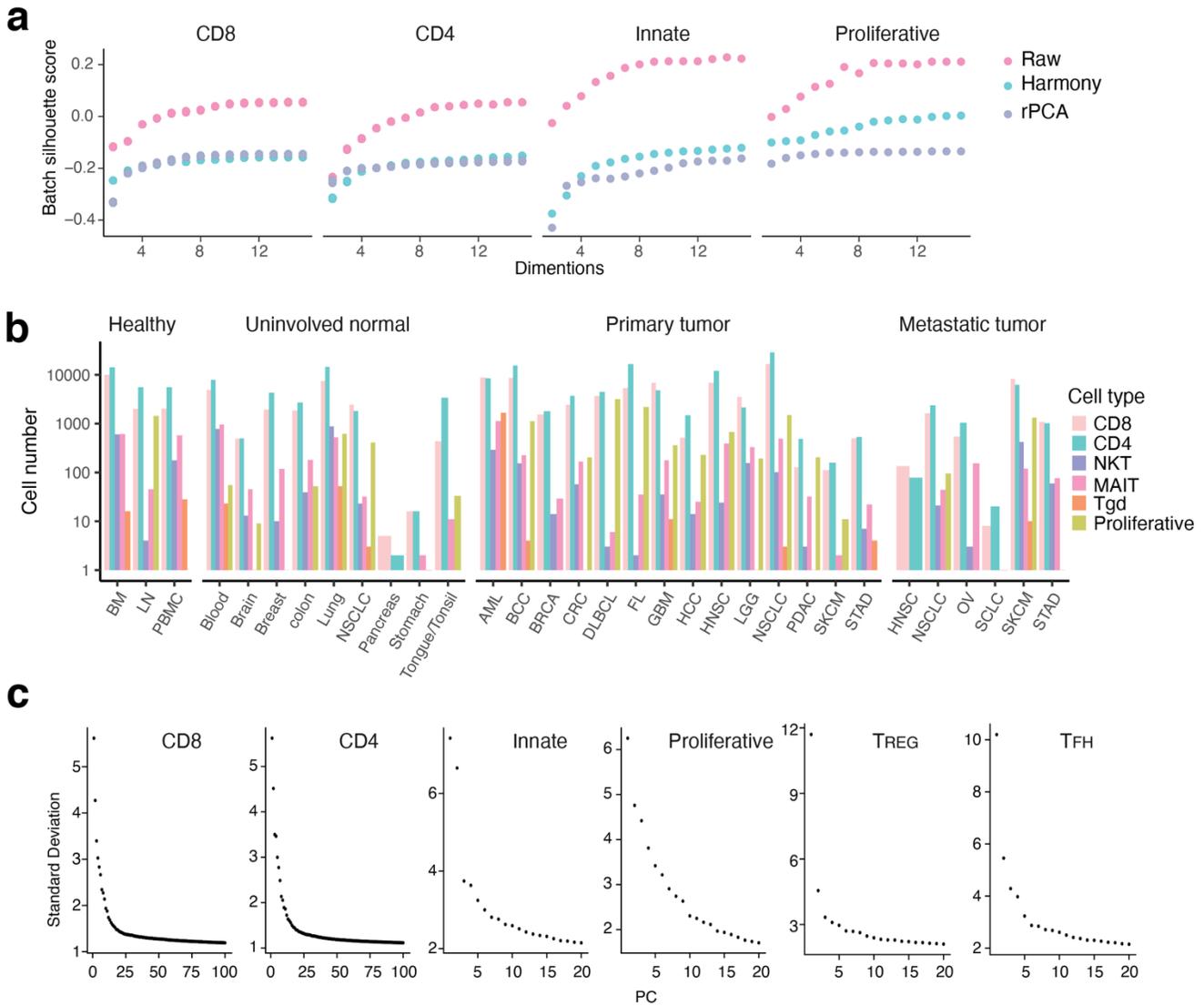


SUPPLEMENTARY FIGURES

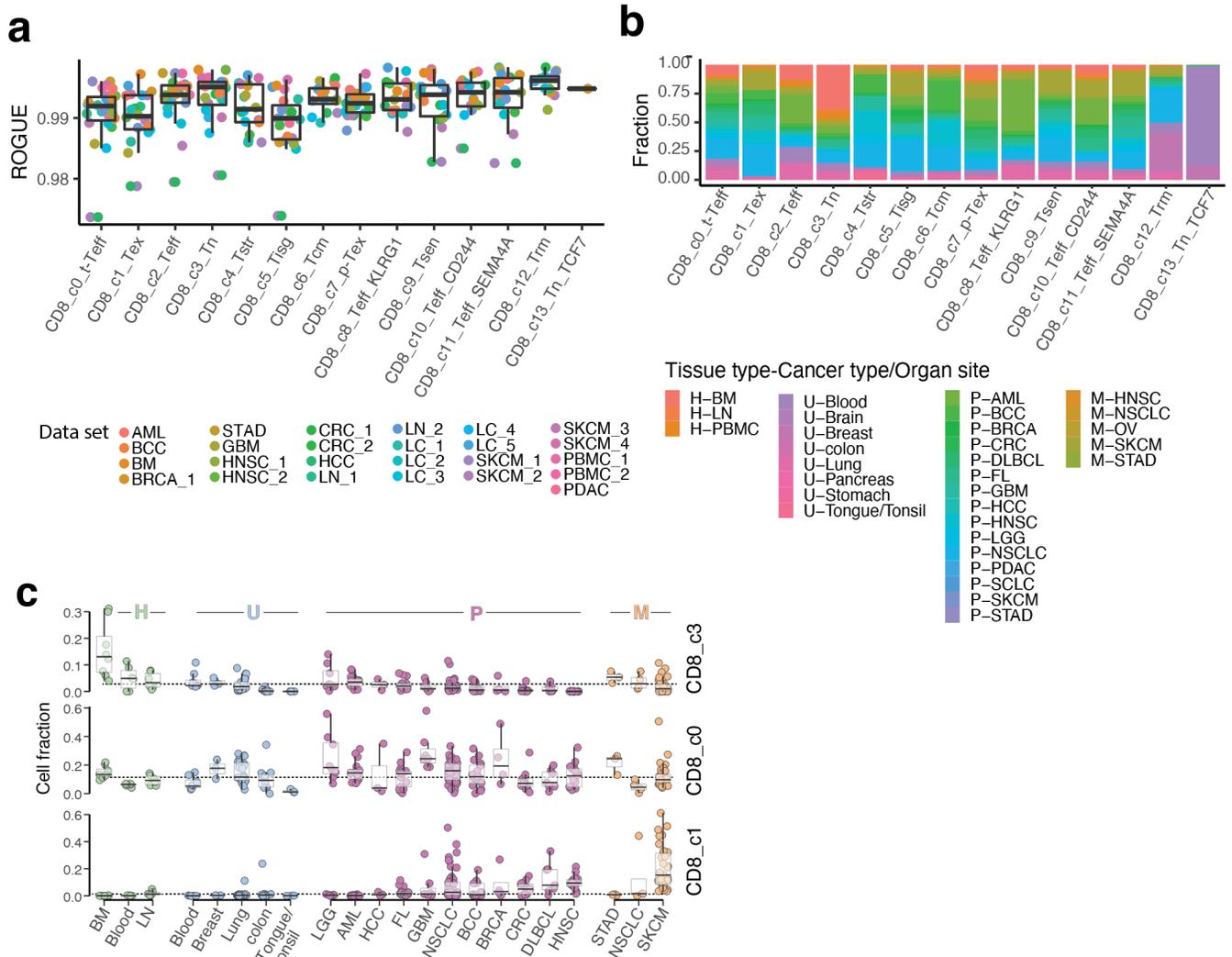
Supplementary Figure 1



Supplementary Figure 1: Batch effect correction and cell distribution across different tissues groups. (a) Dot plots showing the silhouette scores for major cell types without (raw) and with batch effect correction. The silhouette scores range from -1 to $+1$, where a high value indicates that the cell is well matched to its own batch and poorly matched to neighboring batches, while a low or negative value indicates that the cell is well matched to its neighboring batches instead of its own batch. The lower the value the lower the potential batch effect and the better the performance of a batch correction approach. For raw data without any batch correction, x-axis means the number of PCs selected to