulated genes of subgroup 0 were related to purine metabolism. The up-regulated genes of Subgroups 1 and 2 were involved in estrogen signaling pathways. Additionally,

antigen processing and presentation were up-regulated in subgroups 2 and 5. Additionally, subgroup 5 was also associated with autoimmune thyroid disease. These findings reveal the heterogeneity of thyroid tumor cells, including intercellular metabolism, immunity, and tumor-related signaling pathways.

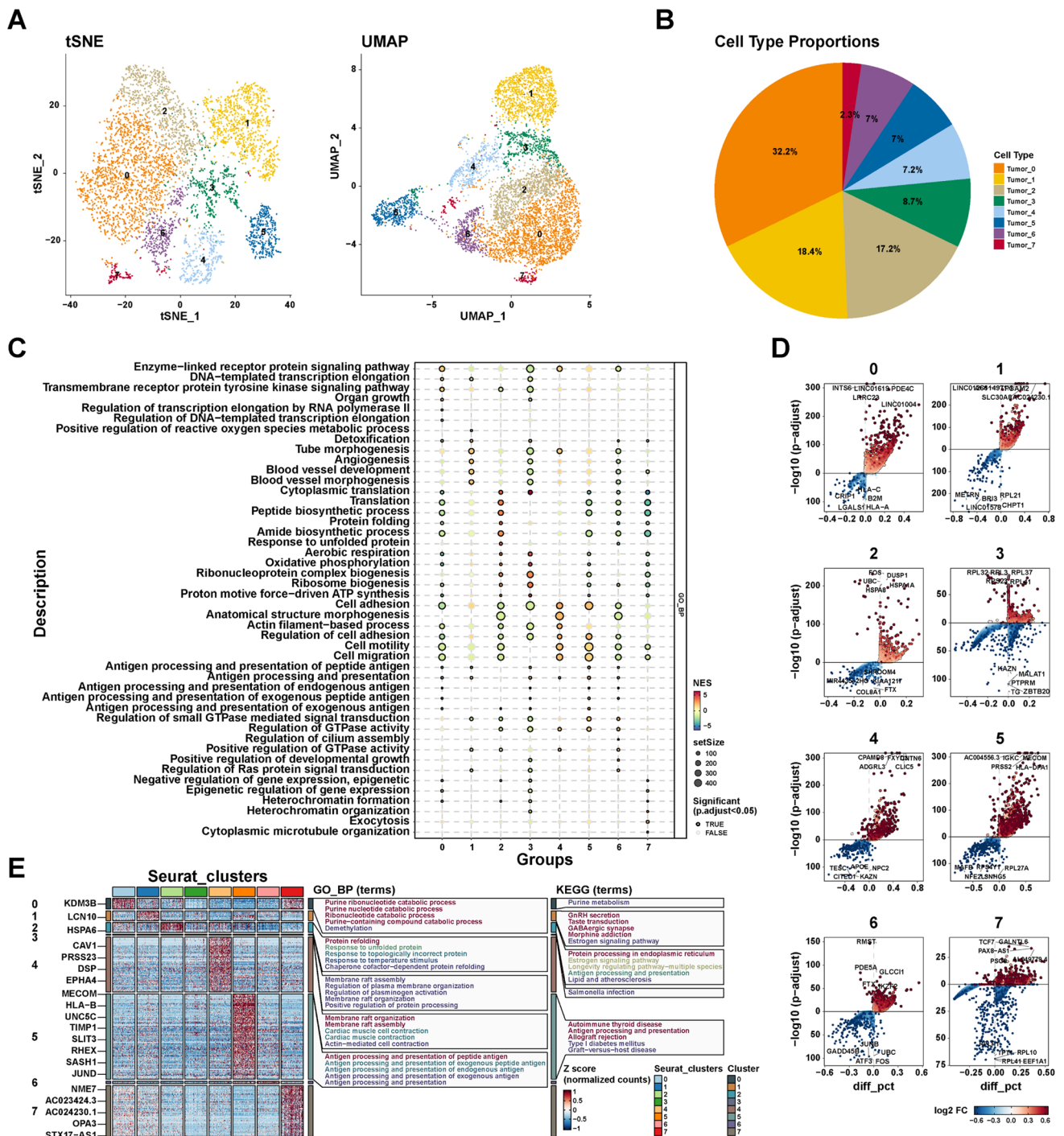
### The crosstalk of tumor cells with other cells in the TME

To further explore the role of tumor cells in the TME, we analyzed all cell–cell interactions using CellChat. As shown in Fig. 5A, the interaction strength between tumor cell subgroups 0–7 and endothelial cells was highest, implying that the endothelial cells may be crucial in promoting tumor cell growth. Additionally, the interaction strength of fibroblasts on tumor cells was significantly enhanced compared with thyrocytes (Fig. 5A). Therefore, we further analyzed the ligand–receptor pair interaction between endothelial cells, thyrocytes, and tumor cells. Ligand–receptor pairs, such as *PROS1-AXL*, *NAMPT-INSR*, *MDK-LRP1*, *MDK-ALK*, *GAS6-MERTK*, and *GAS6-AXL*, between endothelial and tumor cells exhibited a higher communication possibility than that between endothelial cells and thyrocytes (Fig. 5B). These results revealed the potential interactions between tumor cells and endothelial cells in PTC.

