Single-cell RNA sequencing demonstrates the intratumoral heterogeneityof papillary thyroid carcinoma

Zhengshi Wang[1], Youlutuziayi Rixiati[2], Wenli Jiang[3], Caiguo Huang[3], Binghua Jiao[3], Chuangang Tang[4], Zhiqiang Yin[1], Chen Ye[5]

[1] Thyroid Center, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai, P.R.China; Shanghai Center for Thyroid Diseases, Shanghai, 200072, P.R.China;

[2] Department of Pathology, Soochow University Medical School, Suzhou, 215123, P.R.China;

[3] Department of Biochemistry and Molecular Biology, College of Basic Medical, Navy Medical University, Shanghai, 200433, P.R.China;

[4] Department of Breast Surgery, Xuzhou Central Hospital, The Affiliated Xuzhou Hospital of Medical College of Southeast University, Xuzhou, 221009, P.R.China;

[5] Department of Urology, Changhai Hospital, Navy Medical University, Shanghai, 200433, P.R.China.

Abstract

Papillary thyroid carcinoma(PTC) is the most common thyroid malignancy, to investigate the intratumoral heterogeneity of PTC, we analyzed single-cell RNA-sequencing data and identified 10 major cell types from primary papillary thyroid carcinoma, lymph metastatic, or paired normal thyroid tissue samples. In this study, we verified that the increase in the proportion of CD4+Tregs may be key factor responsible for the immunosuppressive property of PTC. Inhibitory checkpoints, such as TIGIT and CD96 may be better targets for immune therapy in lymph metastatic papillary thyroid carcinoma. Our results will further the understanding of the heterogeneity among papillary thyroid carcinoma and provide an essential resource for drug discovery in the future.

Keywords

Papillary thyroid carcinoma, Single-cell RNA sequencing, Tumor microenvironment

Introduction

According to the cancer statistics, thyroid cancer causes 586,202 cases worldwide, ranking 11th in incidence. The global incidence of women is 10.1 per 100,000 people, 3.27 times that of men[1]. In recent decades, the incidence of thyroid cancer has been increasing in many countries. This is mainly due to the increase in the detection rate of papillary thyroid cancer (PTC) through improved detection and diagnosis[2].

Approximately 85% of patients with thyroid cancer are PTC, which is the most common thyroid malignancy[3]. Although PTC is considered to be an indolent tumor, some of the cancer cells will metastasize to the lymph nodes around the thyroid, including central lymph node metastasis (LNM) and cervical LNM. Generally, LNM appears first in the central area, and then in the side areas. LNM is a prognosis of PTC, the scope and method of surgery are important indicators, and it is also an important risk factor for patients with high recurrence rate and low survival rate[4].

Efforts to understand the progression and metastasis of thyroid cancer have mainly focused on the analysis of cancer cells using genetic aberrations[5, 6, 7, 8]. However, the complex and dynamic characteristics around the tumor also affect the progression and metastasis. The analysis of the unique type of tumor microenvironment in advanced cancers can reveal the key elements involved in the susceptibility of tumor-induced immunological changes, and these elements can be used in new immunotherapy strategies.

Genomic and transcriptomics studies have revealed driver mutations, abnormal regulatory programs, and disease subtypes in major human tumors[9, 10, 11, 12]. However, these studies

rely on profiling techniques that can measure a large number of tumors, which limits their ability to capture intratumoral heterogeneity. Compared with bulk sequencing, single-cell RNA sequencing (scRNA-seq) provides the opportunity to sample the entire transcriptome of a single cell, thereby reducing Cell organization is independent of any previous assumptions about surface markers or species conservation, and there is no strict limit on the number of marker genes that can be used for analysis[13].

Single cell atlas of Papillary thyroid carcinoma (PTC) remains to be revealed. Here, we investigate primary Papillary thyroid carcinoma tumors and matched LNs to better understand intra-tumoral heterogeneity, invasion, and metastasis. Transcriptional profiles for 28,205 cells from 15 samples revealed expression programs that distinguish diverse malignant, stromal, and immune cells.

Material and methods

Sample collection and clinical information

Four patients diagnosed with Papillary thyroid carcinoma were obtained from Shanghai Tenth People's Hospital(Supplementary Data 1). The average age was 37.5 years old and 50% of them were female. A total of 15 tissue samples (7 tumor tissue, 6 distant normal tissue, and 2 metastatic lymph node) were obtained from 4 Papillary thyroid carcinoma patients.

Ethics statement

This study has complied with all relevant ethical regulations for the human participants.

Tissue digestion and single cell suspension preparation

Cells of each sample were firstly stained with two fluorescent dye, Calcein AM (Thermo Fisher Scientific Cat. No. C1430) and Draq7 (Cat. No. 564904), for precisely determination of cell concentration and viability via BD Rhapsody™ Scanner before single cell multiplexing labelling. Cell of each sample were sequentially labeled with BD Human Single-Cell Multiplexing Kit (Cat. No. 633781) which utilizing an innovative antibody-oligo technology [14] mainly to provide higher sample throughput and eliminate batch effect for single cell library preparation and sequencing. A set of 12 antibodies in this kit recognize the same universally expressed cell-surface antigen of human cells. Each antibody is conjugated with a Sample Tag, a unique 45-nucleotide barcode sequence. Briefly, cells from each sample were labeled by antibodies with different sample tags respectively and washed twice with BD Pharmingen™ Stain Buffer (FBS) (Cat. No. 554656) before pooling all samples together. BD Rhapsody Express system[15]based on Fan et al was utilized for single cell transcriptome capturation. Pooled samples were then loaded in one BD Rhapsody™ Cartridge that was primed and treated strictly following the manufacturers protocol (BD Biosciences). Cell Capture Beads (BD Biosciences) were then loaded excessively onto the cartridge to ensure that nearly every micro-well contains one bead, and excess beads were washed away from the cartridge. Viable cells with beads captured in wells were detected in BD Rhapsody™ Scanner and the total number was 10871 and cell multiplet rate was 3.00%. Then cells were lysed, Cell Capture Beads were retrieved and washed prior to performing reverse transcription and

treatment with exonuclease I.

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 cDNA library and SampleTag library separately containing cell labels and unique molecular identifiers (UMI) information. Briefly, double strand cDNA was firstly generated from microbeads-captured single cell transcriptome through several steps including reverse transcription, second strand synthesis, end preparation, adapter ligation and whole transcriptome amplification. Then final cDNA library was generated from double strand full length cDNA by random priming amplification with BD Rhapsody cDNA Kit (BD Biosciences, Catalog No.: 633773) and BD Rhapsody Targeted mRNA & AbSeq Amplification Kit (BD Biosciences, Catalog No.: 633774). On the other hand, SampleTag library was generated from microbeads-captured single cell SampleTag sequences through several steps including reverse transcription, nest PCR and final index PCR. All the libraries were sequenced in a PE150 mode (Pair-End for 150bp read) on the NovaSeq platform (Illumina).

Raw data preprocessing and quality control

Raw sequencing reads of cDNA library and SampleTag library were processed through the BD Rhapsody Whole Transcriptome Assay Analysis Pipeline (Early access), which included filtering by reads quality, annotating reads, annotating molecules, determining

putative cells and generating single-cell expression matrix. Briefly, read pairs with low sequencing quality were firstly removed. The quality-filtered R1 reads were analyzed to identify sequence of cell labels, UMI sequence, and poly-dT tail sequence, meanwhile the quality-filtered R2 reads were mapped to Genome Reference Consortium Human Build 38 (GRCh38) using STAR (version 2.5.2b) in reads annotation step. Further adjustment was performed with the recursive substitution error correction (RSEC) and distribution-based error correction (DBEC) algorithms to remove artifact molecules due to amplification bias in molecules annotation step. Putative cells were distinguished from background noise with second derivative analysis in putative cell determination step. Finally, putative cells information was combined together with molecules adjusted by the recursive substitution error correction and distribution-based error correction algorithms to generate a single-cell expression matrix. The pipeline also determined sample origin of every single cells via Sample determination algorithm according to the sequencing reads of SampleTag library. The algorithm classified all putative single cell into three categories: Multiplet, two or more SampleTag exceed their minimum thresholds, indicating more than one actual cell in one micro-well. Undetermined, not enough SampleTag reads for the sample origin. SampleTag01-12, only one SampleTag exceed their minimum thresholds. . Among all the output files, Matrix for UMI counts per cell corrected by DBEC and SampleTag annotation result were used for downstream clustering analysis.

Dimension reduction and clustering

The R package Seurat[15] was utilized for subsequent analysis. Raw gene expression

matrices from the cartridge were read into R (version 3.6.0) and converted to Seurat objects. Cells label as “Undetermined” and “Multiple” were excluded in the following analysis. The gene expression matrix was then normalized to the total cellular UMI count. Top 2000 features were selected as highly variable genes for further clustering analysis. In order to reduce dimensionality, PCA was performed based on the highly variable genes after scaling the data with respect to UMI counts. On top of that, the first 50 principle components were chosen for downstream clustering based on heatmap of principle components, Jackstraw plot, and elbow plot of principle components to further reduce dimensionality using the tSNE algorithm. The transcriptional markers of each cluster were calculated using the FindAllMarkers function with the Wilcoxon Rank-Sum Test under the following criteria: log2 fold change > 0.25; $p < 0.01$; min. percentage > 0.25. Top markers of each cluster were then selected to perform heatmap plot.

