1 Single-cell RNA sequencing reveals a novel cell type and immunotherapeutic

2 targets in papillary thyroid cancer

- 3 Zhengshi Wang^{1,2,†}, Youlutuziayi Rixiati^{3,†}, Wenli Jiang^{4,†}, Chen Ye⁵, Caiguo Huang⁴,
- 4 Chuangang Tang^{6,*}, Zhiqiang Yin^{1,2,*}, Binghua Jiao^{4,*}
- 5
- 6 ¹ Thyroid Center, Shanghai Tenth People's Hospital, Tongji University School of
- 7 Medicine, Shanghai, P.R.China;
- ² Shanghai Center for Thyroid Diseases, Shanghai, 200072, P.R.China;
- ³ Department of Pathology, Soochow University Medical School, Suzhou, 215123,
- 10 P.R.China;
- ⁴ Department of Biochemistry and Molecular Biology, College of Basic Medical,
- 12 Navy Medical University, Shanghai, 200433, P.R.China;
- ⁵ Department of Urology, Changhai Hospital, Navy Medical University, Shanghai,
- 14 200433, P.R.China;
- ⁶ Department of Thyroid and Breast Surgery, Xuzhou Central Hospital, The Affiliated
- 16 Xuzhou Hospital of Medical College of Southeast University, Xuzhou, 221009,
- 17 P.R.China.
- 18
- 19 [†]They contributed equally to the manuscript.
- 20 *Corresponding authors:

21 Binghua Jiao, PhD. Department of Biochemistry and Molecular Biology, College of

- 22 Basic Medical, Navy Medical University, Shanghai, 200433, P.R.China. Tel:
- 23 +86-13801924888, Email: <u>biojiao@163.com</u>.
- 24 Zhiqiang Yin, MD. Shanghai Center for Thyroid Diseases, No. 301, Middle
- 25 Yanchang Road, Shanghai, P.R.China. Tel.: +86-18917689323, Email:
- 26 <u>972683004@qq.com</u>
- 27 Chuangang Tang, MD. Department of Thyroid and Breast Surgery, Xuzhou Central
- 28 Hospital, The Affiliated Xuzhou Hospital of Medical College of Southeast University,
- 29 Xuzhou, 221009, P.R.China. Tel.: +86-18652230137, Email: tcg3711@163.com.

30 Abstract

31 Papillary thyroid cancer (PTC) is the most common thyroid malignancy. Although 32 PTC usually has a favorable prognosis, some aggressive PTC subtypes and lymph 33 node (LN) metastasis contribute to high rates of recurrence and poor clinical 34 outcomes. We analyzed single-cell RNA sequencing (scRNA-seq) data from 15 35 samples, including primary tumors of PTC, metastatic LNs, and paracancerous tissues. 36 After quality filtering, 28,205 cells were detected. Of these, 13,390 cells originated 37 from 7 tumor tissues, 2,869 cells from 2 metastatic LNs, and 11,945 cells from 6 38 paracancerous tissues. The increase in the proportion of CD4⁺ Tregs may be a key 39 factor responsible for the immunosuppressive property of PTC. A novel cell type was identified, named Protective EGR1⁺CD4⁺ T cell, which might be antagonistic to the 40 41 CD4⁺ Tregs and inhibit the formation of the immunosuppressive microenvironment 42 and tumor immune evasion. Inhibitory checkpoints TIGIT and CD96 were found to be 43 better targets than PD-1 for immune therapy in PTC patients with LN metastasis. For 44 PTC patients without LN metastasis, however, PD-1, TIGIT, and CD96 could be 45 suitable targets of immunotherapy. These findings would contribute to the further 46 understanding of molecular mechanisms resulting in occurrence and development of 47 PTC, and provide a theoretical rationale for targeted therapy and immunotherapy. 48 **Keywords**

49 papillary thyroid cancer, single-cell RNA sequencing, tumor microenvironment

50

51

52 Introduction

53 According to the GLOBOCAN database, there were an estimated 586,202 new cases of thyroid cancer worldwide in 2020¹. The global incidence was up to 13.1 per 54 55 100,000 people, ranking 11th in occurrence rate. The most common pathological type of thyroid cancer is papillary thyroid cancer (PTC), accounting for approximately 85% 56 of all cases². Although PTC usually has a favorable prognosis, some aggressive PTC 57 58 subtypes (e.g. tall cell variant and Hürthle cell cancer) and lymph node (LN) metastasis contribute to high rates of recurrence and poor clinical outcomes ^{3, 4}. For 59 60 PTC patients with a high risk of recurrence, radioiodine treatment could be used to 61 improve their prognosis ⁵. However, some PTC patients would develop into radioiodine-refractory PTC, which in turn lead to low survival rate ^{6,7}. Therefore, a 62 deeper understanding of molecular and cellular mechanisms of PTC would help in the 63 64 development of therapeutic strategies.

65 Currently, efforts to study the tumor progression and metastatic process of 66 thyroid cancer have mainly focused on the analysis of cancer cells using genetic aberrations^{8, 9, 10, 11}. However, the tumor progression and metastasis are a complicated 67 biological process, which were not only affected by the characteristic features of 68 cancer cells themselves but also by the tumor microenvironment (TME) $^{12,\ 13,\ 14,\ 15}.$ 69 70 TME refers to a tumor pathology-related environment, comprising stromal cells, 71 extracellular matrix (ECM), and cytokines. A comprehensive analysis of TME in PTC 72 can reveal the key elements involved in the susceptibility of tumor-induced 73 immunological changes, which could be employed to develop new immunotherapy 74 strategies.

75 Genomic and transcriptomic studies have revealed a huge number of driver 76 mutations, abnormal regulatory programs, and disease subtypes in major human tumors^{16, 17, 18, 19}. However, the conventional bulk sequencing frequently adopted in 77 78 these studies only revealed the overall biological characteristics of each tumor and 79 lacked the ability to capture signatures in intratumoral and intercellular 80 heterogeneity. As opposed to bulk sequencing, the emergence of single-cell RNA 81 sequencing (scRNA-seq) provided a new opportunity to enables characterization of 82 cell populations at the single-cell level, and made it more independent of any previous assumptions about surface markers ²⁰. scRNA-seq has been applied to investigate 83 cellular properties in various solid tumor types ²¹. However, the single-cell atlas of 84

PTC remains to be fully revealed. Therefore, we analyzed primary PTC tumors,
paracancerous tissues, and metastatic LNs using the scRNA-seq-based profiling

87 method to better understand the intratumoral heterogeneity and complexity during the

88 development of PTC.