Notably, *MDK/LRP1*, *MDK/ALK*, *GAS6/MERTK*, and *GAS6/AXL* between fibroblasts and tumor cells exhibited a higher communication possibility than that between fibroblasts and thyrocytes (Fig. 5C). In view of our previous findings in cell trajectory analysis that genes that positively regulate fibroblast growth were up-regulated in tumor cells, we analyzed the interactions of thyroid cells and tumor cells on fibroblasts. The probability of interaction between tumor cells and fibroblast receptor–ligand pairs was higher than that of thyroid cells, especially in *TGFBI/ACVRI/TGFBR1*, *FGF18/FGFR1*, *BTC/EGFR*, *BMP8A/BMPRIA/BMPR2*, and *BMP8A/BMPRIA/BMPR2* (Fig. 5D). Consistent with the results of cell trajectory analysis, these results suggested that tumor cells could also have a regulatory effect on fibroblast growth via the aforementioned ligand–receptor pairs. Overall, these results revealed the potential role of endothelial cells and fibroblasts in regulating tumor cell growth in the tumor microenvironment.

### Identification of critical diagnostic genes for PTC

To identify the critical diagnostic genes for PTC, we performed the differential expression gene analysis between thyrocytes and tumor cells in the scRNA-seq dataset, and 22 markers of tumor cells were identified (Table S3). Next, we performed the differential expression gene analysis between TCGA-THCA (PTC samples)