Cell annotation and cell type identification

Cell populations were matched to cell types based on the expression of known marker genes and previously identified expression signatures[15, 16, 17, 18, 19, 20].

Comparison of cell clusters and cell type proportion

The change in fraction of the different cell populations (clusters) was separately computed for each sample across all clusters, as the fraction of cell in each cluster, out of the total number of cell[21]. To assess statistically significant changes in a fraction of a specific population, we first performed the Shapiro–Wilk test for normality (using the shapiro.test function in R), followed by paired t-test (for normally distributed populations) and an

additional paired nonparametric Wilcoxon test (using t.test and wilcox.test functions in R).

Significantly dysregulated genes identification

We identified differentially expressed genes (DEGs) based on analysis of variance (ANOVA) using the R. In brief, only genes that met these criteria were considered as DEGs: 1)P-value of F test < 0.05; 2) |FC| ≥2.

External data validation with TCGA datasets

CIBERSORTx is a new machine-learning method designed for detecting the abundance of cell types in bulk RNA-seq data; this method has been called "digital cytometry" by the authors[22]. To evaluate the roles of the cell clusters identified in the present work, we used CIBERSORTx to analyze TCGA expression matrix, which was first normalized to obtain log2(TPM+1) values. As the signature matrix was constructed, we calculated the relative abundance of each cell cluster and then divided the patients into two groups: high 50% and low 50%. Subsequently, Kaplan-Meier analysis was performed using ggsurvplot function of the R package 'survminer'. One-way ANOVA was used to evaluate whether the occurrence of cell clusters in the different subgroups showed significant differences.

Functional annotation and enrichment

GO enrichment and KEGG enrichment of cluster markers were performed using KOBAS software with Benjamini-Hochberg multiple testing adjustment, using the top 20 markers gene of cluster. The results were visualized using R package.

Gene set variation analysis (GSVA) [23]was performed using 50 hallmark gene sets obtained from the molecular signature database using default sets, as described in the GSVA package.

Identification of hub regulons with SENIC

In order to investigate the gene regulatory network of the samples, we utilized by SCENIC[24], a python-based algorithm for the reconstruction of the transcriptional states and regulatory networks from the scRNA-seq data,37 in the Jupyter notebook software (version1.0.0).

Cell-cell communication analysis based on ligand-receptor pairs

CellPhoneDB[25] is a Python-based computational analysis tool; it enables the analysis of cell-cell communication at the molecular level. A website version was also provided for the analysis of relatively small datasets (https://www.cellphonedb.org/).

Cell trajectory analysis with Monocle

In order to reveal the changes of immune cells in the tumor-educating process, we employed Monocle 2, an R package designed for singlecell trajectories[26]. The following parameters were set: mean expression R0.125, num_cells_expressed R10, qval <0.01 (differentialGeneTest function). The trajectories were visualized as 2D tSNE plots, while dynamical expression heatmaps were constructed using the plot_pseudotime_heatmap function.

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) .

CNV Estimation

CopyKAT[27](Copynumber Karyotyping of Tumors) is a computational tool using integrative Bayesian approaches to identify genome-wide aneuploidy at 5MB resolution in single cells to separate tumor cells from normal cells, and tumor subclones using high-throughput sc-RNAseq data. Cells with extensive genome-wide copy number aberrations (aneuploidy) are considered as tumor cells, wherease stromal normal cells and immune cells often have 2N diploid or near-diploid copy number profiles. In this study, we inferred tumor subclones from using the R package.

Statistical analysis

Box plots were generated using the R base package and default parameters. Hence, the boxes span the interquartile range (IQR; from the 25th to the 75thpercentiles), with the centerline corresponding to the median. Lower whiskers represent the data minimum or the 25th percentile minus 1.5×IQR, whichever is greater. Upper whiskers represent the data maximum or the 75th percentile plus 1.5×IQR (lower), whichever is lower.

Violin plots were generated using the beanplot R package, and data distribution band width was estimated by kernel density estimation, as per the built-in ‘nrd0’ option.

Comparisons between two groups were done using unpaired two-tailed t-tests. One-way analysis of variance (ANOVA) with Tukey's multiple comparisons tests were used for multiple group comparisons.

Results

Single cell atlas and heterogeneity of papillary thyroid carcinoma (PTC)

A total of 15 samples from four papillary thyroid carcinoma patients were involved in this study. After quality filtering, 30986 genes were detected, of these, 13390 cells originated from 7 primary papillary thyroid carcinoma, 2869 cells from 2 lymph metastatic, and 11945 cells from 6 paired normal thyroid tissue samples (Fig. 1A,B). Subsequently, we partitioned the cells into 26 clusters, and classified those clusters into 10 major cell types based on known markers described in previous studies: B cells (CD19+MS4A1+CD38+CD79A+CD79B+); CD4+T cells (CD3D+CD4+); CD8+T cells (CD3D+CD8A+); endothelial cells (CD31+CD34+); epithelial cells (EPCAM+KRT18+); fibroblasts (COL1A1+); myeloid cells (CD14+CD86+ITGAX+CD80+CD83+ITGAM+); naive T cells (CD3D+CCR7+); NKT cells (CD3D+NKG7+); and plasma cells (CD79A+SDC1+) (Fig. 1C, Supplementary Fig. 1).

By calculating the frequency of the cell types, we revealed similar cellular landscapes in primary papillary thyroid carcinoma and normal thyroid tissue samples, with differences in the proportions of several cell types ($P<0.01$; Fig. 1D). There were major differences in immune cells in primary tumor relative to normal tissue, the frequency of NKT cells/plasma cells increased in primary tumor ($P<0.01$; Fig. 1D), while that of CD4+ T

cells/B cells decreased. The most striking changes between lymph metastatic and primary tumor were observed in CD8+T cells. At the cluster level, there were also differences in the frequency of clusters within the same cell group, take myeloid cells as an example, we observed an increase in the frequency of cluster 21 during the disease progression, at the same time, decrease in the proportion of cluster 25 was also be identified. These results demonstrate that the disease progression is associated with major changes in immune cells , heterogeneity is common between samples and cell populations.

To investigate the interaction network in the tumor microenvironment in papillary thyroid carcinoma, CellphoneDB was be used to calculate potential ligand-receptor pairs in Cells. Cytoscape was performed to visualized the cell interaction. We found that myeloid cells and T cell related cells possessed the most interaction pairs with cells from other lineages (Fig. 1E), demonstrated the dominant role of myeloid cells and T cells.

Identification of malignant cells in epithelial cells

In order to distinguished malignant from non-malignant cells within the epithelial cells, CopyKAT was performed to identified papillary thyroid carcinoma genome alterations. CNVs accumulated in most lymph metastatic papillary thyroid carcinoma , despite the heterogeneity, almost all malignant cells possessed deletions from chromosomes 16 and 19 and amplifications in chromosomes 13(Fig. 2A,B). We distinguished 1679 nonmalignant cells, and 450 cells were identified as malignant cells, of which malignant cells were mainly composed of lymphatic metastasis samples, but we also found very few

malignant cells in paired normal thyroid tissue samples(Fig. 2B). In order to support the identification results of the malignant cells, pseudotime trajectory analysis was performed. The results showed that malignant cells was present at the end of the differentiation trajectory(Fig. 2C).

We further characterized the functions of differential genes between lymphatic metastasis and non-metastasis in malignant cells by comparing pathway activities, pathways involved in cellular component assembly, protein containing complex assembly and regulation of microtubule motor activity were relatively upregulated in lymphatic metastasis malignant cells(Fig. 2D).In order to screen for key genes related malignant cells , we conducted SCENIC analysis , Via SCENIC analysis, we identified ATF3 and BHLHE40 motifs were highly activated in malignant cells(Fig. 2E).These results provide potential targets for suppressing cells to develop into malignant cells.

Increased CD4+Tregs in tumor and lymph metastatic PTC and decreased transitional EGR1+CD4+T cells

For discussing the heterogeneity among T cells, 17690 T cells were clustered into 10 subclusters based on known markers described in previous studies: CD4+Tregs(CD3D+CD3E+CD3G+CD4+FOXP3+); Cytotoxic CD8+T cells(CD3D+CD3E+CD3G+CD8A+NKG7+); Exhausted CD4+T cells(CD3D+CD3E+CD3G+CD4+PDCD1+CTLA4+TIGIT+LAG3+); Follicular helper (Tfh) T cels(CD3D+CD3E+CD3G+ICOS+); Naïve CD4+T cells(CD3D+CD3E+CD3G+CD4+CCR7+); Naïve T cells(CD3D+CD3E+CD3G+CCR7+);

NKT cells(CD3D+CD3E+CD3G+NKG7+GZMB+GNLY+) ; Pre-exhausted CD8+ T cells(CD3D+CD3E+CD3G+CD8A+PDCD1+CTLA4+TIGIT+LAG3+); Proliferating CD8+T cells(CD3D+CD3E+CD3G+CD8A+MKI67+TOP2A+CDK1+PCNA+); Transitional EGR1+CD4+T cells(CD3D+CD3E+CD3G+CD8A+MKI67+TOP2A+CDK1+PCNA+)(Fig. 3A,B). In addition to the marker used above, we also found some other makers that can be used for cell population identification In the future, for example, TYMS can be used for Pre-exhausted CD8+ T cells identification, RTKN2 can be used for Pre-exhausted CD4+Tregs identification(Fig. 3C). Through cell frequency analysis, we found an interesting phenomenon, the proportion of CD4+Tregs continues to increase during the development of the disease, while the trend of Transitional EGR1+CD4+T cells is the opposite(Fig. 3D). The trend was consistent with TCGA data calculated by cibersortx(Fig. 3E). The results indicate that both of CD4+Tregs and Transitional EGR1+CD4+T cells may have an important impact on disease progression, and our newly identified cell population(Transitional EGR1+CD4+T cells) may be antagonistic to the CD4+Tregs.