89 Results

90 Single cell atlas and heterogeneity of PTC

91 A total of 15 samples from seven PTC patients were involved in this study. After 92 quality filtering, 28,205 cells were detected. Of these, 13,390 cells originated from 7 93 tumor tissues, 2,869 cells from 2 metastatic LNs, and 11,945 cells from 6 94 paracancerous tissues (Fig. 1a and 1b). Subsequently, we partitioned the cells into 26 95 clusters, which were further classified into 10 major cell types based on known 96 markers described in previous studies: B cells (CD19⁺, MS4A1⁺, CD38⁺, CD79A⁺, 97 $CD79B^+$; $CD4^+T$ cells ($CD3D^+$, $CD4^+$); $CD8^+T$ cells ($CD3D^+$, $CD8A^+$); endothelial cells (CD31⁺, CD34⁺); epithelial cells (EPCAM⁺, KRT18⁺); fibroblasts (COL1A1⁺); 98 99 myeloid cells (CD14⁺, CD86⁺, ITGAX⁺, CD80⁺, CD83⁺, ITGAM⁺); naive T cells 100 $(CD3D^+, CCR7^+)$; natural killer T (NKT) cells $(CD3D^+, NKG7^+)$; and plasma cells 101 (CD79A⁺, SDC1⁺) (Fig. 1c, Supplementary Fig. 1 and 2).

102 Based on the frequencies of cell types, we detected cellular landscapes in 103 primary tumors, paracancerous tissues, and metastatic LNs, respectively. significant 104 differences were observed in several cell types (Fig. 1d). The immune cells accounted 105 for the major differences between primary tumors and paracancerous tissues, with the 106 frequency of NKT cells and plasma cells increasing, and CD4⁺ T cells and B cells 107 decreasing in the primary tumors (P < 0.05). The most striking changes between 108 metastatic LNs and primary tumors were observed in CD8⁺ T cells. At the cluster 109 level, there were also significant differences within the same cell types. As 110 exemplified by myeloid cells, we observed an increase in the frequency of cluster 21 111 and a decrease in the proportion of cluster 25 in metastatic LNs compared with 112 primary tumors and paracancerous tissues. These results demonstrated that the disease 113 progression might also be somehow associated with evolutionary immune cells, 114 reflecting the existence of heterogeneity among samples and cell populations.

To conduct the interaction network of tumor microenvironment in PTC, CellphoneDB was used to calculate potential ligand-receptor pairs in cells. And Cytoscape was performed to visualize the cell interaction. We found that myeloid cells and T cell-related cells possessed more interaction pairs with other cells than

119 others (Fig. 1e), showing the dominant roles of myeloid cells and T cells.

120 Identification of malignant cells in epithelial cells

121 To distinguish malignant from non-malignant cells within the epithelial cells, 122 CopyKAT was performed to identify PTC genome alterations. Despite the 123 heterogeneity, almost all malignant cells possessed deletions from chromosomes 16 124 and 19 and amplifications in chromosomes 13 (Fig. 2a and 2b). We distinguished 125 1,679 non-malignant cells and 450 cells were identified as malignant cells. Very few 126 malignant cells were also found in paracancerous tissues (Fig. 2b). In order to support 127 the identification results of the malignant cells, pseudotime trajectory analysis was 128 performed. The results showed that malignant cells were present at the end of the 129 differentiation trajectory (Fig. 2c).

130 To screen for key regulons in malignant cells, we conducted SCENIC analysis 131 and identified motifs ATF3 and BHLHE40 that were highly activated in malignant 132 cells (Fig. 2d). In a previous study, ATF3 was found to enhance breast cancer metastasis ³⁶. At the same time, decreased activity was found in CREM and ETS1. In 133 134 TCGA THCA cohort, a high level of CREM and ETS1 was significantly related to a 135 good prognosis, and BHLHE40 is highly expressed in patients with LN metastasis 136 compared with patients without LN metastasis (Fig. 2e and 2f). These results provide 137 potential targets for suppressing cells to possess malignant characteristics.

We further characterized the functions of differential genes between metastatic LNs and primary tumors by comparing pathway activities. Pathways involved in cellular component assembly, protein-containing complex assembly, and regulation of microtubule motor activity were relatively upregulated in LN-derived malignant cells (Fig. 2g). The results showed that malignant cells in metastatic LNs have stronger cell proliferative and invasive abilities.

144 Identification of a novel T cell type

145 Reclustering of T cells identified 10 subclusters: CD4⁺ Tregs, Cytotoxic CD8⁺ T 146 cells, Exhausted CD4⁺ T cells, Follicular helper (Tfh) T cels, Naïve CD4⁺ T cells, 147 Naïve T cells; NKT cells, Pre-exhausted CD8⁺ T cells, Proliferating CD8⁺ T cells, and Protective EGR1⁺CD4⁺ T cells (Fig. 3a and 3b). In addition to the markers used above, 148 we also found some other makers that can be used for cell population identification. 149 150 For example, TYMS can be used for Pre-exhausted $CD8^+$ T cells identification, and RTKN2 can be used for Pre-exhausted CD4⁺ Tregs identification (Fig. 3c, 151 152 Supplementary Fig. 3). Notably, a novel cell type was identified, which was totally

153 different from preconceived cellular definitions. The novel cell type specifically 154 highly expressed EGR1, and we named it as Protective EGR1⁺CD4⁺T cells. The 155 mRNA expression level and protein level of EGR1 were significantly decreased in 156 tumor tissues compared with normal tissues (Fig. 3d). Survival analysis showed that 157 patients with high EGR1 levels had a better prognosis than those with low EGR1 158 levels (Fig. 3e). Through cell frequency analysis, we found an interesting 159 phenomenon that the proportion of CD4⁺ Tregs tended to increase with PTC disease progression and Protective EGR1⁺CD4⁺ T cells exhibited a quite opposite trend (Fig. 160 161 3f). The results were validated with TCGA data calculated using Cibersortx (Fig. 3g) 162 and further validated by the changes of marker gene expression levels in the TCGA 163 THCA cohort (Supplementary Fig. 4a). These findings indicated that both of CD4⁺ 164 Tregs and Protective EGR1⁺CD4⁺ T cells may have an important impact on disease 165 progression, and our newly identified cell type (Protective EGR1⁺CD4⁺T cells) may 166 be antagonistic to the CD4⁺ Tregs.

167 We further characterized the functions of CD4⁺ Tregs and Protective EGR1⁺CD4⁺ T cells by comparing pathway activities. In the primary tumors, $CD4^+$ 168 Tregs and Protective EGR1⁺CD4⁺ T cells exhibited vastly different signaling 169 pathways. CD4⁺ Tregs were related to epithelial-mesenchymal transition (EMT) and 170 171 hypoxia, whereas G2M checkpoint and KRAS signaling pathway were significantly down-regulated in Protective EGR1⁺CD4⁺ T cells (Fig. 4a). In the metastatic LNs, 172 173 mitotic spindle and oxidative phosphorylation were also down-regulated in Protective EGR1⁺CD4⁺ T cells (Supplementary Fig. 4b). To screen for key genes related to 174 175 tumorigenesis and tumor development in CD4⁺ Tregs and Protective EGR1⁺CD4⁺ T 176 cells, we conducted SCENIC analysis and identified essential motifs in CD4⁺ Tregs 177 and Protective EGR1⁺CD4⁺ T cells. FOSL2, ATF3, REL, and HES1 were CD4⁺ 178 Tregs-specific motifs. BCLAF1 and ETS1 motifs were highly activated in Protective 179 EGR1⁺CD4⁺ T cells (Fig. 4b). These results provided potential targets for inhibiting 180 or reversing the formation of the immunosuppressive microenvironment.