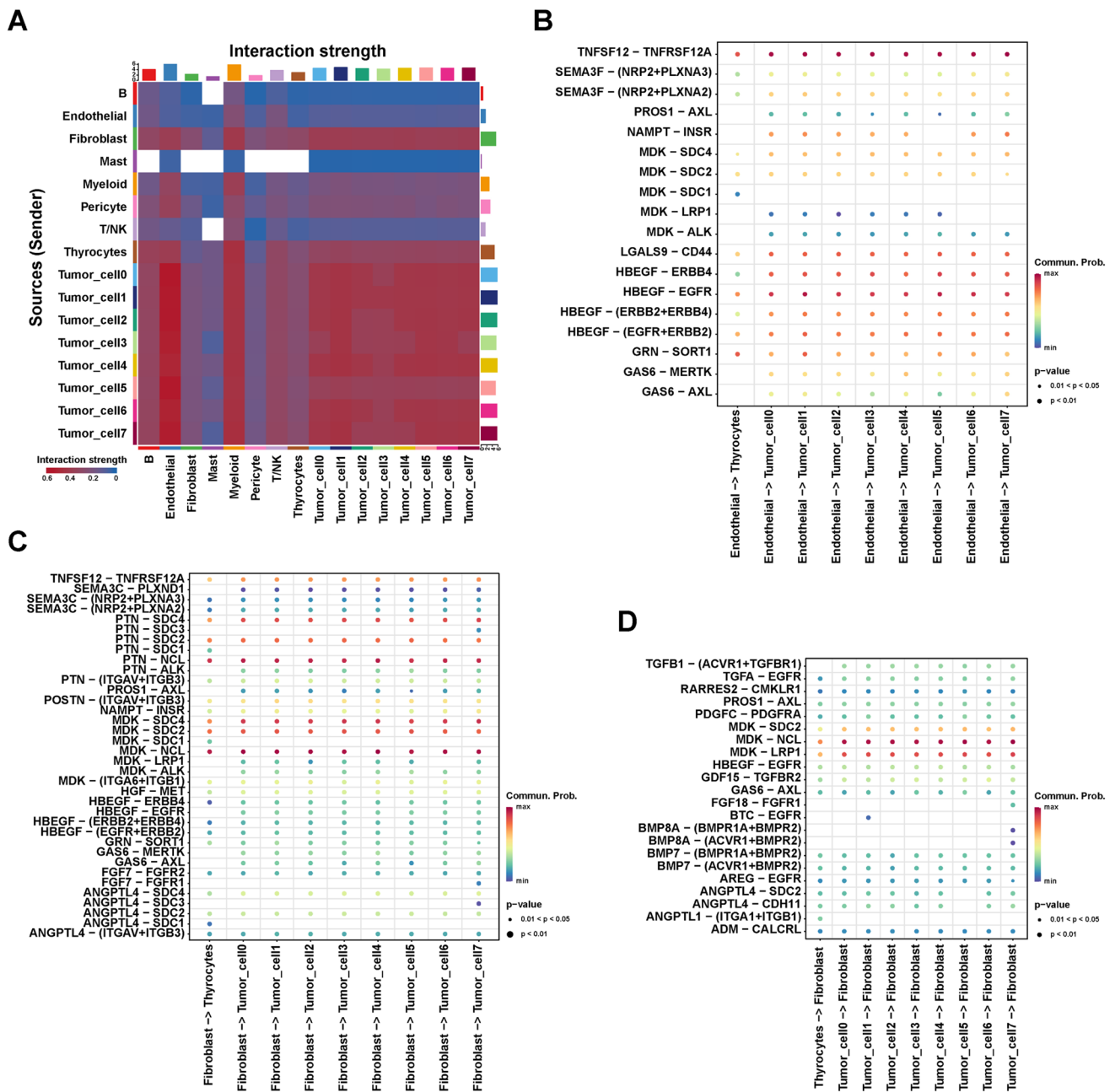


**Fig. 4** Subgroups in tumor cells were sub-clustered using PCA. **A** eight subgroups generated from tumor cells are demonstrated using tSNE and UMAP. **B** Statistics of cell percentage of each subgroup in tumor cells. **C** Dot plot showing the representative biological processes enriched in each subgroup. **D** The volcano plots show differential expressed genes of each subgroup. The horizontal axis represents

$\log_2$ -fold change of genes. The vertical axis represents  $-\log_{10}(P\text{ adjust value})$ . **E** The heatmap shows up-regulated genes of each subgroup, and the representative processes and KEGG pathways of each cluster are shown. Color from blue to red indicates relative expression levels of the genes from low to high

and normal samples, and 1186 up-regulated genes in PTC were obtained (Table S4). After intersection of these two gene sets, we firstly identified eleven genes (*TRPC5*,

*TENM1*, *PDZRN4*, *NELL2*, *FRMD3*, *APOE*, *ARMCX3*, *DMD*, *APOC1*, *SLC35F3*, *AUTS2*) highly expressed at both the tumor cell and tissue levels (Fig. 6A). Boruta

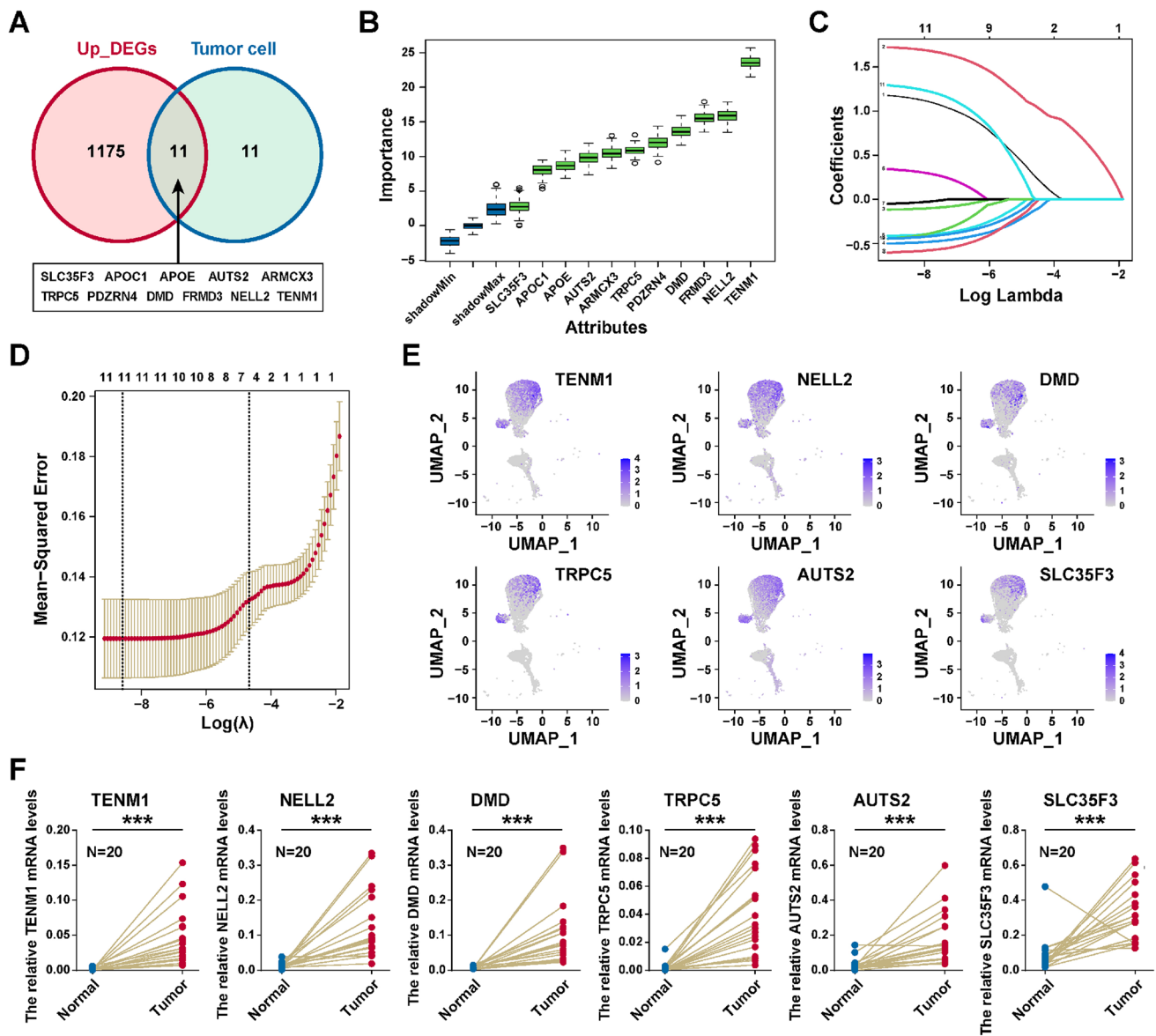


**Fig. 5** The crosstalk between tumor cells and other cells. **A** Heatmap shows the interaction strength among all cells in PTC. Color from blue to red indicates the interaction strength of two cell types from low to high. Blank space represents no significance. Dot plots show potential ligand-receptor pairs for endothelial interaction with tumor