As shown in Fig. 4A, the differentially expressed genes derived from transitional EGR1+CD4+T cells could be identified, 31 mRNAs were found to be up-regulated and 50 were down-regulated in primary tumor compared to normal thyroid tissue samples(Fig. 4A). We further characterized the functions of CD4+Tregs and Transitional EGR1+CD4+T cells by comparing pathway activities. in primary tumor, CD4+Tregs and Transitional EGR1+CD4+T cells exhibited large different activated pathways, pathways involved in epithelial-mesenchymal transition and hypoxia were relatively upregulated in the CD4+Tregs, whereas G2M checkpoint was down regulated in Transitional EGR1+CD4+T

cells(Fig. 4B). In lymph metastatic, mitotic spindle and oxidative phosphorylation were also down regulated in Transitional EGR1+CD4+T cells(Supplementary Fig. 4).In order to screen for key genes related to tumor occurrence and development in CD4+Tregs and transitional EGR1+CD4+T cells，we conducted SCENIC analysis，Via SCENIC analysis, we identified essential motifs in CD4+Tregs and transitional EGR1+CD4+T cells. CREM, FOSL2, ATF3, REL, and HES1 are CD4+Tregs-specific motifs. BCLAF1 and ETS1 motifs were highly activated in Transitional EGR1+CD4+T cells(Fig. 4C).These results provide potential targets for inhibiting or reversing the formation of the immunosup-pressive microenvironment.

We extracted small molecule drugs that can target differential genes related to Transitional EGR1+CD4+T cells from the TCMID database(Fig. 4D). Subsequently, we examined the immune checkpoints in the cell clusters (Fig. 4E). Notably, in lymph metastatic, we found that inhibitory checkpoints,Notably, we found that inhibitory checkpoints, TIGIT was upregulated in cells from Exhausted CD4+T cells, CD96 was upregulated in cells from Pre−exhausted CD8+T cells, since these molecules are markers of T cell exhaustion, these data indicated that cells from Exhausted CD4+T cells and Pre−exhausted CD8+T cells were exhausted in the tumor microenvironment. This result is consistent with our previous cell population classification results. Since the expression of TIGIT and CD96 was much higher than that of PDCD1 in exhausted T cell types, our data indicated that TIGIT and CD96 may be better targets for immune therapy in lymph metastatic papillary thyroid carcinoma. Opposite of this, PDCD1 maybe the most suitable targets of immunotherapy in primary papillary thyroid carcinoma.

Neutrophils cells are extremely reduced in the tumor

For discussing the heterogeneity among myeloid cells, 1607 T cells were clustered into 5 subclusters based on known markers described in previous studies:Macrophage(MSR1+); Dendritic cells(CD83+); monocyte(FCGR3A+); neutrophil(S100A8+S100A9+IFITM2+FCGR3B+CXCR1+CXCR2+); myeloid-derived suppressor cells(HLA.DRB5+) (Fig. 5A,B). Through cell frequency analysis, we found decreased proportion of neutrophils in tumors, however the proportion of macrophage continues to increase during the development of the disease(Fig. 5C), The trend was consistent with TCGA data calculated by cibersortx(Fig. 5D). We also found the DFS of the low-dendritic cells group was longer than that of the high-Dendritic cells group in the TCGA THCA(Fig. 5E), the occurrence of dendritic cells was related to bad prognosis in the case of patients from TCGA THCA cohort . The results suggested that macrophage, neutrophils, and Dendritic cells may be more relevant to papillary thyroid carcinoma.

We further characterized the functions of myeloid cells sub types by comparing pathway activities. in primary tumor, pathways involved in mtorc1 signaling and E2F targets were relatively upregulated in the myeloid-derived suppressor cells, whereas The hallmark related pathways in neutrophils were generally down-regulated(Fig. 6A), which is consistent with the changes in cell proportions we observed earlier(Fig. 5C). However, in lymphatic metastasis samples, myeloid-derived suppressor cells were down-regulated in some hallmark related pathways(Fig. 6B). The results show that in the process of disease progression, myeloid cells are generally up-regulated in hallmark related pathways, but

there is a difference between lymphatic metastasis and primary cancer.In order to screen for key genes related to tumor occurrence and development in myeloid cells , we conducted SCENIC analysis ,by means of SCENIC analysis, we identified essential motifs in myeloid cells sub types. The overall regulatory factors in neutrophil were also inconsistent with other cell types(Fig. 6C). we recognized that the activity of a lot of motifs, such as YY1 and IRF5, were downregulated, while activation of the STAT3, ZNF143, and HCFC1 motifs led to the reduction of neutrophil .

As shown in Fig 6D, cells from cluster 8 uniquely expressed the M1 macrophages marker CCL5, whereas M2 macrophages markers, including CTSB, CTSD, and FN1, were expressed in cluster 0.The proportion of cluster 0 is high in tumor and lymph samples, although the difference has not reached a significant level(Fig. 5C). All of this evidence certified that cluster 0 represented an M2-like TAM cluster, the increase in the proportion of cell subtypes may be related to disease progression. Cells from cluster 9 highly express the cytokine CCL22(Supplementary Fig. 6), cytokine CCL22 binds to CCR4 which expressed on the cell membrane of Tregs and has strong chemotaxis to Tregs. The results indicate that the dendritic cell cluster 9 can recruit tregs into the tumor area, thereby acting as immunosuppressive effect.

Discussion

At present, people's understanding of the progression and metastasis of thyroid cancer is mainly focused on the genetically altered cancer cell map. However, the influence of

tumor microenvironment on disease progression and metastasis has also been confirmed in other diseases. Therefore, studying the intra-tumor heterogeneity of tumors will help to reveal the key factors related to tumor-induced immune changes, thereby deepening the understanding of papillary thyroid cancer, and providing new targets for cancer treatment.

In this study, we discovered a new cell subpopulation, Transitional EGR1+CD4+T cells. By comparing the proportion of cells from different tissues and pathway enrichment analyzing , we speculated that its function is opposite to Ttrg. Transitional EGR1+CD4+T cells specific high expression of EGR1 , the protein encoded by this gene belongs to the EGR family of C2H2-type zinc-finger proteins. It is a nuclear protein and functions as a transcriptional regulator. The products of target genes it activates are required for differentitation and mitogenesis. Studies suggest this is a cancer suppressor gene. CD4+ regulatory T (Treg) cells, which express the X chromosome-linked, linage specific transcription factor Foxp3, are potent immunosuppressive cells and can serve as brakes during immune responses. However, Treg cells play an opposite role in cancer immunity as Treg cells recruited in tumor tissues become accomplices that help cancer cells escape from immunological surveillance. These results indicated that Transitional EGR1+CD4+T cells is a protective subgroup.

In the identification of malignant cells, we initially used a commonly used method, inferCNV for analysis. The results showed that inferCNV could not significantly separate the malignant components in the epithelial cells of this study, and CopyKAT could do it. Provide a new method idea for follow-up similar research. Based on the results of CopyKAT malignant cell identification, we found that there are also a small number of

malignant cells in the adjacent samples. This may be because the pre-tumor microenvironment already exists in the adjacent tissues.

Figure legends

Figure 1 . Overall design and single cell atlas in thyroid carcinoma. A Workflow diagram showing the processing of samples. **B** UMAP plots of total cells, colored by sample origin (primary papillary thyroid carcinoma, lymph metastatic or paired normal thyroid), the corresponding patient, the associated cell type. **C** Expression of cell marker genes. Additional cell marker genes were shown in Supplementary Fig. 1. **D** Changes in frequency of multiple cell types and clusters in primary papillary thyroid carcinoma and lymph metastatic. For each cell type and cluster, the average fraction of cells from each patient group is shown,

Asterisks on the left of the vertical line denote statistically significant differences between primary papillary thyroid carcinoma and normal thyroid tissue, meanwhile, asterisks on the right is used to show statistically significant differences between lymph metastatic andprimary papillary thyroid carcinoma. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, two-tailed t-tests. **E** Interaction network constructed by CellPhoneDB; size of circles and color of arrows represent interaction counts, brighter color and larger size means more interaction with other cell types.

Figure 2 . Identification of malignant cells in epithelial cells. A Clonal substructure of epithelial cells delineated by clustering single-cell copy number profiles inferred from scRNA-seq data by CopyKAT. **B** UMAP plots of total epithelial cells, colored by the associated cell type, sample origin (primary papillary thyroid carcinoma, lymph metastatic or paired normal thyroid), the frequency of cell types and clusters in malignant and non-malignant cells. **C** Differentiation trajectory ofepithelial cells, with each color coded for

pseudotime (right) and clusters (left). **D** GO enrichment analysis of up-regulated genes in malignant lymphatic metastasis cells. while color represents expression level. **E** Heatmap of the area under the curve (AUC) scores of TF motifs estimated per cell by SCENIC. Shown are differentially activated motifs in malignant and non-malignant cells, respectively.