The DEGs derived from Protective EGR1⁺CD4⁺ T cells were identified, including 31 up-regulated mRNAs and 50 down-regulated mRNAs in primary tumors compared with paracancerous tissues (Supplementary Fig. 4c). The DEGs derived from CD4⁺ Tregs were also identified (Supplementary Fig. 4d). Subsequently, we examined the immune checkpoints in the cell clusters (Fig. 4c and 4d). Notably, we found that TIGIT, an inhibitory checkpoint, was upregulated in Exhausted CD4⁺ T

cells and CD96 was upregulated in Pre-exhausted CD8⁺ T cells in metastatic LNs. In
tumor tissues, PDCD1 (PD-1) and TIGHT were significantly upregulated in
Exhausted CD4⁺ T cells, and CD96 was significantly upregulated in Pre-exhausted
CD8⁺ T cells. Since TIGIT, CD96, and PDCD1 (PD-1) are markers of T cell
exhaustion, these data indicated that Exhausted CD4⁺ T cells and Pre-exhausted CD8⁺
T cells were exhausted in the tumor microenvironment, which was consistent with our
previous cell population classification results.

194 Neutrophils cells are extremely reduced in tumor

195 For exploring the heterogeneity among myeloid cells, 1607 myeloid cells were 196 clustered into 5 cell subtypes based on known markers described in previous studies: 197 macrophages, dendritic cells, monocytes, neutrophils, and myeloid-derived suppressor 198 cells (Fig. 5a and 5b). Through cell frequency analysis, a marked decrease of 199 neutrophils was observed in tumors while the proportion of macrophages increased 200 continuously as the disease progressed (Fig. 5c). The trend was consistent with TCGA 201 data calculated by cibersortx (Fig. 5d). Survival analysis showed that disease-free 202 survival (DFS) of patients with low-level dendritic cell was longer than that of the 203 high-level group in the TCGA cohort (Fig. 5e). Similar results were observed in 204 macrophages (Supplementary Fig. 5a). However, low-level neutrophil was observed 205 to be associated with better prognosis (Supplementary Fig. 5b). These results 206 suggested that macrophages, dendritic cells, and neutrophils are more relevant to the 207 development of PTC.

208 We further characterized the functions of myeloid cell subtypes by comparing 209 pathway activities. Compared with paracancerous tissues, the cancer hallmark-related 210 pathways were relatively enriched in myeloid-derived suppressor cells from tumors, 211 whereas they were generally down-regulated in neutrophils from tumors (Fig. 6a). 212 These were consistent with the changes in cell proportions we observed earlier (Fig. 213 5c). Compared with tumor tissues, however, some cancer hallmark-related pathways 214 (i.e. mTORC1 signaling and E2F targets) were down-regulated in myeloid-derived 215 suppressor cells from metastatic LNs (Fig. 6b). The results indicated that 216 myeloid-derived suppressor cells were involved in cancer cell proliferation, but not in 217 tumor metastasis.

In order to screen for key genes related to tumor occurrence and development in myeloid cells, we conducted SCENIC analysis and identified essential motifs in myeloid cells subtypes. Similar to the results of pathway analysis, the overall

regulatory factors in neutrophils were also inconsistent with the other cell types (Fig. 6c). We recognized that the decreased activity of a lot of motifs (i.e. YY1 and IRF5) and activation of the STAT3, ZNF143, and HCFC1 motifs led to the reduction of neutrophils.

225 As shown in Fig. 6d, cells from myeloid cell-derived cluster 8 uniquely 226 expressed the M1 macrophages marker CCL5, whereas M2 macrophages markers 227 including CTSB/CTSD/FN1 were relatively highly expressed in cluster 0. The 228 proportion of cluster 0 was high in tumors and LNs, although the difference has not 229 reached a significant level (Fig. 5c). All of this evidence certified that cluster 0 230 represented an M2-like tumor-associated macrophages (TAM) cluster, the increase of 231 which may be related to disease progression. Cluster 9 highly expressed the cytokines 232 CCL17/CCL19/CCL22. These cytokines could bind to CCR4, a marker on the cell 233 membrane of $CD4^+$ Tregs, and showed strong chemotaxis to $CD4^+$ Tregs ^{37, 38}. It 234 corroborated with our above results that the increasing trend of CD4⁺ Tregs with 235 disease progression was coincident with that of myeloid cell-derived cluster 9 (Fig. 3f 236 and 5c). The findings indicated that the dendritic cells (cluster 9) could recruit $CD4^+$ 237 Tregs into the tumor regions, thereby acting as an immunosuppressive effect.

238 Discussion

239 At present, the study of molecular mechanisms underlying the occurrence and 240 development of PTC mainly focused on the genetical alteration based on bulk 241 sequencing data. We profiled integrated and heterologous transcriptional landscapes 242 of PTC using single-cell sequencing methods, including cancer cells and TME of 243 primary tumors and metastatic lesions. Through sequencing 28,205 cells, we found 244 that the majority of cell clusters possessed strong heterogeneities. Tumor 245 heterogeneity exists in various malignancies and remains not only between tumors but also within a single tumor ³⁹, which is considered as a cause of chemoresistance of 246 247 cancer. In-depth knowledge of tumor heterogeneity of PTC will make molecular 248 testing more reliable and accurate prior to surgery (whether to undergo the surgery or 249 close follow-up only), and benefit to stratify the recurrence risk and personalized 250 precision treatment (thyroidectomy or total thyroidectomy or plus radioiodine 251 treatment). On the other hand, TME on disease progression and metastasis has also been confirmed in numerous diseases^{40, 41}. Investigations of TME-related cellular and 252 253 molecular events will provide a theoretical rationale for drug discovery and 254 development, especially for targeted therapy and immunotherapy.

255 In addition to previously described cell types, we discovered a novel cell 256 subpopulation, named Protective EGR1⁺CD4⁺T cells. EGR1 is a transcription factor 257 primarily mediate cellular functions (i.e. cell growth, cancer progress and apoptosis) via the RAS/RAF/MEK/ERK signaling pathway ⁴², which was a hallmark of PTC ⁴³. 258 259 It has also been previously reported that p53 could bind to EGR1 promoter and subsequently resulted in efficient apoptosis ⁴⁴. EGR1 was usually considered a tumor 260 suppressor in many human malignancies including PTC ^{45, 46, 47, 48}. By comparing the 261 262 proportion of cells from different tissues and pathway enrichment analyzing, we 263 found that KRAS signaling pathway (part of RAS/RAF/MEK/ERK signaling pathway) 264 was significantly down-regulated in Protective EGR1⁺CD4⁺ T cells. It indicated that 265 Protective EGR1⁺CD4⁺T cells might inhibit the formation of the immunosuppressive 266 microenvironment and tumor immune evasion via RAS/RAF/MEK/ERK signaling 267 pathway, thereby restraining tumor growth. Therefore, EGR1 might represent a 268 potential therapeutic target for PTC.