cells (**B**), fibroblast interaction with tumor cells (**C**), and tumor cell interaction with fibroblast (**D**). Color from blue to red indicates relative interaction probability from low to high. Blank space represents no significance

package was used to identify importance of genes, and all eleven genes were confirmed as critical genes (Fig. 6B). We next performed LASSO method to identify six critical genes (*TRPC5*, *TENM1*, *NELL2*, *DMD*, *SLC35F3*, and *AUTS2*; Fig. 6C, D). As shown in Fig. 6E, these candidate diagnostic genes were obviously overexpressed in

the tumor cells. Finally, the RT-qPCR analysis demonstrated that *TRPC5*, *TENM1*, *NELL2*, *DMD*, *SLC35F3*, and *AUTS2* were highly expressed in PTC tissue than in normal tissue (Fig. 6F). In summary, *TRPC5*, *TENM1*, *NELL2*, *DMD*, *SLC35F3*, and *AUTS2* could be potential biomarkers for PTC diagnosis.



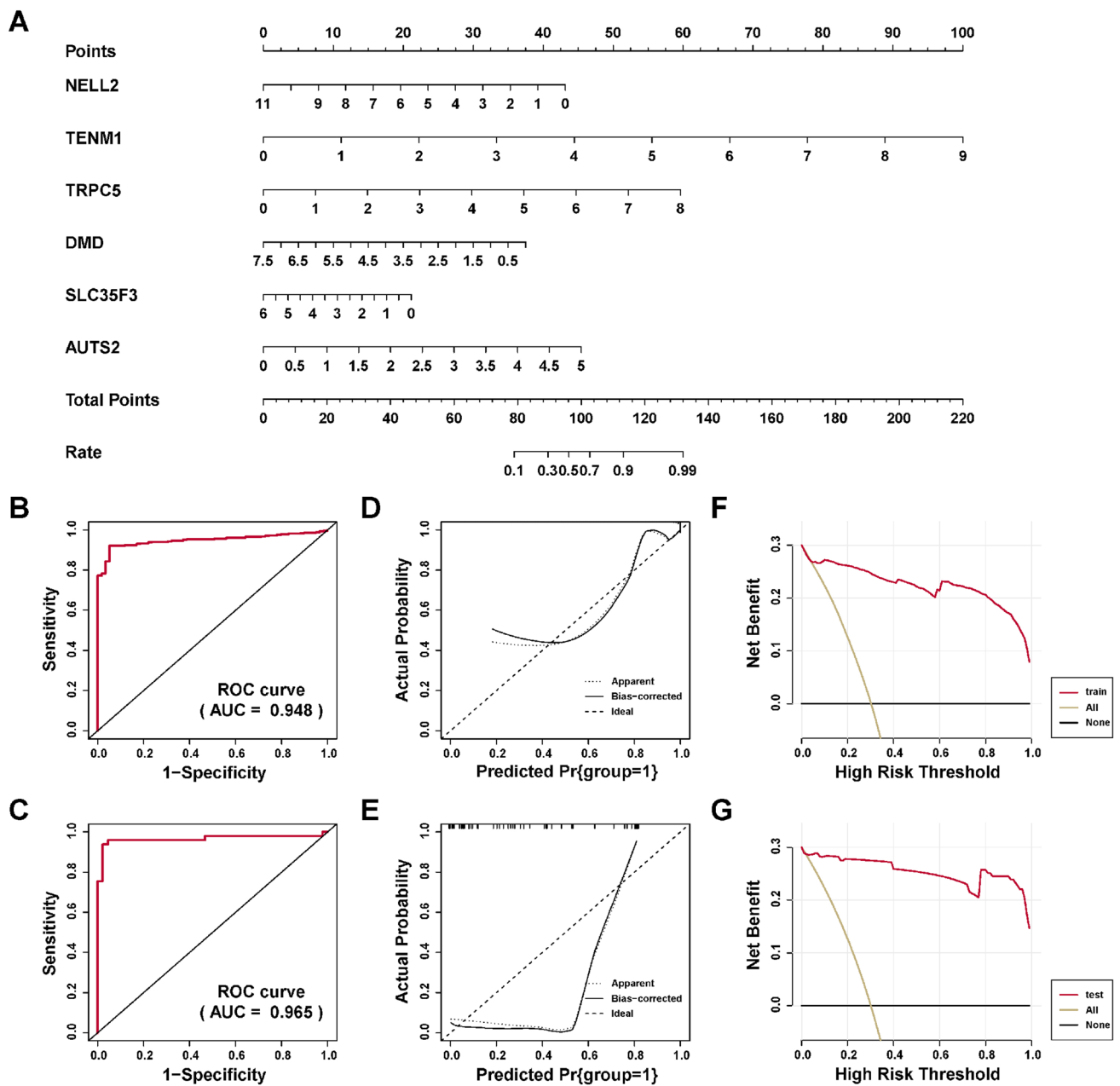
**Fig. 6** Identification of candidate diagnostic genes. **A** Venn diagram showing the intersection of up-related genes in tumor tissue from the TCGA-THCA (PTC) dataset and marker genes of tumor cells from the scRNA-seq dataset. **B** Importance score of candidate genes calculating by Boruta algorithm. **C** LASSO regression analysis: coefficient values at varying levels of penalty. Each curve represents a gene. **D** Ten-fold cross-validation was used to calculate the best lambda, con-