calculate silhouette scores. For the batch corrected data by harmony, x-axis means the number of harmony dimensions used to calculate silhouette scores and for the batch corrected data by rPCA, x-axis means the number of batch corrected PCs used to calculate silhouette scores. **(b)** Bar-plot showing the cell number distribution of each major cell type across various cohorts of 4 main tissue groups. BM, bone marrow; LN, lymph node; PBMC, peripheral blood mononuclear cell. For uninvolved normal tissues, their corresponding organs of sample collection are labeled. For primary and metastatic tumor tissues, their cancer types

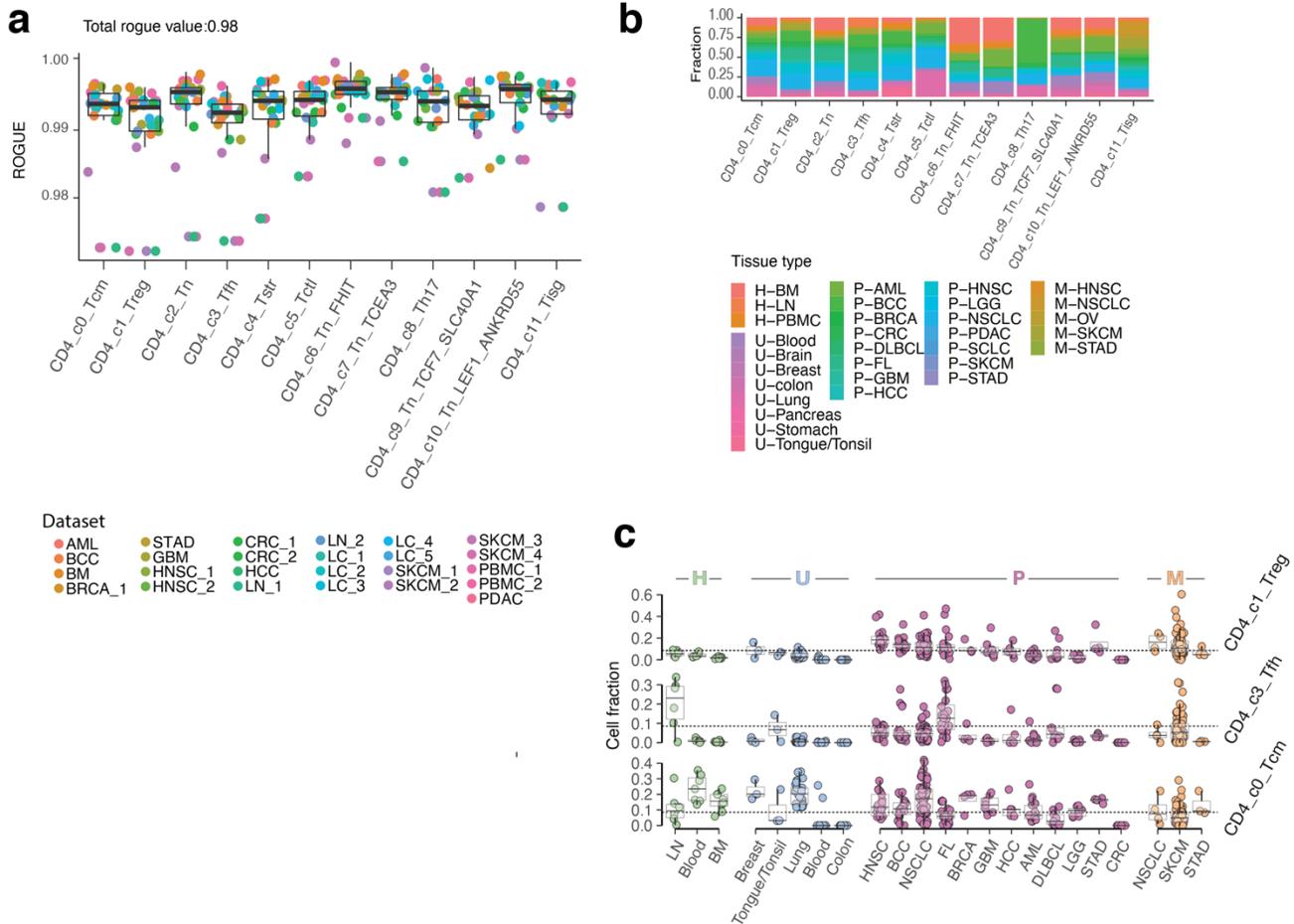
(abbreviations in TCGA format) are labelled. **(c)** Elbow plots showing the rankings of principle components (PCs) based on the percentage of variance explained by each one.