269 CD4⁺ Tregs, were overall considered to disrupt anti-tumor immunity and then 270 help tumor cells to achieve immune evasion, leading to the tumor growth and metastasis ⁴⁹. Our results also showed that CD4⁺ Tregs had the greatest positive 271 272 correlation with the process of EMT, which matched those reported in other studies of PTC patients ⁵⁰. Using pathway enrichment analysis, we found that relevant pathways 273 274 associated with CD4⁺ Tregs and Protective EGR1+CD4+T cells were almost the 275 opposite. The proportion of CD4⁺ Tregs among three samples increased in the disease 276 progression order (paracancerous tissues < tumors < LNs) while the proportion of 277 Protective EGR1⁺CD4⁺T cells decreased sequentially. This strongly suggested that 278 CD4⁺ Tregs and Protective EGR1⁺CD4⁺T cells might act antagonistically to each 279 other.

280 Immunotherapy is an emerging method for cancer treatment and promising results have been obtained in both hematologic and solid tumors ^{51, 52}. T cell 281 282 exhaustion is a critical mechanism of immune evasion. Blockade the PD1/PDL1 283 interaction to reverse T cell exhaustion is considered a milestone achievement in the field of immunotherapy ^{53, 54, 55}. Our results showed that PDCD1 (PD-1), TIGIT, and 284 285 CD96 could be suitable targets of immunotherapy in PTC patients without LN 286 metastasis since they were upregulated in tumor tissues. However, the expression 287 levels of TIGIT and CD96 were much higher than that of PDCD1 (PD-1) in 288 LN-derived exhausted T cell types (Exhausted CD4⁺ T cells and Pre-exhausted CD8⁺

289 T cells). Therefore, TIGIT and CD96 might be better therapeutic targets for immunotherapy in PTC patients with LN metastasis than PDCD1 (PD-1). Jill et al. ⁵⁶ 290 291 investigated the CD4⁺ and CD8⁺ T cell exhaustion in PTC patients with LN metastasis 292 using flow cytometry, and found that CD8+ T cell exhaustion was incomplete. This 293 was most likely because TIGHT and CD96 were not included into analysis in their 294 study, both of which were very important inhibitory immune checkpoints. However, 295 their findings supported our conclusions to some extent that PDCD1 (PD-1) was not a 296 suitable molecular target for the treatment of PTC with LN metastasis.

297 We initially applied a commonly used method inferCNV for the identification of malignant cells ⁵⁷, which was shown to be unable to identify efficiently malignant 298 299 tumor cells from epithelial cells in this study. Thus, we adapted an integrative 300 Bayesian segmentation approach called CopyKAT and solved the conundrum 301 successfully ³⁵. As a result, more than half of tumor-derived epithelial cells were 302 non-malignant cells, which might well explain the reason for the indolent clinical 303 behavior of PTC. Additionally, there were also a very small number of malignant 304 epithelial cells in paracancerous tissues, possibly due to the presence of the pre-tumor 305 microenvironment in paracancerous tissues. There were several possible reasons. First, 306 the pre-metastatic niche, a favorable microenvironment to tumor metastasis, formed in paracancerous tissues ⁵⁸, and in turn lead to intra-thyroid metastasis. The concept of 307 308 intra-thyroid metastasis of thyroid cancer has not been described yet. The reason is 309 possibly caused by the small size of the thyroid gland itself and the narrow lumen of 310 artery/vein nourishing the thyroid gland. As is well-known, liver cancer tends to 311 present with intrahepatic metastatic disease, mainly due to the large diameter of the 312 portal vein and hepatic artery. Second, multifocal PTC has already existed within the 313 thyroid gland, whereas the extremely tiny lesion cannot be detected under 314 microscopic examination. Based on the two points above, thyroidectomy is 315 necessary to be performed for patients with PTC at least to ensure a total removal of 316 potential lesions. Third, CopyKAT is not able to discriminate malignant cells from 317 non-malignant cells with hundred percent accuracy.

In summary, we initially investigated the single-cell level heterogeneity of primary PTC tumors and metastatic lesions. A novel cell type, named Protective EGR1+CD4+T cells, was identified, which might inhibit the formation of the immunosuppressive microenvironment and tumor immune evasion. TIGIT and CD96 might be better therapeutic targets for immunotherapy in PTC patients with LN

- 323 metastasis than PDCD1 (PD-1). These findings would contribute to the further
- 324 understanding of molecular mechanisms resulting in occurrence and development of
- 325 PTC, and provide a theoretical rationale for targeted therapy and immunotherapy.

326

327 Methods

328 Sample collection and clinical information

Seven patients diagnosed with PTC were obtained from Shanghai Tenth People's Hospital. A total of 15 tissue samples (7 tumor tissues, 6 paracancerous tissues, and 2 metastatic LNs) were obtained from 7 PTC patients. Written informed consent was obtained from all patients in the present study. Histological diagnosis of all the samples was confirmed by the pathology department.

334 Tissue digestion and single-cell suspension preparation

335 Cells of each sample were firstly stained with two fluorescent dye, Calcein AM 336 (Thermo Fisher Scientific Cat. No. C1430) and Draq7 (Cat. No. 564904), for 337 precisely determination of cell concentration and viability via BD RhapsodyTM 338 Scanner before single-cell multiplexing labeling. The cells of each sample were 339 sequentially labeled with BD Human Single-Cell Multiplexing Kit (Cat. No. 633781) which utilizing an innovative antibody-oligo technology ²² mainly to provide higher 340 341 sample throughput and eliminate batch effect for single-cell library preparation and 342 sequencing. A set of 12 antibodies in this kit recognize the same universally expressed 343 cell-surface antigen of human cells. Each antibody is conjugated with a Sample Tag, a 344 unique 45-nucleotide barcode sequence. Briefly, cells from each sample were labeled 345 by antibodies with different sample tags respectively and washed twice with BD PharmingenTM Stain Buffer (FBS) (Cat. No. 554656) before pooling all samples 346 together. BD Rhapsody Express system²³ based on Fan et al. was utilized for 347 348 single-cell transcriptome capturation. Pooled samples were then loaded in one BD 349 RhapsodyTM Cartridge that was primed and treated strictly following the 350 manufacturer's protocol (BD Biosciences). Cell Capture Beads (BD Biosciences) 351 were then loaded excessively onto the cartridge to ensure that nearly every micro-well 352 contains one bead, and excess beads were washed away from the cartridge. Viable 353 cells with beads captured in wells were detected in BD RhapsodyTM. Then cells were 354 lysed, Cell Capture Beads were retrieved and washed before performing reverse 355 transcription and treatment with exonuclease I.