tributing to the minimum mean cross-validated error. **E** Expression levels of six candidate diagnostic genes plotted onto the UMAP in the scRNA-seq dataset. Color key from gray to blue indicates relative expression levels from low to high. **F** The mRNA relative expression of seven candidate diagnostic genes in normal and cancerous thyroid tissues. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$

### Construction of a diagnostic model for PTC

To construct a diagnostic model for diagnosis for PTC, logistic regression was performed to construct a nomogram to predict the risk of PTC based on six critical genes (Fig. 7A). ROC R package was utilized to assess the discriminatory ability of the diagnostic model. The AUCs were 0.948 (Fig. 7B) and 0.965 (Fig. 7C) in the train and test sets, respectively, suggesting the effective role of the model

to distinguish the PTC and normal tissues. In addition, the calibration curves in the train and test set suggested good predictive accuracy (Fig. 7D, E), with a  $P$  value for HL tests were 0.94 and 0.14 in train and test sets. Finally, the DCA were performed to analyze the clinical benefit of the model we constructed. As shown in Fig. 7F, G, the model we established could confirm the good benefits of clinical intervention. In a word, our model with good discriminatory ability could be served as a potential tool for PTC diagnosis.



**Fig. 7** Construction and validation of a diagnostic model for PTC. **A** Nomogram to estimate the risk of PTC tissues. The area under ROC curve was utilized to estimate the discrimination of the model in training set (**B**) and test set (**C**). Calibration curves of the diagnostic

model in the **D** training (HL test,  $P=0.94$ ) and **E** validation cohorts (HL test,  $P=0.14$ ). DCA of the diagnostic model in the **F** train and **G** test cohorts

### Discussion

A large number of transcriptomic and genomic studies have been performed to identify diagnostic and prognostic biomarkers for THCA. Several effective diagnostic and prognostic biomarkers are prepared to apply in clinical practice for THCA, including somatic mutations and other molecular changes (Nikiforov and Nikiforova 2011). However, PTC, as a main type of THCA, is a highly heterogeneous cancer.

Previous studies have explored RNA expression in tumor tissue based on transcriptomics, which could not accurately reflect differences in gene expression between different cells within the tissue. Single-cell sequencing has emerged to identify individual-cell-level changes in gene expression. Therefore, with the progress of sequencing technology, we can better characterize the tumor microenvironment of PTC and find more specific tumor-related biomarkers for precise diagnosis and individual treatment.

Here, we comprehensively analyzed the tumor microenvironment of PTC. Our results revealed the potential aberrant interactions between the T/NK and other cells (endothelial cells, fibroblasts, and myeloid cells) through *FNI/ITGB1*. Fibronectin is confirmed to be a critical role of wound healing, cellular differentiation and growth, migration, and adhesion (Rick et al. 2019). And integrin- $\beta$  (*ITGB*), a member of the integrin superfamily, is essential for cell proliferation, adhesion, and differentiation (Miranti and Brugge 2002). Gu et al. (2023) found that overexpression of *ITGB1* in gastric cancer was related to a poor prognosis and immune suppression. Also, increased *ITGB1* expression was confirmed to be associated with poor prognosis and increased fibroblast infiltration in pancreatic ductal adenocarcinoma (Benesch et al. 2022). Consistently with *ITGB1*, high *FNI* expression was also related to M2 macrophage infiltration and poor prognosis in THCA (Geng et al. 2021). These findings suggested the immune suppression role of *FNI/ITGB1* in tumor microenvironment. Taken together, T/NK cells in PTC may form the immunosuppressive tumor microenvironment via *FNI/ITGB1* interaction with other cells. Therefore, *FNI/ITGB1* may be a potential therapeutic target to target the tumor microenvironment in PTC.