Figure 3 . T cell population heterogeneity. A UMAP plot of T cells color-coded by their associated clusters. **B** Dotplot of cell markers; sizes of dots represent abundance, while color represents expression level. **C** Potential new cell markers, additional new cell marker genes were shown in Supplementary Fig. 2. **D** Changes in frequency of multiple cell types and clusters in T cell population. For each cell type and cluster, the average fraction of cells from each patient group is shown, Asterisks on the left of the vertical line denote statistically significant differences between primary papillary thyroid carcinoma and normal thyroid tissue, meanwhile, asterisks on the right is used to show statistically significant differences between lymph metastatic andprimary papillary thyroid carcinoma. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, two-tailed t-tests. **E** Violin of cell abundance predicted from TCGA THCA cohort by CIBERSORTx.

Figure 4 . CD4+Tregs and transitional EGR1+CD4+T cells in tumor and lymph metastatic PTC and potential therapeutic target. A Volcano plot of differentially expressed genes (DEGs) between lymph metastatic and primary tumor in transitional EGR1+CD4+T cells. Upregulated genes and downregulated genes were colored in red and blue, respectively. FC≥2. Additional Volcano plot were shown in Supplementary Fig. 3. **B** Differences in 50 hallmark pathway activities scored with GSVA software. Shown are t values calculated by

a linear model. **C** Heatmap of the area under the curve (AUC) scores of TF motifs estimated per cell by SCENIC. Shown are differentially activated motifs in CD4+Tregs and transitional EGR1+CD4+T cells, respectively. **D** Targeted drug prediction corresponding to differential genes in TCMID database, Additional Drug-Target Genes were shown in Supplementary Fig 5. **E** Heatmap of positive and negative immune checkpoint expression on T cells in lymph metastatic PTC. Additional immune checkpoint expression were shown in Supplementary Fig 5

Figure 5 . Myeloid cells population heterogeneity. A UMAP plot of Myeloid cells color-coded by their associated clusters. **B** Dotplot of cell markers; sizes of dots represent abundance, while color represents expression level. **C** Changes in frequency of multiple cell types and clusters in Myeloid cell population. For each cell type and cluster, the average fraction of cells from each patient group is shown, Asterisks on the left of the vertical line denote statistically significant differences between primary papillary thyroid carcinoma and normal thyroid tissue, meanwhile, asterisks on the right is used to show statistically significant differences between lymph metastatic andprimary papillary thyroid carcinoma. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, two-tailed t-tests. **D** Violin of cell abundance predicted from TCGA THCA cohort by CIBERSORTx. **E** Kaplan-Meier curves of DFS for patients with THCA based on the percentage of dendritic cells in the TCGA datebase.

Figure 6 . M2 macrophages are strongly enriched in the tumor. A,B Heatmap shows difference in pathway activities scored by GSVA per cell between different sample groups. Shown are t values calculated by a linear model. **C** Heatmap of the area under the curve

(AUC) scores of TF motifs estimated per cell by SCENIC. Shown are differentially activated motifs in myeloid cells sub types, respectively. **D** Dotplot of M1 and M2 macrophages markers; sizes of dots represent abundance, while color represents expression level.

Reference

1. Sung H, *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. (2021).

2. Li M, Maso LD, Vaccarella S. Global trends in thyroid cancer incidence and the impact of overdiagnosis. *The Lancet Diabetes & Endocrinology* **8**, 468-470 (2020).

3. Fagin JA, Wells SA, Jr. Biologic and Clinical Perspectives on Thyroid Cancer. *N Engl J Med* **375**, 1054-1067 (2016).

4. Yu J, *et al.* Lymph node metastasis prediction of papillary thyroid carcinoma based on transfer learning radiomics. *Nat Commun* **11**, 4807 (2020).

5. Shen X, *et al.* Patient Age-Associated Mortality Risk Is Differentiated by BRAF V600E Status in Papillary Thyroid Cancer. **36**, 438-445 (2018).

6. Pozdeyev N, *et al.* Genetic Analysis of 779 Advanced Differentiated and Anaplastic Thyroid Cancers. **24**, 3059-3068 (2018).

7. Liu R, Bishop J, Zhu G, Zhang T, Ladenson P, Xing MJJo. Mortality Risk Stratification

by Combining BRAF V600E and TERT Promoter Mutations in Papillary Thyroid Cancer: Genetic Duet of BRAF and TERT Promoter Mutations in Thyroid Cancer Mortality. **3**, 202-208 (2017).

8. Li Y*, et al.* CYP2S1 is a synthetic lethal target in BRAF-driven thyroid cancers. **5**, 191 (2020).

9. Robertson A*, et al.* Identification of Differential Tumor Subtypes of T1 Bladder Cancer. **78**, 533-537 (2020).

10. Mer A*, et al.* Biological and therapeutic implications of a unique subtype of NPM1 mutated AML. **12**, 1054 (2021).

11. Job S*, et al.* Identification of Four Immune Subtypes Characterized by Distinct Composition and Functions of Tumor Microenvironment in Intrahepatic Cholangiocarcinoma. **72**, 965-981 (2020).

12. Wright G*, et al.* A Probabilistic Classification Tool for Genetic Subtypes of Diffuse Large B Cell Lymphoma with Therapeutic Implications. **37**, 551-568.e514 (2020).

13. Guruprasad P, Lee Y, Kim K, Ruella MJTJoem. The current landscape of single-cell transcriptomics for cancer immunotherapy. **218**, (2021).

14. Stoeckius M, *et al.* Simultaneous epitope and transcriptome measurement in single cells. **14**, 865-868 (2017).

15. Butler A, Hoffman P, Smibert P, Papalexi E, Satija RJNb. Integrating single-cell transcriptomic data across different conditions, technologies, and species. **36**, 411-420 (2018).

16. Jerby-Arnon L, *et al.* A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade. **175**, 984-997.e924 (2018).

17. Peng J, *et al.* Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. **29**, 725-738 (2019).

18. Zhang X, *et al.* CellMarker: a manually curated resource of cell markers in human and mouse. **47**, D721-D728 (2019).

19. Zilionis R, *et al.* Single-Cell Transcriptomics of Human and Mouse Lung Cancers Reveals Conserved Myeloid Populations across Individuals and Species. **50**, 1317-1334.e1310 (2019).

20. Luoma A, *et al.* Molecular Pathways of Colon Inflammation Induced by Cancer

Immunotherapy. **182**, 655-671.e622 (2020).

21. Habib N*, et al.* Disease-associated astrocytes in Alzheimer's disease and aging. *Nat Neurosci* **23**, 701-706 (2020).

22. Steen CB, Liu CL, Alizadeh AA, Newman AM. Profiling Cell Type Abundance and Expression in Bulk Tissues with CIBERSORTx. *Methods Mol Biol* **2117**, 135-157 (2020).

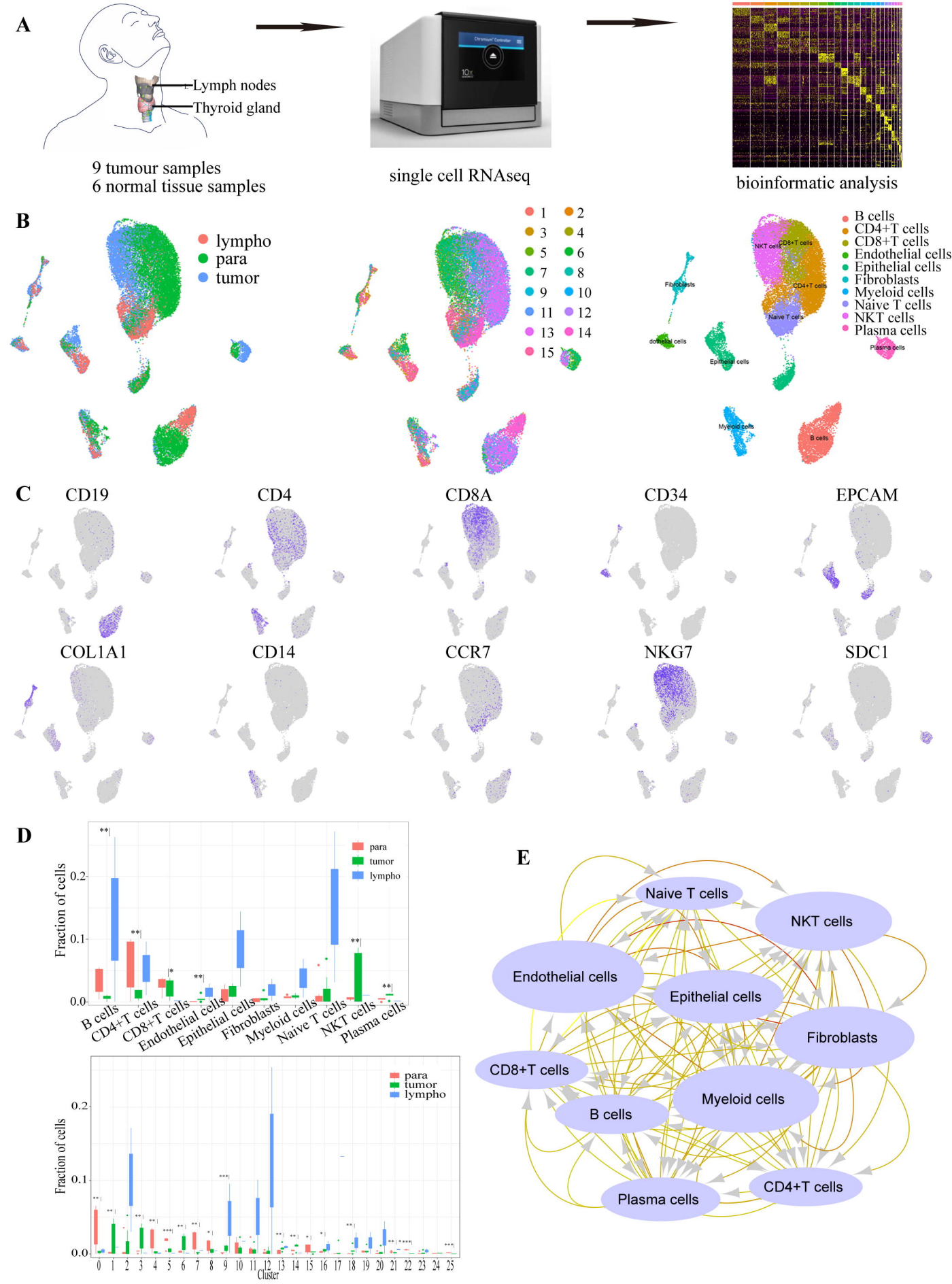
23. Hänzelmann S, Castelo R, Guinney JJBb. GSVA: gene set variation analysis for microarray and RNA-seq data. **14**, 7 (2013).