356 Library construction and sequencing

Transcriptome and SampleTag information of single-cells was obtained through BD
Rhapsody System. Microbeads-captured single cell transcriptome and SampleTag
sequences were generated into the cDNA library and SampleTag library separately

360 containing cell labels and unique molecular identifiers (UMI) information. Briefly, 361 double-strand cDNA was firstly generated from microbeads-captured single-cell 362 transcriptome through several steps including reverse transcription, second-strand 363 synthesis, end preparation, adapter ligation, and whole transcriptome amplification. 364 Then final cDNA library was generated from double-strand full-length cDNA by 365 random priming amplification with BD Rhapsody cDNA Kit (BD Biosciences, 366 Catalog No.: 633773) and BD Rhapsody Targeted mRNA & AbSeq Amplification Kit 367 (BD Biosciences, Catalog No.: 633774). On the other hand, the SampleTag library 368 was generated from microbeads-captured single-cell SampleTag sequences through 369 several steps including reverse transcription, nest PCR, and final index PCR. Libraries 370 were sequenced on the NovaSeq platform (Illumina).

371 Raw data preprocessing and quality control

372 Raw sequencing reads of the cDNA library and SampleTag library were processed 373 through the BD Rhapsody Whole Transcriptome Assay Analysis Pipeline (Early 374 access), which included filtering by reads quality, annotating reads, annotating 375 molecules, determining putative cells, and generating single-cell expression matrix. 376 Briefly, read pairs with low sequencing quality were firstly removed. The 377 quality-filtered R1 reads were analyzed to identify the sequence of cell labels, UMI 378 sequence, and poly-dT tail sequence, meanwhile the quality-filtered R2 reads were 379 mapped to Genome Reference Consortium Human Build 38 (GRCh38) using STAR 380 (version 2.5.2b) in the reads annotation step. Further adjustment was performed with 381 the recursive substitution error correction (RSEC) and distribution-based error 382 correction (DBEC) algorithms to remove artifact molecules due to amplification bias 383 in the molecules annotation step. Putative cells were distinguished from background 384 noise with second derivative analysis in the putative cell determination step. Finally, 385 putative cells information was combined with molecules adjusted by the recursive 386 substitution error correction and distribution-based error correction algorithms to 387 generate a single-cell expression matrix. The pipeline also determined the sample 388 origin of every single cells via the sample determination algorithm according to the 389 sequencing reads of the SampleTag library. The algorithm classified all putative single 390 cell into three categories: Multiplet, two or more SampleTag exceed their minimum 391 thresholds, indicating more than one actual cell in one micro-well. Undetermined, not 392 enough SampleTag reads for the sample origin. SampleTag01-12, only one 393 SampleTag exceed their minimum thresholds. Among all the output files, Matrix for

394 UMI counts per cell corrected by DBEC and SampleTag annotation result were used

395 for downstream clustering analysis.

396 Dimension reduction and clustering

The R package Seurat²³ was utilized for subsequent analysis. Raw gene expression 397 398 matrices from the cartridge were read into R (version 3.6.0) and converted to Seurat 399 objects. Cells label as "Undetermined" and "Multiple" were excluded in the following 400 analysis. The gene expression matrix was then normalized to the total cellular UMI 401 count. Top 2000 features were selected as highly variable genes for further clustering 402 analysis. In order to reduce dimensionality, PCA was performed based on the highly 403 variable genes after scaling the data with respect to UMI counts. On top of that, the 404 first 9 principal components were chosen for downstream clustering based on the 405 heatmap of principle components, and the elbow plot of principle components to 406 further reduce dimensionality using the UMAP algorithm. The transcriptional markers 407 of each cluster were calculated using the FindAllMarkers function with theMAST 408 package to run the DE testing under the following criteria: $\log 2$ fold change > 0.1; p < 409 0.05; min. percentage > 0.25. Top 500 markers of each cluster were then selected to 410 perform a heatmap plot.

411 Cell annotation and cell type identification

412 Cell populations were matched to cell types based on the expression of known marker

413 genes and previously identified expression signatures ^{23, 24, 25, 26, 27, 28}.

414 Comparison of cell clusters and cell type proportion

The change in the fraction of the different cell types was separately computed for each sample across all clusters, as the fraction of cell in each cluster, out of the total number of cells ²⁹. To assess a statistically significant difference in a fraction of a specific cell type, we performed paired t-test.

419 Significantly dysregulated genes identification

420 We identified differentially expressed genes (DEGs) based on analysis of the MAST

- 421 package using the R. The threshold of DEGs was set as: 1)P-value of F test < 0.05; 2)
- 422 |FC| ≥1.5.

423 External data validation with TCGA datasets

424 CIBERSORTx was used to detect the abundance of cell types identified in the present

- 425 work in bulk RNA-seq data ³⁰. After calculated the relative abundance of each cell
- type, we divided the patients into two groups: high 50% and low 50%. Subsequently,
- 427 Kaplan-Meier analysis was performed using ggsurvplot function of the R package

428 'survminer'. Whether the occurrence of cell types in the different subgroups showed

429 significant differences was evaluated by Wilcoxon test.

430 Functional annotation and enrichment

- 431 GO enrichment and KEGG enrichment of DEGs were performed using Fisher exact
- test with Benjamini-Hochberg multiple testing adjustment. The results were
 visualized using R package. Gene set variation analysis (GSVA) ³¹ was performed
- using 50 hallmark-related gene sets, as described in the GSVA package.
- 435 Identification of hub regulons with SENIC
- In order to investigate the gene regulatory network of different sample groups, we
 utilized SCENIC ³² using the R.
- 438 Cell-cell communication analysis based on ligand-receptor pairs
- 439 Cell-cell communication at the molecular level was analyzed with CellPhoneDB 33 .
- 440 Cell trajectory analysis with Monocle
- 441 In order to reveal the development of malignant cells, we employed Monocle 2, an R
- package designed for single cell trajectories ³⁴. After obtaining the differentially
 expressed genes through differential Gene Test function, the trajectories were
 visualized as 2D tSNE plots.
- 445 Drug-target analysis with TCMID database
- After obtaining the differentially expressed genes, we obtain the targeted relationship
 between drugs and genes from the TCMID database
 (http://www.megabionet.org/tcmid).
- 449 CNV Estimation
- 450 CopyKAT ³⁵ (Copynumber Karyotyping of Tumors), a R package was performed to 451 separate tumor cells from normal cells using high-throughput sc-RNAseq data. Cells 452 with extensive genome-wide copy number aberrations (aneuploidy) are considered as 453 tumor cells, whereas stromal normal cells and immune cells often have 2N diploid or 454 near-diploid copy number profiles.
- 455 Statistical analysis
- Box plots were drawn with the R base package. Hence, the boxes span from the 25th
 to the 75th percentiles. Violin plots were generated by the ggplot2 R package.
 Unpaired two-tailed t-tests was used to compare the difference between two groups.
 While One-way analysis of variance (ANOVA) was used for multiple group
 comparisons.
- 461 Data availability