Supplementary Figure 2



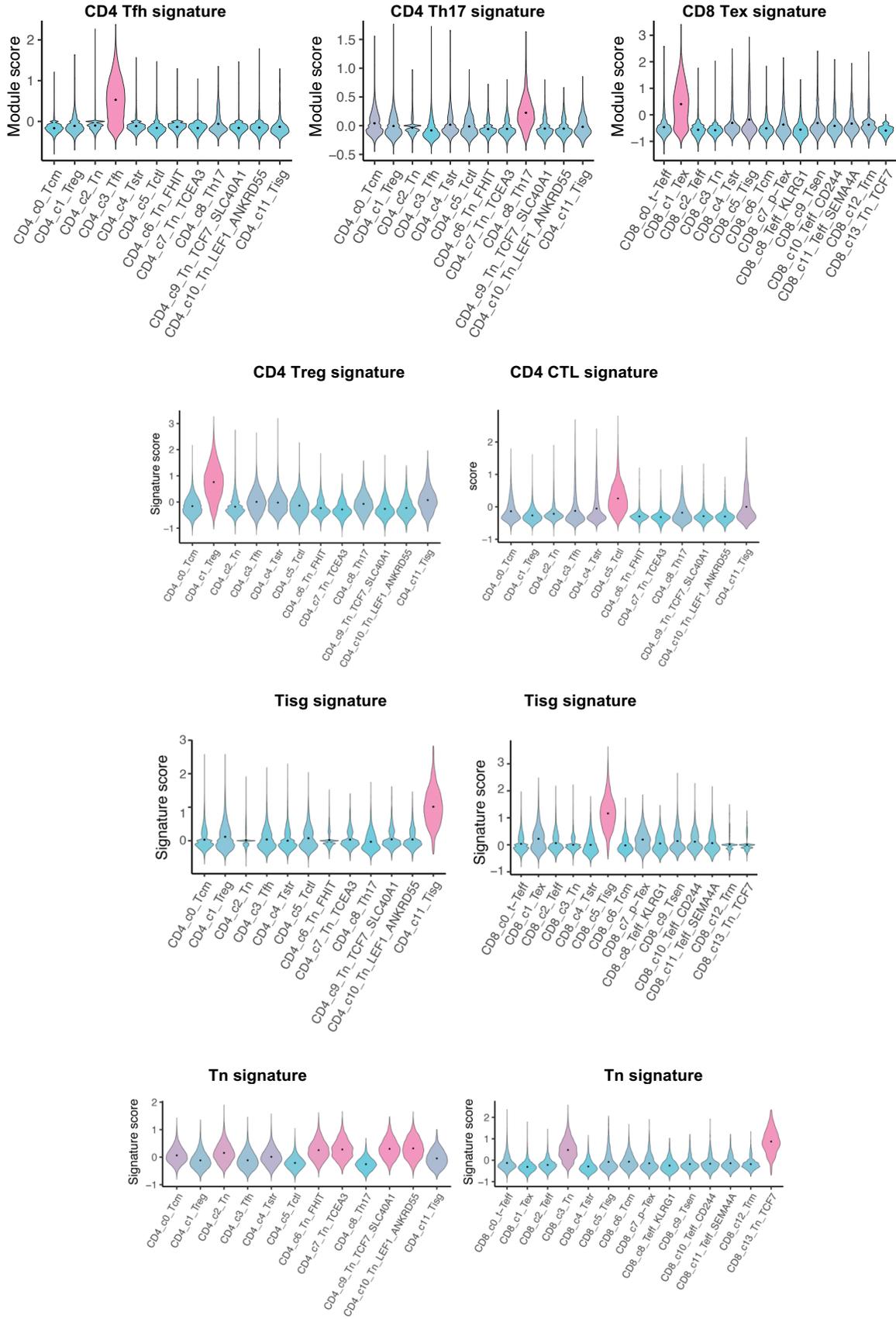
Supplementary Figure 2: CD8+ T cell clusters. (a) Box-plot showing the distribution of ROGUE (Ratio of Global Unshifted Entropy) scores across 14 CD8 T cell clusters. The R package ROGUE (v1.0) was applied to calculate the ROGUE scores. Each dot represents a single cell dataset (Supplementary Table 1). The sample number n of each CD8 T cell clusters from c0 to c13 are 27, 27, 27, 26, 25, 25, 26, 26, 24, 24, 24, 25, 24, 4. Boxes, median \pm interquartile range; whiskers, $1.5 \times$ interquartile range. (b) Bar graph showing the tissue compositions (i.e., the fractions of cells belonging to each of the 4 different tissue types) of 14 defined CD8 T cell clusters. (c) Box plots showing cellular fraction distribution of three CD8 T cell subsets across various cohorts of 4 tissue types. BM, bone marrow; LN, lymph node. For uninvolved normal tissues, their corresponding organs of sample collection are labeled. For primary and metastatic tumour tissues, their cancer types (abbreviations in TCGA format) are labelled. Sample number n of each x axis from left to right are 8, 6, 6, 8, 3, 29, 6, 3, 8, 16, 3, 17, 7, 46, 16, 4, 13, 7, 16, 3, 4, 31. Boxes, median \pm the interquartile range; whiskers, $1.5 \times$ interquartile range.

Supplementary Figure 3



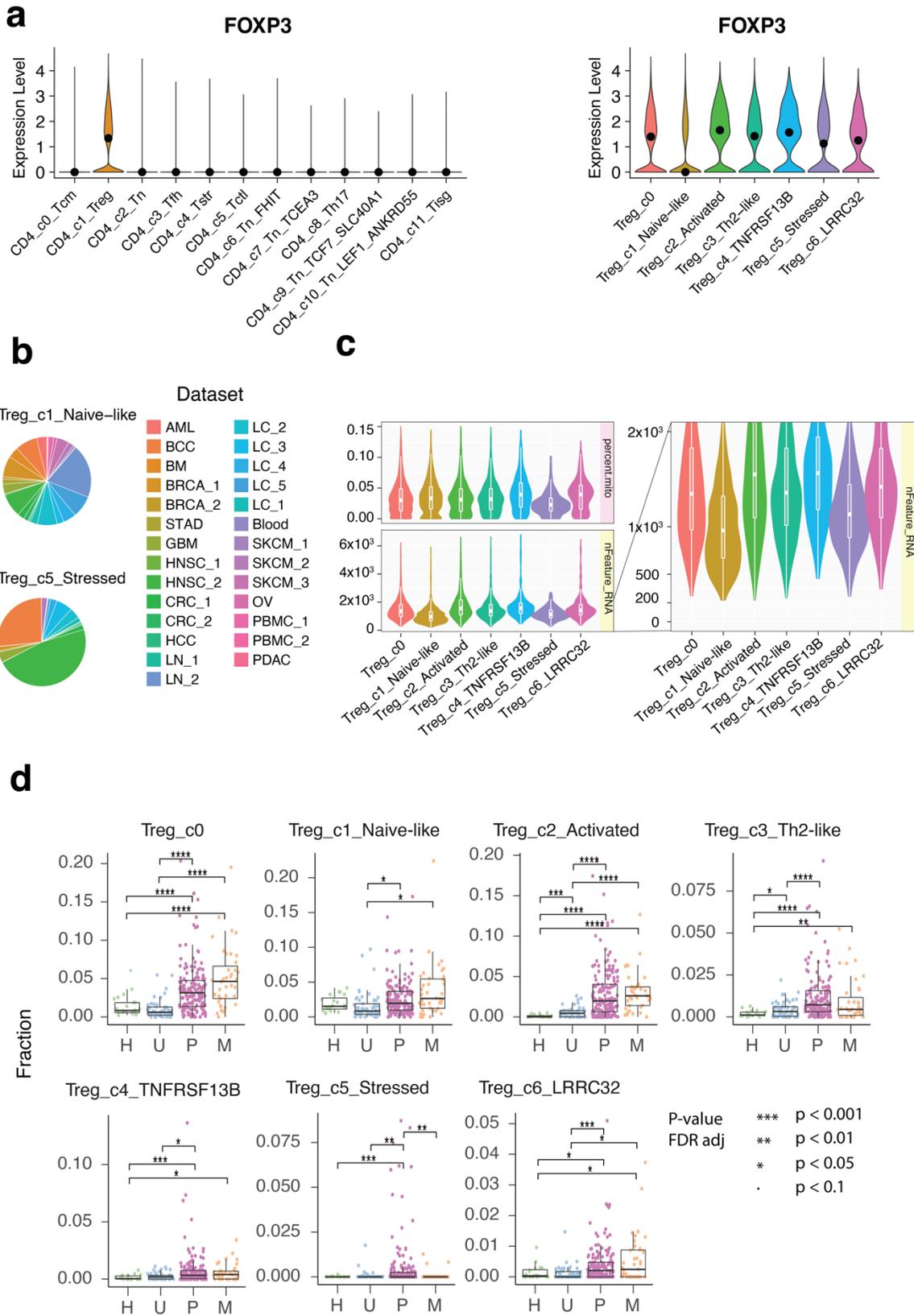
Supplementary Figure 3. CD4+ T cell clusters. (a) Box-plot showing the distribution of ROGUE scores across 12 CD4 T cell clusters. Each dot represents a single cell dataset (Supplementary Table 1). The sample number n of each CD8 T cell clusters from c0 to c11 are 27, 27, 27, 27, 27, 27, 26, 23, 27, 27, 24. (b) Bar graph showing the cohort compositions of 12 defined CD4 T cell clusters. H, normal tissues from healthy donors; U, tumor adjacent uninvolved normal tissues; P, primary tumor tissues; and M, metastatic tumor tissue; BM, bone marrow; LN, lymph node; PBMC, peripheral blood mononuclear cell. For uninvolved normal tissues, their corresponding organs of sample collection are labeled. For primary and metastatic tumor tissues, their cancer types ([abbreviations in TCGA format](#)) are labelled. (c) Box plot showing cell fractions of three representative CD4 cell subsets across 4 tissue types, stratified by tissue (H and U) and disease (P and M). Each dot represents a sample. Sample number n of each x axis from left to right are 6, 7, 8, 3, 3, 29, 8, 6, 16, 16, 46, 17, 4, 7, 5, 16, 8, 8, 4, 3, 4, 31, 3. Boxes, median \pm the interquartile range; whiskers, $1.5 \times$ interquartile range.