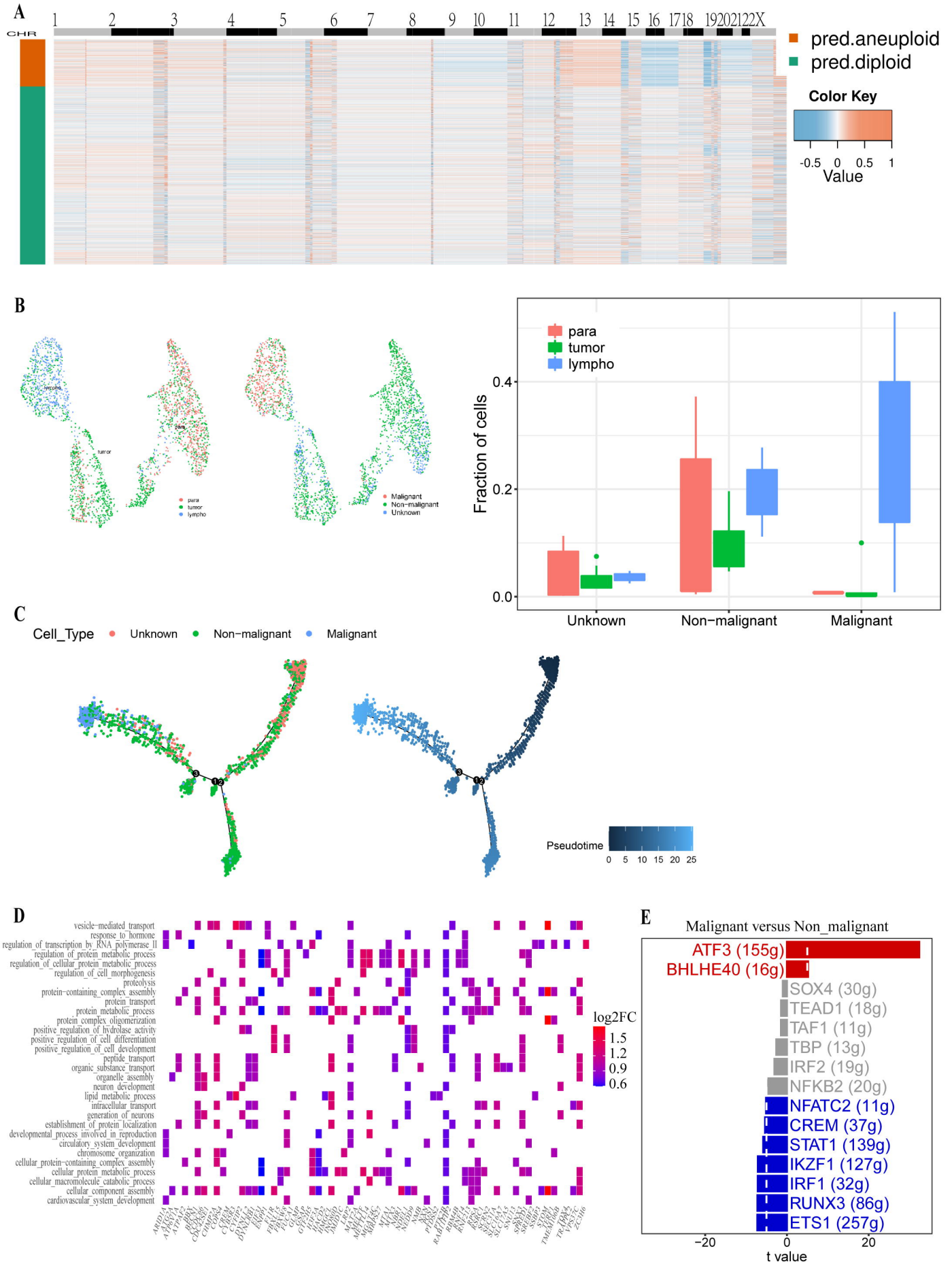
24. Aibar S*, et al.* SCENIC: single-cell regulatory network inference and clustering. **14**, 1083-1086 (2017).

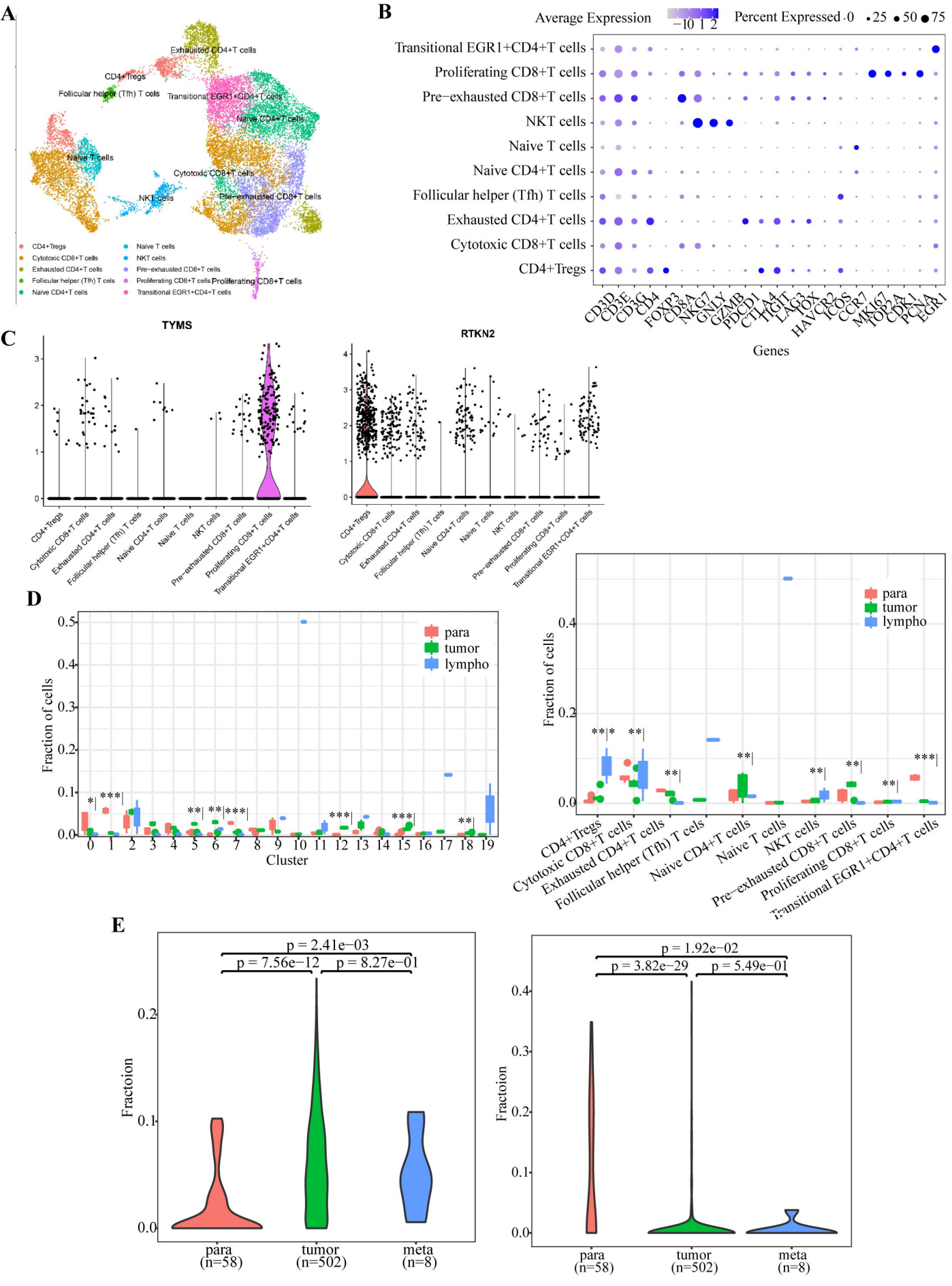
25. Efremova M, Vento-Tormo M, Teichmann S, Vento-Tormo RJNp. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. **15**, 1484-1506 (2020).