462	The ca	ncer genome atlas (TCGA) thyroid cancer (THCA) cohort were downloaded	
463	from UCSC XENA (http://xena.ucsc.edu/).		
464 465	Refere	ences	
466	1.	Sung H, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and	
467		mortality worldwide for 36 cancers in 185 countries. (2021).	
468			
469	2.	Fagin JA, Wells SA, Jr. Biologic and Clinical Perspectives on Thyroid Cancer. N Engl J	
470		<i>Med</i> 375 , 1054-1067 (2016).	
471			
472	3.	Yu J, et al. Lymph node metastasis prediction of papillary thyroid carcinoma based on	
473		transfer learning radiomics. Nat Commun 11, 4807 (2020).	
474			
475	4.	Cohen EE, et al. Axitinib is an active treatment for all histologic subtypes of advanced	
476		thyroid cancer: results from a phase II study. <i>J Clin Oncol</i> 26 , 4708-4713 (2008).	
477			
478	5.	Jiao C, et al. REGgamma ablation impedes dedifferentiation of anaplastic thyroid	
479		carcinoma and accentuates radio-therapeutic response by regulating the	
480		Smad7-TGF-beta pathway. <i>Cell Death Differ</i> 27 , 497-508 (2020).	
481			
482	6.	Laetitia G, Sven S, Fabrice J. Combinatorial Therapies in Thyroid Cancer: An	
483		Overview of Preclinical and Clinical Progresses. <i>Cells</i> 9 , (2020).	
484			
485	7.	Hong CM, Ahn BC. Redifferentiation of Radioiodine Refractory Differentiated Thyroid	

486		Cancer for Reapplication of I-131 Therapy. Front Endocrinol (Lausanne) 8, 260
487		(2017).
488		
489	8.	Shen X, et al. Patient Age-Associated Mortality Risk Is Differentiated by BRAF V600E
490		Status in Papillary Thyroid Cancer. 36 , 438-445 (2018).
491		
492	9.	Pozdeyev N, et al. Genetic Analysis of 779 Advanced Differentiated and Anaplastic
493		Thyroid Cancers. 24 , 3059-3068 (2018).
494		
495	10.	Liu R, Bishop J, Zhu G, Zhang T, Ladenson P, Xing MJJo. Mortality Risk Stratification
496		by Combining BRAF V600E and TERT Promoter Mutations in Papillary Thyroid
497		Cancer: Genetic Duet of BRAF and TERT Promoter Mutations in Thyroid Cancer
498		Mortality. 3 , 202-208 (2017).
499		
500	11.	Li Y, et al. CYP2S1 is a synthetic lethal target in BRAF-driven thyroid cancers. 5, 191
501		(2020).
502		
503	12.	Tekpli X, et al. An independent poor-prognosis subtype of breast cancer defined by a
504		distinct tumor immune microenvironment. Nat Commun 10, 5499 (2019).
505		
506	13.	Roehlecke C, Schmidt MHH. Tunneling Nanotubes and Tumor Microtubes in Cancer.
507		<i>Cancers (Basel)</i> 12 , (2020).

508		
509	14.	Oczko-Wojciechowska M, et al. Impact of the Tumor Microenvironment on the Gene
510		Expression Profile in Papillary Thyroid Cancer. <i>Pathobiology</i> 87, 143-154 (2020).
511		
512	15.	Jolly LA, et al. Fibroblast-Mediated Collagen Remodeling Within the Tumor
513		Microenvironment Facilitates Progression of Thyroid Cancers Driven by BrafV600E
514		and Pten Loss. <i>Cancer Res</i> 76, 1804-1813 (2016).
515		
516	16.	Robertson A, et al. Identification of Differential Tumor Subtypes of T1 Bladder Cancer.
517		78 , 533-537 (2020).
518		
519	17.	Mer A, et al. Biological and therapeutic implications of a unique subtype of NPM1
520		mutated AML. 12 , 1054 (2021).
521		
522	18.	Job S, et al. Identification of Four Immune Subtypes Characterized by Distinct
523		Composition and Functions of Tumor Microenvironment in Intrahepatic
524		Cholangiocarcinoma. 72 , 965-981 (2020).
525		
526	19.	Wright G, et al. A Probabilistic Classification Tool for Genetic Subtypes of Diffuse
527		Large B Cell Lymphoma with Therapeutic Implications. 37 , 551-568.e514 (2020).
528		
529	20.	Guruprasad P, Lee Y, Kim K, Ruella MJTJoem. The current landscape of single-cell

530		transcriptomics for cancer immunotherapy. 218 , (2021).
531		
532	21.	Lin W, et al. Single-cell transcriptome analysis of tumor and stromal compartments of
533		pancreatic ductal adenocarcinoma primary tumors and metastatic lesions. Genome
534		<i>Med</i> 12 , 80 (2020).
535		
536	22.	Stoeckius M, et al. Simultaneous epitope and transcriptome measurement in single
537		cells. 14 , 865-868 (2017).
538		
539	23.	Butler A, Hoffman P, Smibert P, Papalexi E, Satija RJNb. Integrating single-cell
540		transcriptomic data across different conditions, technologies, and species. 36,
541		411-420 (2018).
542		
543	24.	Jerby-Arnon L, et al. A Cancer Cell Program Promotes T Cell Exclusion and
544		Resistance to Checkpoint Blockade. 175, 984-997.e924 (2018).
545		
546	25.	Peng J, et al. Single-cell RNA-seq highlights intra-tumoral heterogeneity and
547		malignant progression in pancreatic ductal adenocarcinoma. 29 , 725-738 (2019).
548		
549	26.	Zhang X, et al. CellMarker: a manually curated resource of cell markers in human and
550		mouse. 47 , D721-D728 (2019).
551		