Supplementary Figure 4



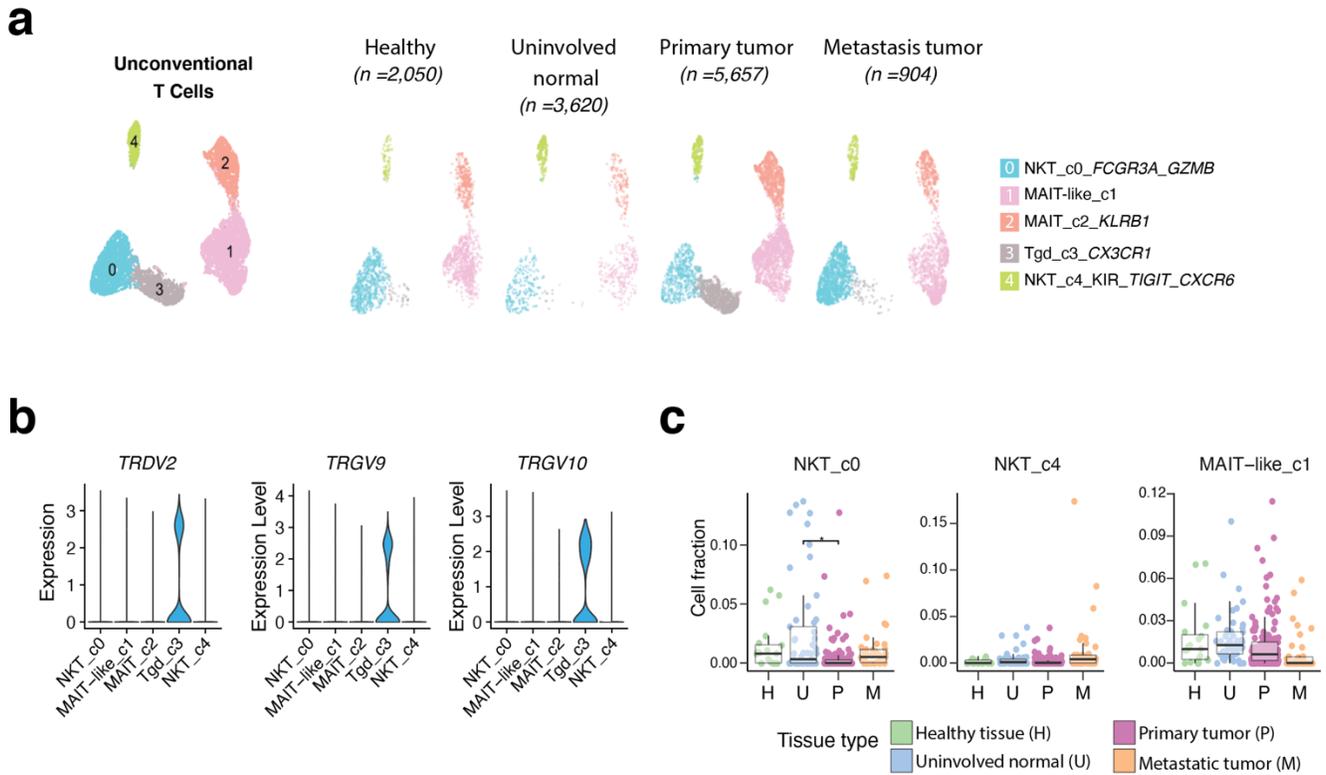
Supplementary Figure 4. Expression of previously published gene signatures across CD4 and CD8 T cell clusters. These gene signatures were downloaded from three recently published studies (*Liangtao Zheng, et al. Science. 2021. PMID: 34914499; Bassez et al. Nature Medicine 2021. PMID: 33958794; Han et al. Blood Cancer Research. 2022. PMID:35687817*). P values were calculated by the Kruskal-Wallis test. FDR adjusted p-values <0.0001 for all comparisons.