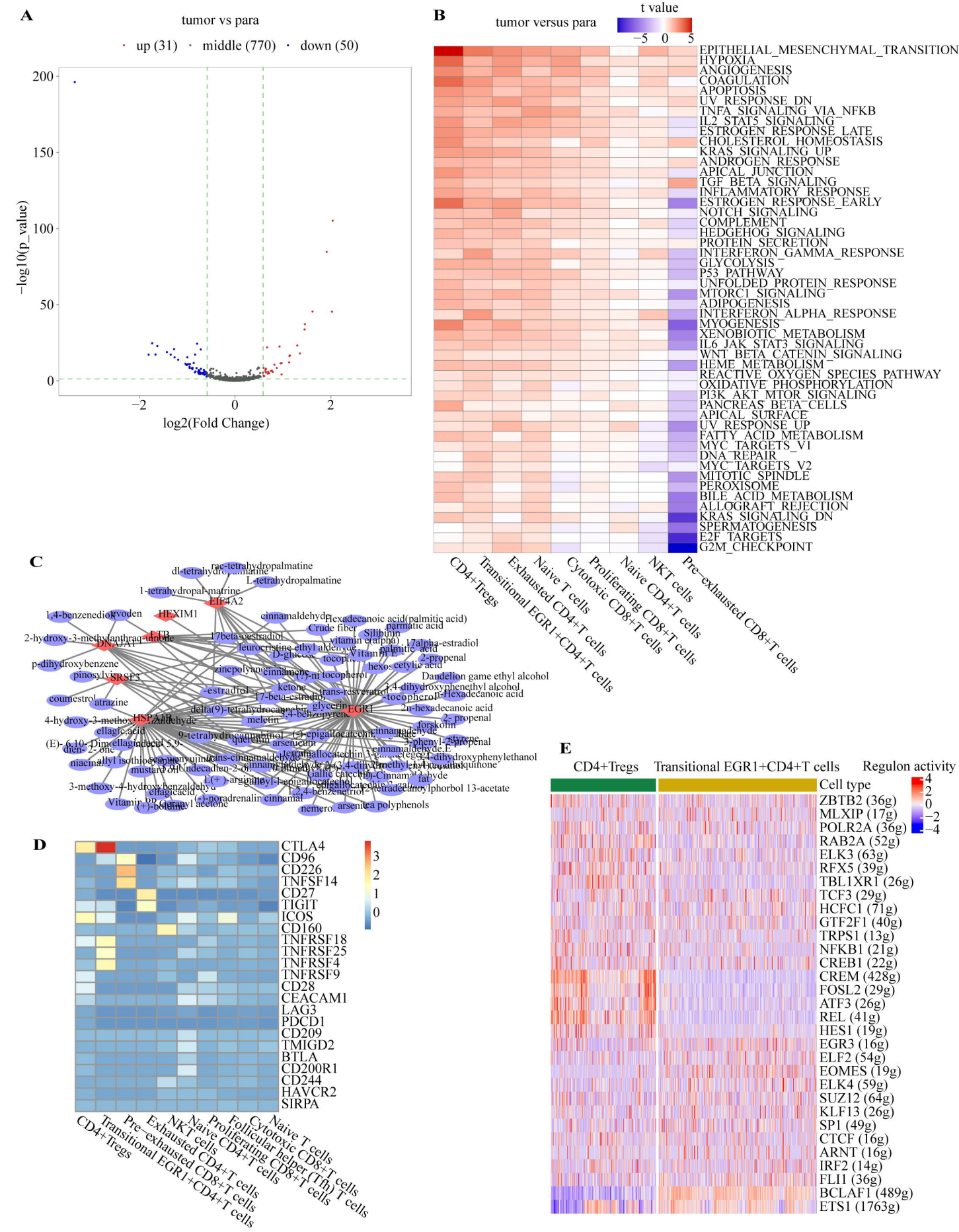
26. Trapnell C*, et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386 (2014).

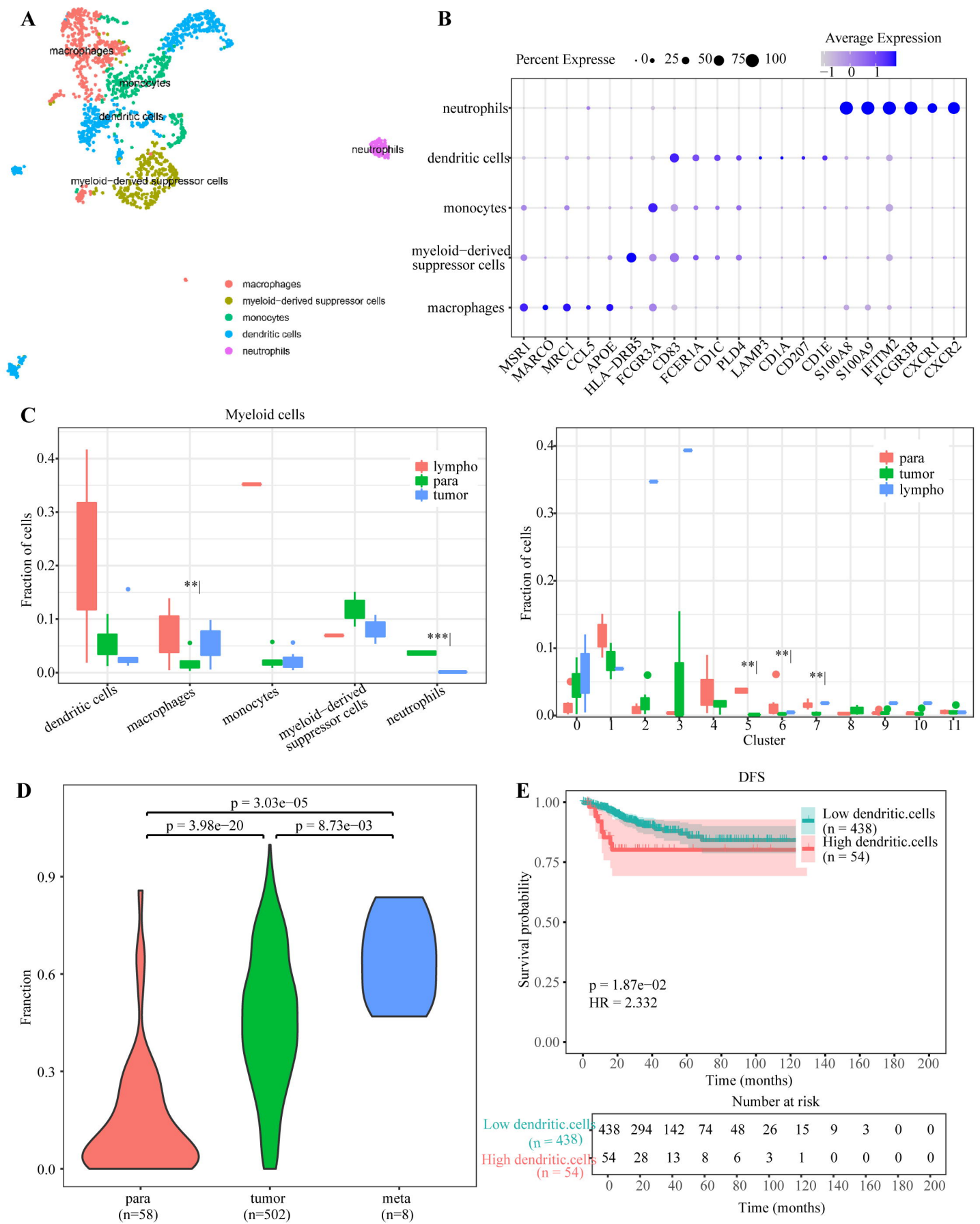
27. Gao R, *et al.* Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat Biotechnol*, (2021).