Through the CNV analysis, we identified thyroid tumor cells and analyzed the potential trajectories of thyroid cell differentiation into tumor cells. Several genes related to tumor cell differentiation were identified: *CYR61*, *SERPINA1*, *TIMP1*, *FNI*, *S100A6*, *APOE*, and *APOC1*. Studies have confirmed the prognostic values of *CYR61* (Ren et al. 2021), *SERPINA1* (Wu et al. 2021), and *FNI* in THCA. By further subdividing the tumor cells, eight distinct tumor subgroups were identified. Tumor subgroups were involved in different biological processes, indicating a high degree of heterogeneity among tumor subgroups. The cell–cell interaction analysis revealed a strong interaction possibility between tumor cells and fibroblast/endothelial cells. Cancer-associated fibroblasts (CAFs) play essential roles in tumor development. They secrete extracellular matrix proteins, inflammatory ligands, and growth factors which promote cancer cell proliferation, immune exclusion, and therapy resistance (Biffi and Tuveson 2021). Fibroblasts and endothelial cells are the primary sources of CAFs. Endothelial cells are involved in intravasation, which allows invasive cancer cells to translocate into the blood vessel lumen (Sobierajska et al. 2020). These studies have revealed the critical role of endothelial cells and fibroblasts in TME. Further, our results identified potential ligand-receptor pairs (*MDK/LRP1*, *MDK/ALK*, *GAS6/MERTK*, and *GAS6/AXL*) for cell communication between these two types of cells and tumor cells. The *MDK* and *GAS* signaling pathway has been confirmed to regulate several biological processes in cells, including proliferation, survival and migration in tumor microenvironment by binding to their receptors (Wu et al.

2018; Filippou et al. 2020). Therefore, endothelial cells and fibroblasts may regulate the development of tumor cells and promote tumor microenvironment through these pairs.

To identify effective diagnostic markers, we intersected the up-regulated genes in tumor cells and up-regulated genes in PTC tissues. *TRPC5*, *TENM1*, *NELL2*, *DMD*, *SLC35F3*, and *AUTS2* were identified as diagnostic biomarkers for THCA. To validate our findings, we performed RT-qPCR on 20 cases of paired PTC and para-cancer tissue. These gene expression levels confirmed the accuracy of our analysis. Although further study and clinical experiment are needed, the potential of the six genes to be successful diagnostic biomarkers for PTC has been consolidated in published studies. For example, Kechin et al. (2022) found that *AUTS2* might be as one of important genes which can classify PTC in relation to the presence of driver NTRK-chimeric TRK genes. In addition, *AUTS2* was also reported to be positively associated with other cancer progression via TGF-beta pathway activation, HEDGEHOG and WNT signaling pathway (Han et al. 2015). These pathways play an essential role in tumor metabolism and immunity (Hanna and Shevde 2016; Zhao et al. 2020; Zou and Park 2023). Cheng et al. (2017) reported that *TEMN1* was highly expressed in cancerous thyroid tissues, and *TENM1* expression in PTC was related to an advanced stage, *BRAF* V600E mutation, extra-thyroidal invasion and the classical subtype. These findings demonstrate the diagnostic efficacy and potential of candidate genes in differentiating benign and malignant thyroid tissue. In addition, the diagnostic model we constructed shows good differentiation and accuracy in both train and external test set, which indicates that the model is stable and practical. However, our study has some limitations. First, the scRNA-seq data in our study was limited. More single-cell data must be collected to reveal the prevalence of microenvironmental features in PTC. Additionally, in vitro and in vivo experiments are needed to further verify the mechanism of the receptor–ligand we screened for in PTC. Finally, we just analyzed and validated the mRNA levels of six candidate biomarkers; their application must be confirmed through further experiments and clinical studies.

In conclusion, we characterized the tumor microenvironment of thyroid papillary carcinoma and identified essential receptor and ligand pairs. Notably, we comprehensively explored the heterogeneity of thyroid tumor cells. Further, we established a diagnostic model based on six candidate diagnostic markers for PTC. Our findings provide new insights into the heterogeneity of PTC and the theoretical basis for diagnosing PTC.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00432-023-05420-8>.

**Author contributions** Performed literature search and bioinformatics analysis: HXH, WYZ, TY; Data collection: SWB, LYX; Figures

preparation: TY; writing—original draft preparation: WYZ; Writing—review and editing: JLN, WXN, JRY; Funding acquisition: LYX; Investigation: LYX, WYZ; Supervision: TY, HXH. All authors read and approved the final manuscript.

**Funding** This work was supported by Tianjin Municipal Education Commission (grant number 2021KJ248).

**Availability of data and materials** Datasets related to this article are from public database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE191288>, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE33630>). All data generated or analyzed during this study are included in this article and its supplementary information files.

## Declarations

**Conflict of interest** The authors declare that they have no competing interests.

**Ethics approval and consent to participate** For patient samples, written informed consent was obtained from each patient and the study was approved by the hospital ethics committee (Tianjin Medical University General Hospital). The study was performed in accordance with the Declaration of Helsinki.