Supplementary Figure 5



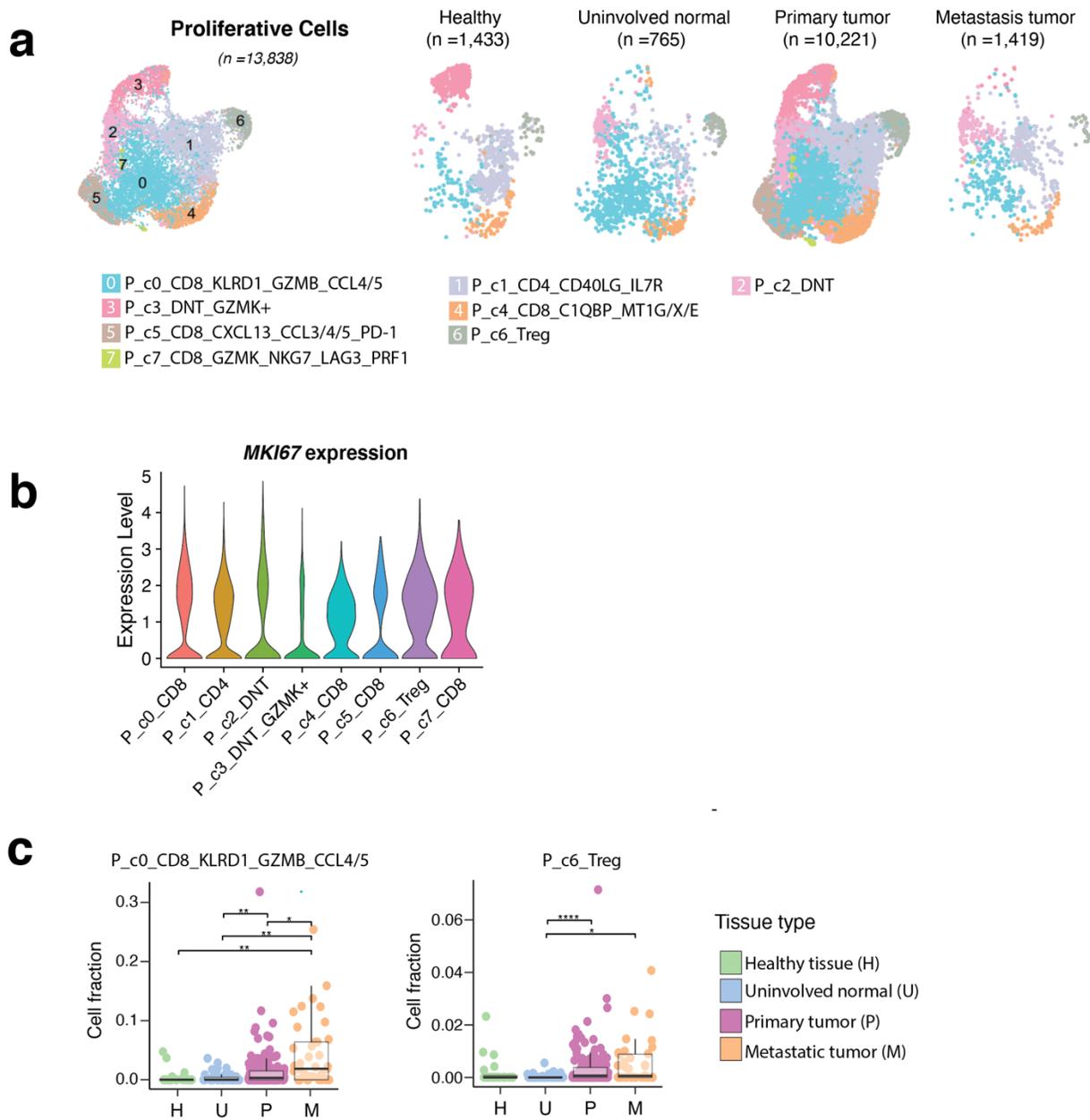
Supplementary Figure 5. CD4+ Treg cell subclusters. (a) Violin plots show the expression levels of *FOXP3* across 12 CD4 cell clusters (left) and 7 Treg cell subsets (right). (b) Pie plots show the dataset proportion in Treg_c1 and Treg_c5. (c) Violin and box plots show the count and feature levels of each Treg subsets. Cell number of each Treg subset from c0 to c6 are 9393, 7557, 6115, 2631, 1732, 1091, 1038. (d) Box plot comparing the fractions of 7 Treg cell subsets across 4 tissue types. Each dot represents a sample. H, normal tissues from healthy donors; U, tumor adjacent uninvolved normal tissues; P, primary tumor tissues; and M, metastatic tumor tissues. The one-sided Games-Howell test was applied to calculate the p values (sample number of each x axis from c0 to c6 are 20, 53, 158, 39.), followed by FDR (false discovery rate) correction. Boxes, median \pm the interquartile range; whiskers, 1.5 \times interquartile range.

Supplementary Figure 6



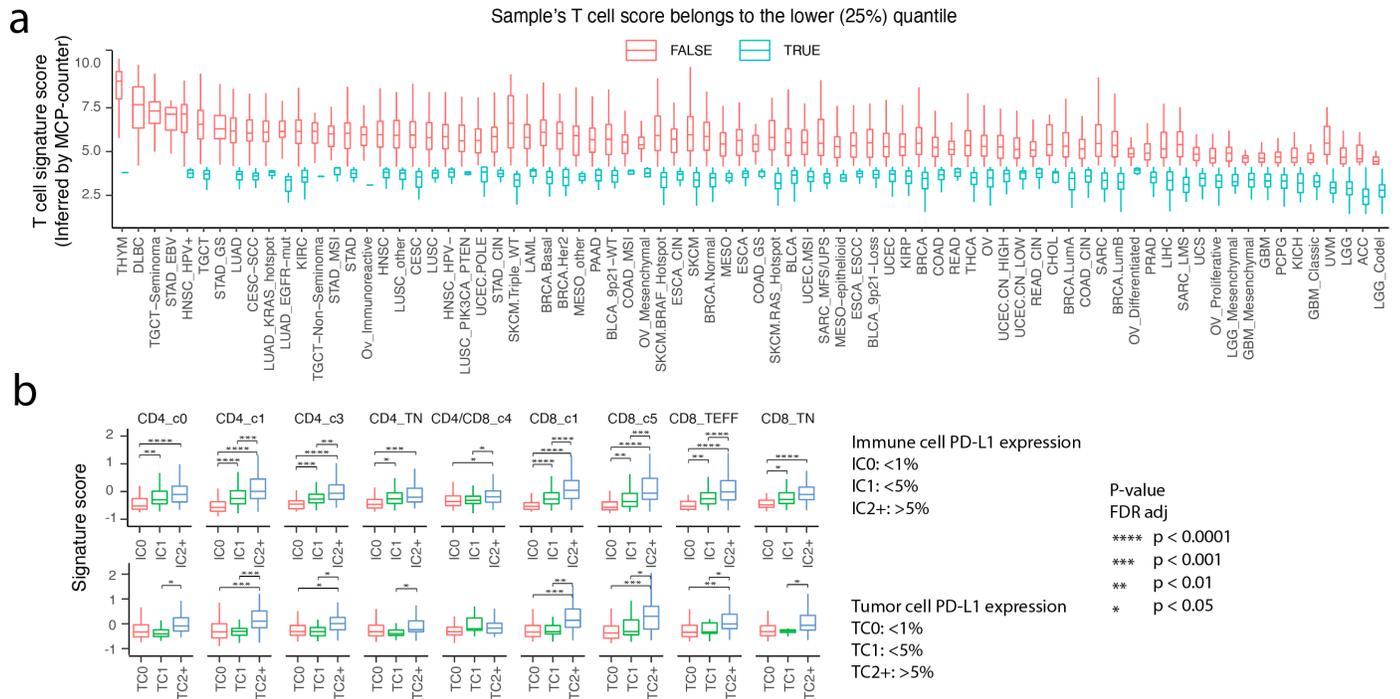
Supplementary Figure 6. Characterization of unconventional T cells. (a) UMAP view of 5 unconventional T cell clusters by tissue type. **(b)** Expression of Tgd related marker genes in 5 unconventional T cell clusters. **(c)** Box plot showing cell fractions of three innate cell subsets across 4 tissue groups. H, normal tissues from healthy donors; U, tumor adjacent uninvolved normal tissues; P, primary tumor tissues; and M, metastatic tumor tissues. The one-sided Games-Howell test was applied to calculate the p values (sample number of each x axis from left to right are 20, 53, 158, 39.), followed by FDR (false discovery rate) correction ($p_{U\text{vs}P} = 0.018$). FDR adjusted p value: *0.05; **0.01; ***0.001, ****0.0001. Boxes, median \pm the interquartile range; whiskers, 1.5 \times interquartile range.