552	27.	Zilionis R, et al. Single-Cell Transcriptomics of Human and Mouse Lung Cancers
553		Reveals Conserved Myeloid Populations across Individuals and Species. 50,
554		1317-1334.e1310 (2019).
555		
556	28.	Luoma A, et al. Molecular Pathways of Colon Inflammation Induced by Cancer
557		Immunotherapy. 182 , 655-671.e622 (2020).
558		
559	29.	Habib N, et al. Disease-associated astrocytes in Alzheimer's disease and aging. Nat
560		<i>Neurosci</i> 23 , 701-706 (2020).
561		
562	30.	Steen CB, Liu CL, Alizadeh AA, Newman AM. Profiling Cell Type Abundance and
563		Expression in Bulk Tissues with CIBERSORTx. <i>Methods Mol Biol</i> 2117, 135-157
564		(2020).
565		
566	31.	Hänzelmann S, Castelo R, Guinney JJBb. GSVA: gene set variation analysis for
567		microarray and RNA-seq data. 14, 7 (2013).
568		
569	32.	Aibar S, et al. SCENIC: single-cell regulatory network inference and clustering. 14,
570		1083-1086 (2017).
571		
572	33.	Efremova M, Vento-Tormo M, Teichmann S, Vento-Tormo RJNp. CellPhoneDB:
573		inferring cell-cell communication from combined expression of multi-subunit

574		ligand-receptor complexes. 15 , 1484-1506 (2020).
575		
576	34.	Trapnell C, et al. The dynamics and regulators of cell fate decisions are revealed by
577		pseudotemporal ordering of single cells. <i>Nat Biotechnol</i> 32 , 381-386 (2014).
578		
579	35.	Gao R, et al. Delineating copy number and clonal substructure in human tumors from
580		single-cell transcriptomes. Nat Biotechnol, (2021).
581		
582	36.	Wolford C, et al. Transcription factor ATF3 links host adaptive response to breast
583		cancer metastasis. 123 , 2893-2906 (2013).
584		
585	37.	Pere H, et al. A CCR4 antagonist combined with vaccines induces antigen-specific
586		CD8+ T cells and tumor immunity against self antigens. 118 , 4853-4862 (2011).
587		
588	38.	Berlato C, et al. A CCR4 antagonist reverses the tumor-promoting microenvironment
589		of renal cancer. 127 , 801-813 (2017).
590		
591	39.	Vinci M, et al. Functional diversity and cooperativity between subclonal populations of
592		pediatric glioblastoma and diffuse intrinsic pontine glioma cells. <i>Nat Med</i> 24,
593		1204-1215 (2018).
594		
595	40.	Rubio-Perez C, et al. Immune cell profiling of the cerebrospinal fluid enables the

596		characterization of the brain metastasis microenvironment. 12 , 1503 (2021).
597		
598	41.	Chen YP, et al. Single-cell transcriptomics reveals regulators underlying immune cell
599		diversity and immune subtypes associated with prognosis in nasopharyngeal
600		carcinoma. <i>Cell Res</i> 30 , 1024-1042 (2020).
601		
602	42.	Cook PJ, et al. Cox-2-derived PGE2 induces Id1-dependent radiation resistance and
603		self-renewal in experimental glioblastoma. <i>Neuro Oncol</i> 18 , 1379-1389 (2016).
604		
605	43.	Jin S, Borkhuu O, Bao W, Yang YT. Signaling Pathways in Thyroid Cancer and Their
606		Therapeutic Implications. <i>J Clin Med Res</i> 8 , 284-296 (2016).
607		
608	44.	Yu J, Baron V, Mercola D, Mustelin T, Adamson ED. A network of p73, p53 and Egr1
609		is required for efficient apoptosis in tumor cells. <i>Cell Death Differ</i> 14 , 436-446 (2007).
610		
611	45.	Jung SN, et al. EGR1/GADD45alpha Activation by ROS of Non-Thermal Plasma
612		Mediates Cell Death in Thyroid Carcinoma. <i>Cancers (Basel)</i> 13 , (2021).
613		
614	46.	Crawford NT, McIntyre AJ, McCormick A, D'Costa ZC, Buckley NE, Mullan PB. TBX2
615		interacts with heterochromatin protein 1 to recruit a novel repression complex to
616		EGR1-targeted promoters to drive the proliferation of breast cancer cells. Oncogene
617		38 , 5971-5986 (2019).

618

619	47.	Ryu JW, et al. Paradoxical induction of growth arrest and apoptosis by EGF via the
620		up-regulation of PTEN by activating Redox factor-1/Egr-1 in human lung cancer cells.
621		<i>Oncotarget</i> 8 , 4181-4195 (2017).
622		
623	48.	Peng WX, et al. Egr-1 regulates irradiation-induced autophagy through Atg4B to
624		promote radioresistance in hepatocellular carcinoma cells. <i>Oncogenesis</i> 6 , e292
625		(2017).
626		
627	49.	Gonzalez-Navajas JM, et al. The Impact of Tregs on the Anticancer Immunity and the
628		Efficacy of Immune Checkpoint Inhibitor Therapies. Front Immunol 12, 625783 (2021).
629		
630	50.	Antonelli A, Ferrari SM, Fallahi P. Current and future immunotherapies for thyroid
631		cancer. Expert Rev Anticancer Ther 18, 149-159 (2018).
632		
633	51.	Lv M, et al. Manganese is critical for antitumor immune responses via cGAS-STING
634		and improves the efficacy of clinical immunotherapy. <i>Cell Res</i> 30 , 966-979 (2020).
635		
636	52.	Marigo I, et al. T Cell Cancer Therapy Requires CD40-CD40L Activation of Tumor
637		Necrosis Factor and Inducible Nitric-Oxide-Synthase-Producing Dendritic Cells.
638		<i>Cancer Cell</i> 30 , 377-390 (2016).
639		

640	53.	Cortez MA, et al. Bone morphogenetic protein 7 promotes resistance to
641		immunotherapy. <i>Nat Commun</i> 11, 4840 (2020).
642		
643	54.	Huang Z, et al. IL-27 promotes the expansion of self-renewing CD8(+) T cells in
644		persistent viral infection. J Exp Med 216, 1791-1808 (2019).
645		
646	55.	Lei Q, Wang D, Sun K, Wang L, Zhang Y. Resistance Mechanisms of Anti-PD1/PDL1
647		Therapy in Solid Tumors. Front Cell Dev Biol 8, 672 (2020).
648		
649	56.	Severson JJ, et al. PD-1+Tim-3+ CD8+ T Lymphocytes Display Varied Degrees of
650		Functional Exhaustion in Patients with Regionally Metastatic Differentiated Thyroid
651		Cancer. <i>Cancer Immunol Res</i> 3 , 620-630 (2015).
652		
653	57.	Tirosh I, et al. Dissecting the multicellular ecosystem of metastatic melanoma by
654		single-cell RNA-seq. 352 , 189-196 (2016).
655		
656	58.	Liu Y, Cao X. Characteristics and Significance of the Pre-metastatic Niche. Cancer
657		<i>Cell</i> 30 , 668-681 (2016).
658		
659		

660 Acknowledgments

- 661 This work was supported by Shanghai Municipal Commission of Health and Family
- 662 Planning Commission (2019SY062).

663 Authors' contributions

- 664 BH J, ZS W, ZQ Y, and Y R conceived and designed the study. WL J, CG T, and ZS
- 665 W contributed to carry out the experiments. BH J, CG H, and C Y contributed to data
- analysis. ZS W, BH J, and ZQ Y wrote the manuscript. YR, WL J, and ZQ Y revised
- the paper. All authors read and approved the final manuscript.