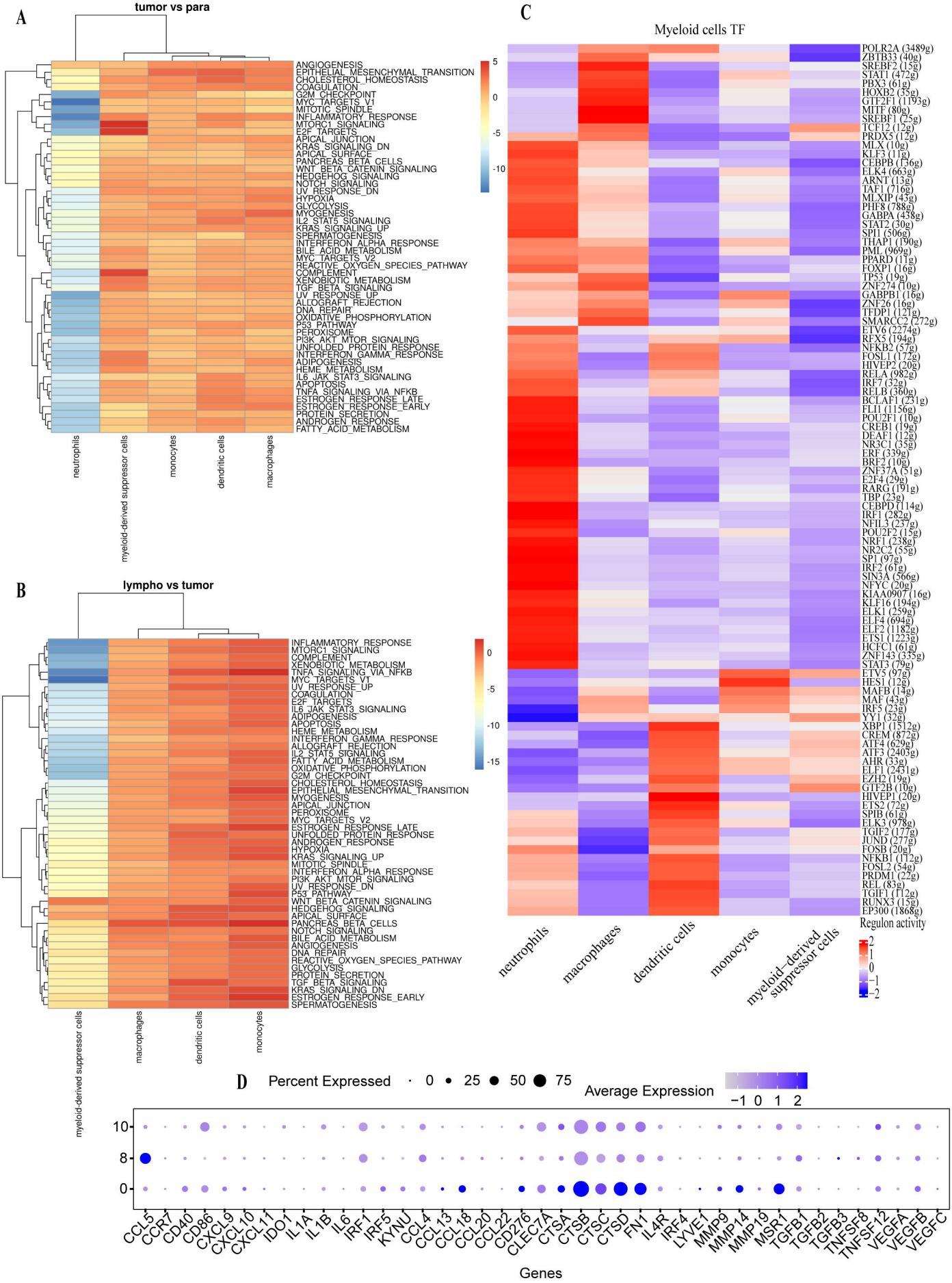

A
Lymph nodes
Thyroid gland
9 tumour samples
6 normal tissue samples
single cell RNAseq
bioinformatic analysis
B
lympho
para
tumor
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
B cells
CD4+T cells
CD8+T cells
Endothelial cells
Epithelial cells
Fibroblasts
Myeloid cells
Naive T cells
NKT cells
Plasma cells
C
CD19
CD4
CD8A
CD34
EPCAM
COL1A1
CD14
CCR7
NKG7
SDC1
D
Fraction of cells
para
tumor
lympho
B cells
CD4+T cells
CD8+T cells
Endothelial cells
Epithelial cells
Fibroblasts
Myeloid cells
Naive T cells
NKT cells
Plasma cells
Cluster
E
Naive T cells
NKT cells
Endothelial cells
Epithelial cells
Fibroblasts
CD8+T cells
Myeloid cells
B cells
Plasma cells
CD4+T cells


A
CHR
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
pred.aneuploid
pred.diploid
Color Key
-0.5 0 0.5 1
Value
B
lympho
tumor
para
para
tumor
lympho
Malignant
Non−malignant
Unknown
para
tumor
lympho
Fraction of cells
0.0
0.2
0.4
Unknown
Non−malignant
Malignant
C
Cell_Type
Unknown
Non−malignant
Malignant
Pseudotime
0 5 10 15 20 25
D
vesicle−mediated transport
response to hormone
regulation of transcription by RNA polymerase II
regulation of protein metabolic process
regulation of cellular protein metabolic process
regulation of cell morphogenesis
proteolysis
protein−containing complex assembly
protein transport
protein metabolic process
protein complex oligomerization
positive regulation of hydrolase activity
positive regulation of cell differentiation
positive regulation of cell development
peptide transport
organic substance transport
organelle assembly
neuron development
lipid metabolic process
intracellular transport
generation of neurons
establishment of protein localization
developmental process involved in reproduction
circulatory system development
chromosome organization
cellular protein−containing complex assembly
cellular protein metabolic process
cellular macromolecule catabolic process
cellular component assembly
cardiovascular system development
log2FC
1.5
1.2
0.9
0.6
E
Malignant versus Non_malignant
ATF3 (155g)
BHLHE40 (16g)
SOX4 (30g)
TEAD1 (18g)
TAF1 (11g)
TBP (13g)
IRF2 (19g)
NFKB2 (20g)
NFATC2 (11g)
CREM (37g)
STAT1 (139g)
IKZF1 (127g)
IRF1 (32g)
RUNX3 (86g)
ETS1 (257g)
−20
0
20
t value


A
Exhausted CD4+T cells
CD4+Tregs
Follicular helper (Tfh) T cells
Transitional EGR1+CD4+T cells
Naive CD4+T cells
Naive T cells
Cytotoxic CD8+T cells
NKT cells
Pre−exhausted CD8+T cells
Proliferating CD8+T cells
B
Average Expression
Percent Expressed
Transitional EGR1+CD4+T cells
Proliferating CD8+T cells
Pre−exhausted CD8+T cells
NKT cells
Naive T cells
Naive CD4+T cells
Follicular helper (Tfh) T cells
Exhausted CD4+T cells
Cytotoxic CD8+T cells
CD4+Tregs
CD3D
CD3E
CD3G
CD4
FOXP3
CD8A
NKG7
GNLY
GZMB
PDCD1
CTLA4
TIGIT
LAG3
TOX
HAVCR2
ICOS
CCR7
MKI67
TOP2A
CDK1
PCNA
EGR1
Genes
C
TYMS
RTKN2
D
para
tumor
lympho
Fraction of cells
Cluster
E
p = 2.41e−03
p = 7.56e−12
p = 8.27e−01
p = 1.92e−02
p = 3.82e−29
p = 5.49e−01
Fractoion
para (n=58)
tumor (n=502)
meta (n=8)


A
tumor vs para
up (31)
middle (770)
down (50)
−log10(p_value)
log2(Fold Change)
B
tumor versus para
t value
−5
0
5
EPITHELIAL_MESENCHYMAL_TRANSITION
HYPOXIA
ANGIOGENESIS
COAGULATION
APOPTOSIS
UV_RESPONSE_DN
TNFA_SIGNALING_VIA_NFKB
IL2_STAT5_SIGNALING
ESTROGEN_RESPONSE_LATE
CHOLESTEROL_HOMEOSTASIS
KRAS_SIGNALING_UP
ANDROGEN_RESPONSE
APICAL_JUNCTION
TGF_BETA_SIGNALING
INFLAMMATORY_RESPONSE
ESTROGEN_RESPONSE_EARLY
NOTCH_SIGNALING
COMPLEMENT
HEDGEHOG_SIGNALING
PROTEIN_SECRETION
INTERFERON_GAMMA_RESPONSE
GLYCOLYSIS
P53_PATHWAY
UNFOLDED_PROTEIN_RESPONSE
MTORC1_SIGNALING
ADIPOGENESIS
INTERFERON_ALPHA_RESPONSE
MYOGENESIS
XENOBIOTIC_METABOLISM
IL6_JAK_STAT3_SIGNALING
WNT_BETA_CATENIN_SIGNALING
HEME_METABOLISM
REACTIVE_OXYGEN_SPECIES_PATHWAY
OXIDATIVE_PHOSPHORYLATION
PI3K_AKT_MTOR_SIGNALING
PANCREAS_BETA_CELLS
APICAL_SURFACE
UV_RESPONSE_UP
FATTY_ACID_METABOLISM
MYC_TARGETS_V1
DNA_REPAIR
MYC_TARGETS_V2
MITOTIC_SPINDLE
PEROXISOME
BILE_ACID_METABOLISM
ALLOGRAFT_REJECTION
KRAS_SIGNALING_DN
SPERMATOGENESIS
E2F_TARGETS
G2M_CHECKPOINT
CD4+Tregs
Transitional EGR1+CD4+T cells
Exhausted CD4+T cells
Naive T cells
Cytotoxic CD8+T cells
Proliferating CD8+T cells
Naive CD4+T cells
NKT cells
Pre−exhausted CD8+T cells
C
D
CTLA4
CD96
CD226
TNFSF14
CD27
TIGIT
ICOS
CD160
TNFRSF18
TNFRSF25
TNFRSF4
TNFRSF9
CD28
CEACAM1
LAG3
PDCD1
CD209
TMIGD2
BTLA
CD200R1
CD244
HAVCR2
SIRPA
CD4+Tregs
Transitional EGR1+CD4+T cells
Pre−exhausted CD8+T cells
Exhausted CD4+T cells
NKT cells
Naive CD4+T cells
Proliferating CD8+T cells
Follicular helper (Tfh) T cells
Cytotoxic CD8+T cells
Naive T cells
E
CD4+Tregs
Transitional EGR1+CD4+T cells
Regulon activity
Cell type
ZBTB2 (36g)
MLXIP (17g)
POLR2A (36g)
RAB2A (52g)
ELK3 (63g)
RFX5 (39g)
TBL1XR1 (26g)
TCF3 (29g)
HCFC1 (71g)
GTF2F1 (40g)
TRPS1 (13g)
NFKB1 (21g)
CREB1 (22g)
CREM (428g)
FOSL2 (29g)
ATF3 (26g)
REL (41g)
HES1 (19g)
EGR3 (16g)
ELF2 (54g)
EOMES (19g)
ELK4 (59g)
SUZ12 (64g)
KLF13 (26g)
SP1 (49g)
CTCF (16g)
ARNT (16g)
IRF2 (14g)
FLI1 (36g)
BCLAF1 (489g)
ETS1 (1763g)