Supplementary Figure 7



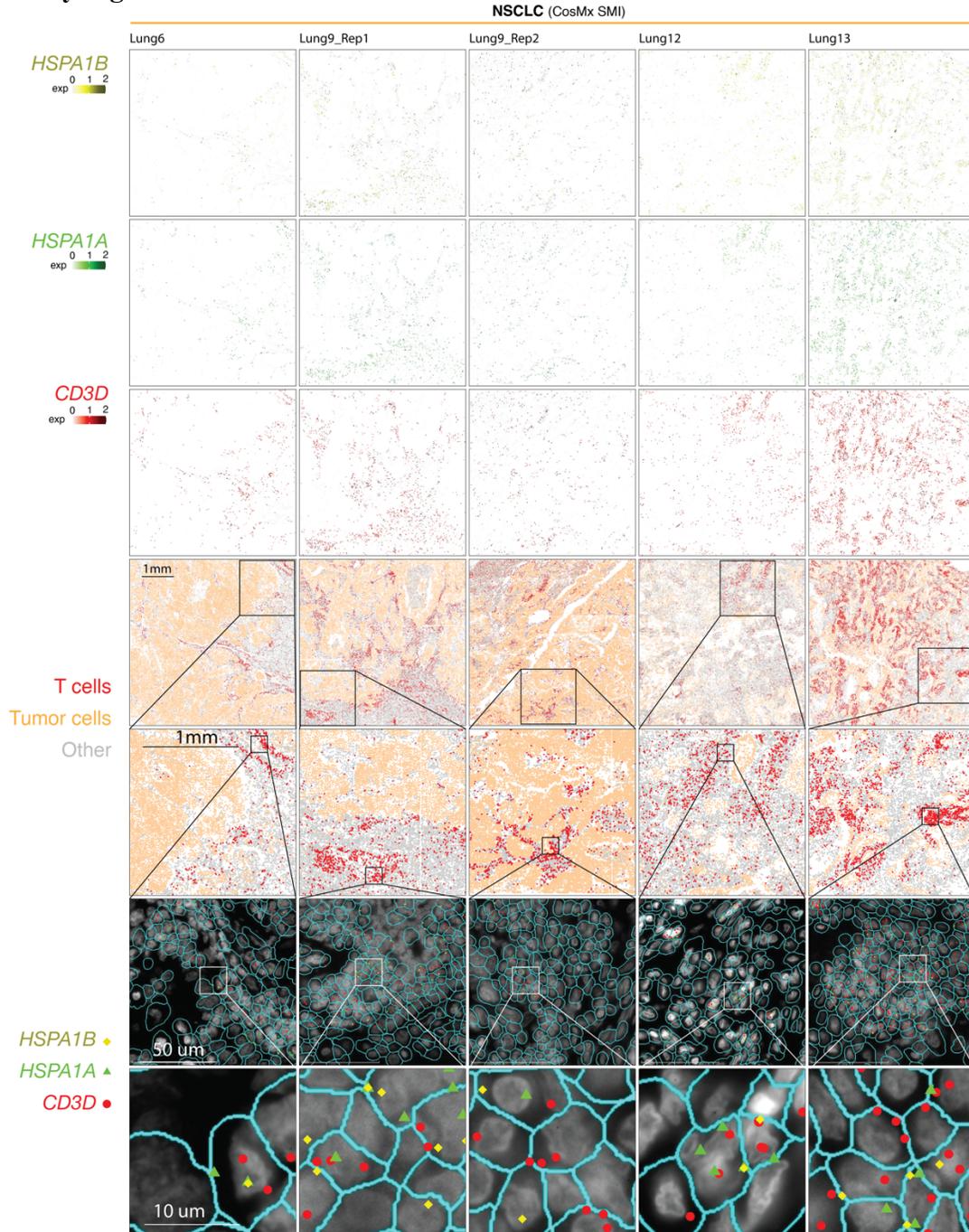
Supplementary Figure 7. Characterization of proliferative T cells. (a) UMAP view of 8 proliferative T cell clusters by tissue type. (b) Violin plot showing the expression levels of *MKI67* across proliferative T cell subclusters. (c) Box plot showing cell fractions of two proliferative cell subsets across 4 tissue groups. H, normal tissues from healthy donors; U, tumor adjacent uninvolved normal tissues; P, primary tumor tissues; and M, metastatic tumor tissues. The one-sided Games-Howell test was applied to calculate the p values (sample number of each x axis from left to right are 20, 53, 158, 39.), followed by FDR (false discovery rate) correction. FDR adjusted p value: *0.05; **0.01; ***0.001, ****0.0001. Boxes, median ± the interquartile range; whiskers, 1.5× interquartile range.

Supplementary Figure 8



Supplementary Figure 8: Correlation analysis using bulk RNA sequencing datasets. (a) The TCGA cohorts. Box plot showing the distribution of MCP-counter inferred signature scores for T cells across 52 TCGA cancer types and subtypes. For each cancer type and subtype, two boxes are shown, one (in blue) is for samples whose T cell signature scores belongs to the lower quantile (the bottom 25% of the ranked data among the 11,051 TCGA tumors with bulk RNA-seq data available, see the list in Supplementary Table 13) which was excluded from T cell deconvolution analysis, and the other (in red) is for the rest of samples. Sample number *n* of each box from left to right are 121, 1, 48, 0, 74, 0, 31, 0, 68, 7, 147, 8, 52, 1, 555, 14, 63, 3, 97, 4, 132, 3, 30, 1, 75, 3, 282, 24, 78, 1, 415, 32, 526, 61, 506, 55, 497, 39, 282, 26, 215, 15, 391, 43, 31, 15, 218, 24, 146, 27, 150, 40, 80, 5, 29, 5, 163, 19, 67, 14, 61, 4, 86, 19, 111, 40, 66, 6, 340, 133, 68, 12, 75, 12, 158, 35, 43, 7, 67, 25, 329, 98, 112, 29, 95, 31, 56, 26, 102, 34, 46, 7, 74, 19, 445, 119, 391, 92, 409, 157, 139, 27, 871, 335, 243, 75, 235, 70, 86, 21, 123, 42, 118, 31, 178, 65, 399, 175, 155, 106, 27, 18, 141, 77, 43, 23, 339, 208, 225, 194, 42, 41, 31, 26, 43, 36, 16, 29, 33, 56, 20, 33, 59, 111, 54, 132, 14, 35, 17, 63, 41, 483, 16, 62, 4, 170. **(b)** The cohort of urothelial carcinoma from *Mariathasan et al.* Box plot showing associations between the signature scores of 9 unique T cell states and levels of PD-L1 expression in immune cells (top, sample number *n* of *x* axis from left to right for each panel are 83, 112, 102) or tumor cells (bottom, sample number *n* of *x* axis from left to right for each panel are 238, 17, 42) as measured via immunohistochemistry. IC, immune cell; TC, tumor cell. The one-sided Games-Howell test was applied to calculate the p values, followed by FDR (false discovery rate) correction. FDR adjusted p value: * ≤ 0.05 ; ** ≤ 0.01 ; *** ≤ 0.001 , **** ≤ 0.0001 . Boxes (b-c), median \pm the interquartile range; whiskers, $1.5 \times$ interquartile range.