668 **Competing Interests**

669 The authors have declared that no competing interest exists.

670

671 Figure legends

672 Fig.1 Overall design and single cell atlas in papillary thyroid carcinoma. a 673 Workflow diagram showing the processing of samples. **b** UMAP plots of total cells, 674 colored by the sample origin (tumor, metastatic LN, and paracancerous tissues). c 675 Expression of cell marker genes. d Changes in frequency of multiple cell types and 676 clusters in tumors and metastatic LNs. Asterisks on the left of the vertical line denote 677 statistically significant differences between tumors and paracancerous tissues while 678 asterisks on the right are used to show statistically significant differences between 679 metastatic LNs and tumors. *p < 0.05, **p < 0.01, ***p < 0.001, two-tailed t-tests. e 680 Interaction network constructed by CellPhoneDB. Size of circles and color of arrows 681 represent interaction counts, and brighter color and larger size mean more interaction 682 with other cell types.

683 Fig.2 Identification of malignant cells in epithelial cells. a Clonal substructure of 684 epithelial cells delineated by clustering single-cell copy number profiles inferred from 685 scRNA-seq data by CopyKAT. **b** UMAP plots of total epithelial cells, colored by the 686 sample origin (left and middle). Cell counts of malignant and non-malignant cells, 687 colored by the sample origin (right). c Differentiation trajectory of epithelial cells, 688 with each color coded for pseudotime (right) and clusters (left). **d** Heatmap of the area 689 under the curve (AUC) scores of transcription factor (TF) motifs estimated by 690 SCENIC. Shown are differentially activated motifs in malignant and non-malignant 691 cells, respectively. e High level of CREM and ETS1 predicted good prognosis in the 692 TCGA THCA cohort. Log-rank p value < 0.05 was considered as statistically 693 significant. **f** BHLHE40 was highly expressed in LN samples (p < 0.001). **g** GO 694 enrichment analysis of up-regulated genes in malignant lymphatic metastasis cells.

695 Fig.3 Heterogeneity of T cell populations. a UMAP plot of T cells color-coded by 696 their associated cell types. **b** Dotplot of cell markers; sizes of dots represent 697 abundance while color represents expression level. c Potential new cell markers. d 698 The protein level of EGR1 was decreased in tumor tissues compared with normal 699 tissues. e High level of EGR1 predicted good prognosis in the TCGA THCA cohort. 700 **f** Changes in frequency of multiple cell types in T cell populations. Asterisks on the 701 left of the vertical line denote statistically significant differences between tumors and 702 paracancerous tissues while asterisks on the right are used to show statistically 703 significant differences between LNs and tumors. *p < 0.05, **p < 0.01, ***p < 0.001, 704 two-tailed t-tests. g Violin of cell abundance predicted from the TCGA THCA cohort

705 by CIBERSORTx.

Fig.4 Novel cell type, named Protective EGR1⁺CD4⁺ T cells, and potential therapeutic targets. a Differences in 50 hallmark pathway activities scored with GSVA software. Shown are t values calculated by a linear model. **b** Heatmap of the AUC scores of TF motifs estimated by SCENIC. Shown are differentially activated motifs in CD4⁺ Tregs and Protective EGR1⁺CD4⁺T cells, respectively. **c-d** Heatmap of positive and negative immune checkpoint expression on T cells in LNs (**c**) and tumors (**d**), respectively.

713 Fig.5 Heterogeneity of Myeloid cell populations. a UMAP plot of Myeloid cells 714 color-coded by their associated cell types. **b** Dotplot of cell markers; sizes of dots 715 represent abundance while color represents expression level. c Changes in frequency 716 of multiple cell types and clusters in Myeloid cell populations. Asterisks on the left of 717 the vertical line denote statistically significant differences between tumors and 718 paracancerous tissues while asterisks on the right are used to show statistically 719 significant differences between LNs and tumors. *p < 0.05, **p < 0.01, ***p < 0.001, 720 two-tailed t-tests. **d** Violin of Macrophage abundance predicted from the TCGA 721 THCA cohort by CIBERSORTx. e Kaplan-Meier curves of DFS based on the 722 percentage of dendritic cells in the TCGA database.

Fig.6 M2 macrophages were strongly enriched in the tumor tissues. a-b Heatmap shows the differences in pathway activities scored by GSVA between different sample origins. Shown are t values calculated by a linear model. **c** Heatmap of the AUC scores of TF motifs estimated by SCENIC. Shown are differentially activated motifs in subtypes of Myeloid cells, respectively. **d** Violin plot shows the expression level of related genes in Macrophage subclusters (left) and cytokines in dendritic cells (right).

729 Supplementary information

Supplementary Fig. 1 Identification of major cell types in PTC. UMAP plots for
expression of the marker genes from Fig 1c and of additional marker genes for major
cell types. For B cells: CD19, MS4A1(CD20), CD38, CD79A, and CD79B; for CD4⁺
T cells: CD4 and CD3D; for CD8⁺T cells: CD8A and CD3D; for Endothelial cells:
CD34 and CD31(PECAM); for Epithelial cells: EPCAM and KRT18; for Fibroblasts:
COL1A1; for Myeloid cells: CD14, CD86, ITGAX, CD80, CD83, and ITGAM; for
Naive T cells: CCR7 and CD3D; for NKT cells: NKG7 and CD3D; for Plasma cells:

737 CD79A and SDC1.

738 Supplementary Fig. 2 Overview of the 28,205 single cells from tumors, 739 paracancerous tissues, and metastatic LNs. a The fraction of multiple cell types 740 originating from 7 tumor tissues, 6 paracancerous tissues, and 2 metastatic LNs. b The 741 fraction of multiple cell types originating from 15 samples. c The number of multiple 742 cell types. d Box plots of the number of transcripts.

743 Supplementary Fig. 3 Potential new cell markers in T cells.

744 Supplementary Fig. 4 DEGs related to CD4⁺Tregs and transitional EGR1⁺CD4⁺T

cells. a The expression of the maker genes in the TCGA THCA cohort. b Differences in 50 hallmark pathway activities scored with GSVA software. Shown are t values calculated by a linear model. c DEGs related to Protective EGR1⁺CD4⁺ T cells in tumors and LNs, respectively. d DEGs related to CD4⁺Tregs in tumors and LNs, respectively. Upregulated genes and downregulated genes were colored in red and blue, respectively. FC \geq 2.

- 751 Supplementary Fig. 5 External verification based on the TCGA data. a High level
- of Macrophage abundance predicted poor prognosis in the TCGA THCA cohort, and
- neutrophil was the opposite. **b** Kaplan-Meier survival curve for IFITM2.





cardi

SPREA FRANK GERMANNE 18101 19101 19101 19101















С













Macrophages