A
macrophages
monocytes
dendritic cells
neutrophils
myeloid−derived suppressor cells
macrophages
myeloid−derived suppressor cells
monocytes
dendritic cells
neutrophils
B
Percent Expresse
0
25
50
75
100
Average Expression
−1
0
1
neutrophils
dendritic cells
monocytes
myeloid−derived suppressor cells
macrophages
MSR1
MARCO
MRC1
CCL5
APOE
HLA−DRB5
FCGR3A
CD83
FCER1A
CD1C
PLD4
LAMP3
CD1A
CD207
CD1E
S100A8
S100A9
IFITM2
FCGR3B
CXCR1
CXCR2
C
Myeloid cells
Fraction of cells
lympho
para
tumor
**
***
dendritic cells
macrophages
monocytes
myeloid−derived suppressor cells
neutrophils
Cluster
D
p = 3.03e−05
p = 3.98e−20
p = 8.73e−03
Franction
para (n=58)
tumor (n=502)
meta (n=8)
E
DFS
Low dendritic.cells (n = 438)
High dendritic.cells (n = 54)
Survival probability
p = 1.87e−02
HR = 2.332
Time (months)
Number at risk
Low dendritic.cells (n = 438)
High dendritic.cells (n = 54)
438 294 142 74 48 26 15 9 3 0 0
54 28 13 8 6 3 1 0 0 0 0
0 20 40 60 80 100 120 140 160 180 200
Time (months)


A
tumor vs para
ANGIOGENESIS
EPITHELIAL_MESENCHYMAL_TRANSITION
CHOLESTEROL_HOMEOSTASIS
COAGULATION
G2M_CHECKPOINT
MYC_TARGETS_V1
MITOTIC_SPINDLE
INFLAMMATORY_RESPONSE
MTORC1_SIGNALING
E2F_TARGETS
APICAL_JUNCTION
KRAS_SIGNALING_DN
APICAL_SURFACE
PANCREAS_BETA_CELLS
WNT_BETA_CATENIN_SIGNALING
HEDGEHOG_SIGNALING
NOTCH_SIGNALING
UV_RESPONSE_DN
HYPOXIA
GLYCOLYSIS
MYOGENESIS
IL2_STAT5_SIGNALING
KRAS_SIGNALING_UP
SPERMATOGENESIS
INTERFERON_ALPHA_RESPONSE
BILE_ACID_METABOLISM
MYC_TARGETS_V2
REACTIVE_OXYGEN_SPECIES_PATHWAY
COMPLEMENT
XENOBIOTIC_METABOLISM
TGF_BETA_SIGNALING
UV_RESPONSE_UP
ALLOGRAFT_REJECTION
DNA_REPAIR
OXIDATIVE_PHOSPHORYLATION
P53_PATHWAY
PEROXISOME
PI3K_AKT_MTOR_SIGNALING
UNFOLDED_PROTEIN_RESPONSE
INTERFERON_GAMMA_RESPONSE
ADIPOGENESIS
HEME_METABOLISM
IL6_JAK_STAT3_SIGNALING
APOPTOSIS
TNFA_SIGNALING_VIA_NFKB
ESTROGEN_RESPONSE_LATE
ESTROGEN_RESPONSE_EARLY
PROTEIN_SECRETION
ANDROGEN_RESPONSE
FATTY_ACID_METABOLISM
neutrophils
myeloid-derived suppressor cells
monocytes
dendritic cells
macrophages
5
0
-5
-10
B
lympho vs tumor
INFLAMMATORY_RESPONSE
MTORC1_SIGNALING
COMPLEMENT
XENOBIOTIC_METABOLISM
TNFA_SIGNALING_VIA_NFKB
MYC_TARGETS_V1
UV_RESPONSE_UP
COAGULATION
E2F_TARGETS
IL6_JAK_STAT3_SIGNALING
ADIPOGENESIS
APOPTOSIS
HEME_METABOLISM
INTERFERON_GAMMA_RESPONSE
ALLOGRAFT_REJECTION
IL2_STAT5_SIGNALING
FATTY_ACID_METABOLISM
OXIDATIVE_PHOSPHORYLATION
G2M_CHECKPOINT
CHOLESTEROL_HOMEOSTASIS
EPITHELIAL_MESENCHYMAL_TRANSITION
MYOGENESIS
APICAL_JUNCTION
PEROXISOME
MYC_TARGETS_V2
ESTROGEN_RESPONSE_LATE
UNFOLDED_PROTEIN_RESPONSE
ANDROGEN_RESPONSE
HYPOXIA
KRAS_SIGNALING_UP
MITOTIC_SPINDLE
INTERFERON_ALPHA_RESPONSE
PI3K_AKT_MTOR_SIGNALING
UV_RESPONSE_DN
P53_PATHWAY
WNT_BETA_CATENIN_SIGNALING
HEDGEHOG_SIGNALING
APICAL_SURFACE
PANCREAS_BETA_CELLS
NOTCH_SIGNALING
BILE_ACID_METABOLISM
ANGIOGENESIS
DNA_REPAIR
REACTIVE_OXYGEN_SPECIES_PATHWAY
GLYCOLYSIS
PROTEIN_SECRETION
TGF_BETA_SIGNALING
KRAS_SIGNALING_DN
ESTROGEN_RESPONSE_EARLY
SPERMATOGENESIS
myeloid-derived suppressor cells
macrophages
dendritic cells
monocytes
0
-5
-10
-15
C
Myeloid cells TF
POLR2A (3489g)
ZBTB33 (40g)
SREBF2 (15g)
STAT1 (472g)
PBX3 (61g)
HOXB2 (35g)
GTF2F1 (1193g)
MITF (80g)
SREBF1 (25g)
TCF12 (12g)
PRDX5 (12g)
MLX (10g)
KLF3 (11g)
CEBPB (136g)
ELK4 (663g)
ARNT (13g)
TAF1 (716g)
MLXIP (43g)
PHF8 (788g)
GABPA (438g)
STAT2 (30g)
SPI1 (506g)
THAP1 (190g)
PML (969g)
PPARD (11g)
FOXP1 (16g)
TP53 (19g)
ZNF274 (10g)
GABPB1 (16g)
ZNF26 (16g)
TFDP1 (121g)
SMARCC2 (272g)
ETV6 (2274g)
RFX5 (194g)
NFKB2 (57g)
FOSL1 (172g)
HIVEP2 (20g)
RELA (982g)
IRF7 (32g)
RELB (360g)
BCLAF1 (231g)
FLI1 (1156g)
POU2F1 (10g)
CREB1 (19g)
DEAF1 (12g)
NR3C1 (35g)
ERF (339g)
BRF2 (10g)
ZNF37A (51g)
E2F4 (29g)
RARG (191g)
TBP (23g)
CEBPD (114g)
IRF1 (282g)
NFIL3 (237g)
POU2F2 (15g)
NRF1 (238g)
NR2C2 (55g)
SP1 (97g)
IRF2 (61g)
SIN3A (566g)
NFYC (20g)
KIAA0907 (16g)
KLF16 (194g)
ELK1 (259g)
ELF4 (694g)
ELF2 (1182g)
ETS1 (1223g)
HCFC1 (61g)
ZNF143 (335g)
STAT3 (79g)
ETV5 (97g)
HES1 (12g)
MAFB (14g)
MAF (43g)
IRF5 (23g)
YY1 (32g)
XBP1 (1512g)
CREM (872g)
ATF4 (629g)
ATF3 (2403g)
AHR (33g)
ELF1 (2431g)
EZH2 (19g)
GTF2B (10g)
HIVEP1 (20g)
ETS2 (72g)
SPIB (61g)
ELK3 (978g)
TGIF2 (177g)
JUND (277g)
FOSB (20g)
NFKB1 (112g)
FOSL2 (54g)
PRDM1 (22g)
REL (83g)
TGIF1 (112g)
RUNX3 (15g)
EP300 (1868g)
neutrophils
macrophages
dendritic cells
monocytes
myeloid-derived suppressor cells
Regulon activity
2
1
0
−1
−2
D
Percent Expressed
0
25
50
75
Average Expression
−1
0
1
2
10
8
0
CCL5
CCR7
CD40
CD86
CXCL9
CXCL10
CXCL11
IDO1
IL1A
IL1B
IL6
IRF1
IRF5
KYNU
CCL4
CCL13
CCL18
CCL20
CCL22
CD276
CLEC7A
CTSA
CTSB
CTSC
CTSD
FN1
IL4R
IRF4
LYVE1
MMP9
MMP14
MMP19
MSR1
TGFB1
TGFB2
TGFB3
TNFSF8
TNFSF12
VEGFA
VEGFB
VEGFC
Genes