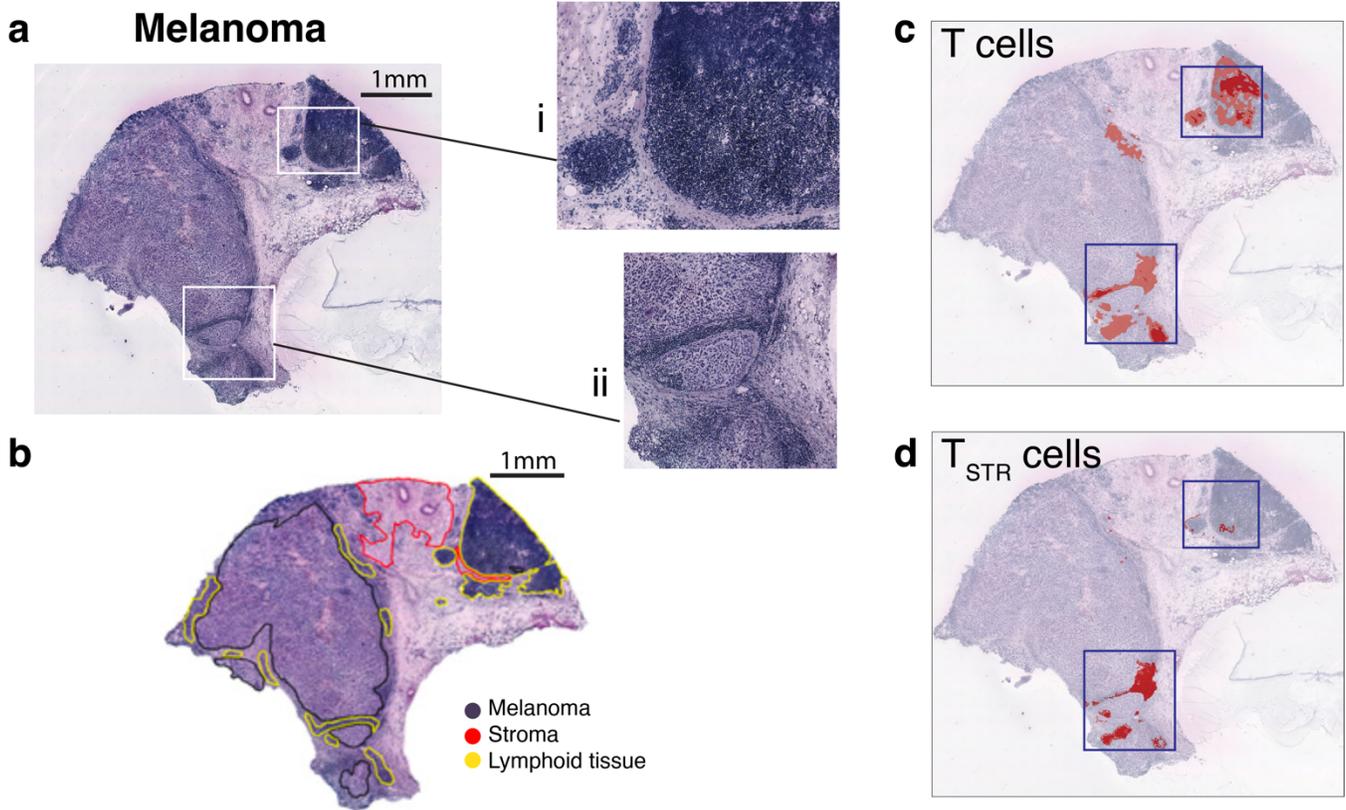
Supplementary Figure 9



Supplementary Figure 9: Detection of *in situ* HSPA1A and HSPA1B expression in tumor-infiltrating T cells in five NSCLC samples by CosMx. The representative tissue sections from two NSCLC (section “Lung 6”, “Lung9_Rep1”, “Lung5_Rep2”, “Lung12”, “Lung13”) were profiled. The sample “Lung9” included two consecutive tissue sections. (**Row 4**) Cells in physical locations (x, y coordinates). Color denotes cell type. Spatial mapping of *CD3D* (**Row 3**), *HSPA1A* (**Row 2**), and *HSPA1B* (**Row 1**) expression in T cells in “Lung 6”, “Lung9_Rep1”, “Lung5_Rep2”, “Lung12”, “Lung13” (the same area as **Row 4**). (**Row 5**) A zoom-in view of a representative area of (i) showing two lymphocyte aggregates enriched with T cells. (**Row 6**) a zoom-in view of (**Row 5**) showing subcellular localization of *CD3D*, *HSPA1A*, and *HSPA1B* transcripts. (**Row 7**) a zoom-in view of (**Row 6**) showing co-localization of *CD3D*, *HSPA1A*, and *HSPA1B* transcripts in the same cells. For (**Row 6**) and (**Row 7**), cell segmentation was done by the original study. The outlines of cell nuclei were determined based