**Consent for publication** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Agianian B, Gavathiotis E (2018) Current insights of BRAF inhibitors in cancer. *J Med Chem* 61(14):5775–5793
- Ahlmann-Eltze C, Huber W (2021) glmGamPoi: fitting Gamma-Poisson generalized linear models on single cell count data. *Bioinformatics* 36(24):5701–5702
- Alexander EK, Cibas ES (2022) Diagnosis of thyroid nodules. *Lancet Diabetes Endocrinol* 10(7):533–539
- Bartoschek M, Oskolkov N, Bocci M, Lovrot J, Larsson C, Sommarin M, Madsen CD, Lindgren D, Pekar G, Karlsson G, Ringner M, Bergh J, Bjorklund A, Pietras K (2018) Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing. *Nat Commun* 9(1):5150
- Benesch MG, Wu R, Menon G, Takabe K (2022) High beta integrin expression is differentially associated with worsened pancreatic ductal adenocarcinoma outcomes. *Am J Cancer Res* 12(12):5403–5424
- Biffi G, Tuveson DA (2021) Diversity and biology of cancer-associated fibroblasts. *Physiol Rev* 101(1):147–176
- Chen Z, Zhou L, Liu L, Hou Y, Xiong M, Yang Y, Hu J, Chen K (2020) Single-cell RNA sequencing highlights the role of inflammatory cancer-associated fibroblasts in bladder urothelial carcinoma. *Nat Commun* 11(1):5077
- Chen DW, Lang BHH, McLeod DSA, Newbold K, Haymart MR (2023) Thyroid cancer. *Lancet* 401(10387):1531–1544
- Cheng SP, Chen MJ, Chien MN, Lin CH, Lee JJ, Liu CL (2017) Overexpression of teneurin transmembrane protein 1 is a potential marker of disease progression in papillary thyroid carcinoma. *Clin Exp Med* 17(4):555–564
- Costa AM, Herrero A, Fresno MF, Heymann J, Alvarez JA, Cameselle-Teijeiro J, Garcia-Rostan G (2008) BRAF mutation associated with other genetic events identifies a subset of aggressive papillary thyroid carcinoma. *Clin Endocrinol (oxf)* 68(4):618–634
- Fagin JA, Wells SA Jr (2016) Biologic and clinical perspectives on thyroid cancer. *N Engl J Med* 375(11):1054–1067
- Fallahi P, Ferrari SM, Galdiero MR, Varricchi G, Elia G, Ragusa F, Paparo SR, Benvenuta S, Antonelli A (2022) Molecular targets of tyrosine kinase inhibitors in thyroid cancer. *Semin Cancer Biol* 79:180–196
- Filippou PS, Karagiannis GS, Constantinidou A (2020) Midkine (MDK) growth factor: a key player in cancer progression and a promising therapeutic target. *Oncogene* 39(10):2040–2054
- Geng QS, Huang T, Li LF, Shen ZB, Xue WH, Zhao J (2021) Overexpression and prognostic significance of FN1, correlating with immune infiltrates in thyroid cancer. *Front Med (lausanne)* 8:812278
- Gu W, Sun H, Zhang M, Mo S, Tan C, Ni S, Yang Z, Wang Y, Sheng W, Wang L (2023) ITGB1 as a prognostic biomarker correlated with immune suppression in gastric cancer. *Cancer Med* 12(2):1520–1531
- Han Y, Ru GQ, Mou X, Wang HJ, Ma Y, He XL, Yan Z, Huang D (2015) AUTS2 is a potential therapeutic target for pancreatic cancer patients with liver metastases. *Med Hypotheses* 85(2):203–206
- Hanna A, Shevde LA (2016) Hedgehog signaling: modulation of cancer properties and tumor microenvironment. *Mol Cancer* 15:24
- Hwang B, Lee JH, Bang D (2018) Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med* 50(8):1–14
- Ito Y, Miyauchi A, Kihara M, Fukushima M, Higashiyama T, Miya A (2018) Overall survival of papillary thyroid carcinoma patients: a single-institution long-term follow-up of 5897 patients. *World J Surg* 42(3):615–622
- Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, Myung P, Plikus MV, Nie Q (2021) Inference and analysis of cell-cell communication using Cell Chat. *Nat Commun* 12(1):1088
- Kechin AA, Ivanov AA, Kel AE, Kalmykov AS, Oskorbin IP, Boyarskikh UA, Kharporov EA, Bakharev SY, Oskina NA, Samuilenkova OV, Vikhlyanov IV, Kushlinskii NE, Filipenko ML (2022) Prediction of EVT6-NTRK3-dependent papillary thyroid cancer using minor expression profile. *Bull Exp Biol Med* 173(2):252–256
- Kobaly K, Kim CS, Mandel SJ (2022) Contemporary management of thyroid nodules. *Annu Rev Med* 73:517–528
- Li X, Sun Z, Peng G, Xiao Y, Guo J, Wu B, Li X, Zhou W, Li J, Li Z, Bai C, Zhao L, Han Q, Zhao RC, Wang X (2022) Single-cell RNA sequencing reveals a pro-invasive cancer-associated fibroblast subgroup associated with poor clinical outcomes in patients with gastric cancer. *Theranostics* 12(2):620–638
- Ma X, Guo J, Liu K, Chen L, Liu D, Dong S, Xia J, Long Q, Yue Y, Zhao P, Hu F, Xiao Z, Pan X, Xiao K, Cheng Z, Ke Z, Chen ZS, Zou C (2020) Identification of a distinct luminal subgroup diagnosing and stratifying early stage prostate cancer by tissue-based single-cell RNA sequencing. *Mol Cancer* 19(1):147
- McGinnis CS, Murrow LM, Gartner ZJ (2019) DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst* 8(4):329–337 e324
- Miranti CK, Brugge JS (2002) Sensing the environment: a historical perspective on integrin signal transduction. *Nat Cell Biol* 4(4):E83–90

- Nikiforov YE, Nikiforova MN (2011) Molecular genetics and diagnosis of thyroid cancer. *Nat Rev Endocrinol* 7(10):569–580
- Pu W, Shi X, Yu P, Zhang M, Liu Z, Tan L, Han P, Wang Y, Ji D, Gan H, Wei W, Lu Z, Qu N, Hu J, Hu X, Luo Z, Li H, Ji Q, Wang J, Zhang X, Wang YL (2021) Single-cell transcriptomic analysis of the tumor ecosystems underlying initiation and progression of papillary thyroid carcinoma. *Nat Commun* 12(1):6058
- Ren H, Liu X, Li F, He X, Zhao N (2021) Identification of a six gene prognosis signature for papillary thyroid cancer using multi-omics methods and bioinformatics analysis. *Front Oncol* 11:624421
- Rick JW, Chandra A, Dalle Ore C, Nguyen AT, Yagnik G, Aghi MK (2019) Fibronectin in malignancy: cancer-specific alterations, protumoral effects, and therapeutic implications. *Semin Oncol* 46(3):284–290
- Siegel RL, Miller KD, Wagle NS, Jemal A (2023) Cancer statistics, 2023. *CA Cancer J Clin* 73(1):17–48
- Sobierajska K, Ciszewski WM, Sacewicz-Hofman I, Niewiarowska J (2020) Endothelial cells in the tumor microenvironment. *Adv Exp Med Biol* 1234:71–86
- Song E, Jin M, Jang A, Jeon MJ, Song DE, Yoo HJ, Kim WB, Shong YK, Kim WG (2021) Mutation in genes encoding key functional groups additively increase mortality in patients with BRAF(V600E)-mutant advanced papillary thyroid carcinoma. *Cancers (basel)* 13(22):5846
- Sui X, Chen C, Zhou X, Wen X, Shi C, Chen G, Liu J, He Z, Yao Y, Li Y, Gao Y (2023) Integrative analysis of bulk and single-cell gene expression profiles to identify tumor-associated macrophage-derived CCL18 as a therapeutic target of esophageal squamous cell carcinoma. *J Exp Clin Cancer Res* 42(1):51
- Wu G, Ma Z, Cheng Y, Hu W, Deng C, Jiang S, Li T, Chen F, Yang Y (2018) Targeting Gas6/TAM in cancer cells and tumor microenvironment. *Mol Cancer* 17(1):20
- Wu L, Zhou Y, Guan Y, Xiao R, Cai J, Chen W, Zheng M, Sun K, Chen C, Huang G, Zhang X, Qian Z, Shen S (2021) Seven genes associated with lymphatic metastasis in thyroid cancer that is linked to tumor immune cell infiltration. *Front Oncol* 11:756246
- Xia C, Dong X, Li H, Cao M, Sun D, He S, Yang F, Yan X, Zhang S, Li N, Chen W (2022) Cancer statistics in China and United States, 2022: profiles, trends, and determinants. *Chin Med J (engl)* 135(5):584–590
- Xiao Y, Yu D (2021) Tumor microenvironment as a therapeutic target in cancer. *Pharmacol Ther* 221:107753
- Zhang X, Lan Y, Xu J, Quan F, Zhao E, Deng C, Luo T, Xu L, Liao G, Yan M, Ping Y, Li F, Shi A, Bai J, Zhao T, Li X, Xiao Y (2019) Cell marker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res* 47(D1):D721–D728
- Zhao H, Wei J, Sun J (2020) Roles of TGF-beta signaling pathway in tumor microenvironment and cancer therapy. *Int Immunopharmacol* 89(Pt B):107101
- Zou G, Park JI (2023) Wnt signaling in liver regeneration, disease, and cancer. *Clin Mol Hepatol* 29(1):33–50

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Yizeng Wang<sup>1</sup> · Wenbin Song<sup>1</sup> · Yingxi Li<sup>2</sup> · Zhaoyi Liu<sup>1</sup> · Ke Zhao<sup>1</sup> · Lanning Jia<sup>1</sup> · Xiaoning Wang<sup>1</sup> · Ruoyu Jiang<sup>1</sup> · Yao Tian<sup>1</sup> · Xianghui He<sup>1</sup>

✉ Yao Tian  
tianyao@tmu.edu.cn

✉ Xianghui He  
hexh88@tmu.edu.cn

<sup>1</sup> Department of General Surgery, Tianjin Medical University General Hospital, 154 Anshan Road, Heping District, Tianjin 300052, People's Republic of China

<sup>2</sup> Key Laboratory of Immune Microenvironment and Disease (Ministry of Education), Tianjin Medical University, Tianjin 300070, People's Republic of China