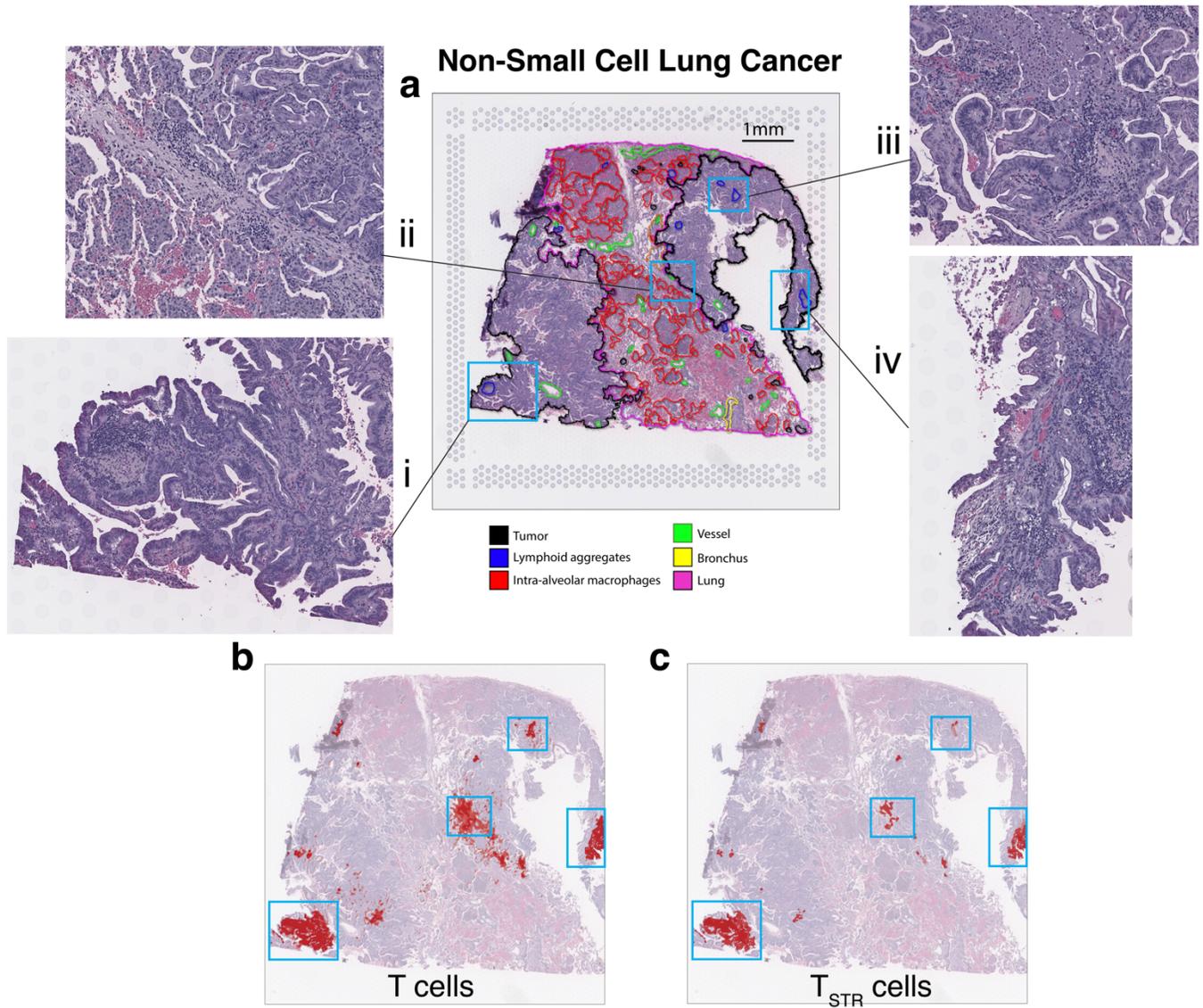
on DAPI staining and the cell boundaries were determined based on morphology markers for membrane (e.g., CD298) combined with machine learning.

Supplementary Figure 10



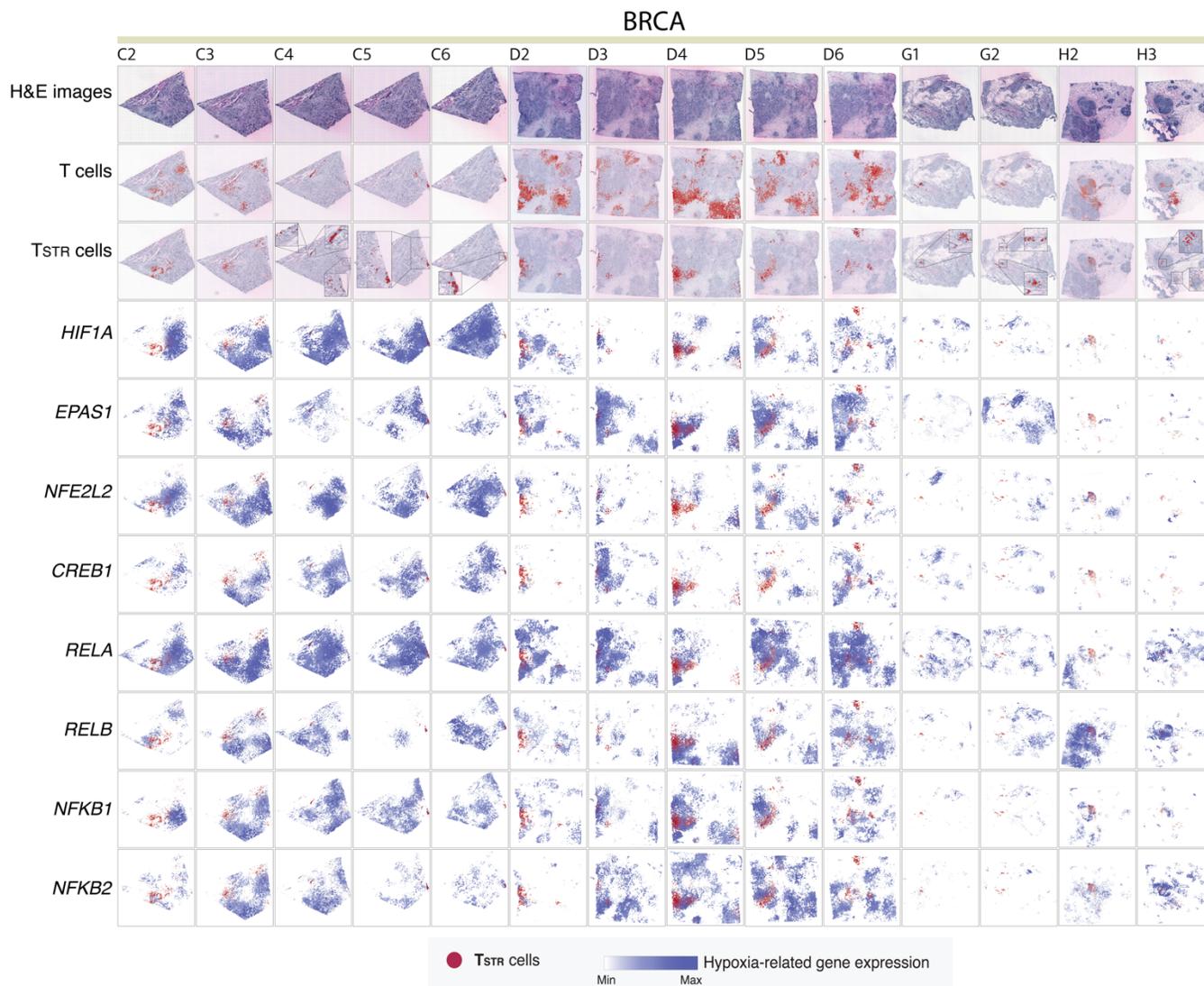
Supplementary Figure 10: Detection of T_{STR} cells in a melanoma sample using spatial transcriptomics. (a) H&E stained tissue image with zoom-in view of two regions (while squares) showing the location of lymphocytes (i and ii). (b) The same image as in (a) but with pathologic annotation (obtained from the original study). (c) Mapping of T cells and (d) the T_{STR} subset on the same histology image as in (a) based on corresponding spatial transcriptomics data generated on the same tissue section (spatial transcriptomics, ST, the slide dimension is 6,200 x 6,600 μm). No replicates are available for this tissue section.

Supplementary Figure 11



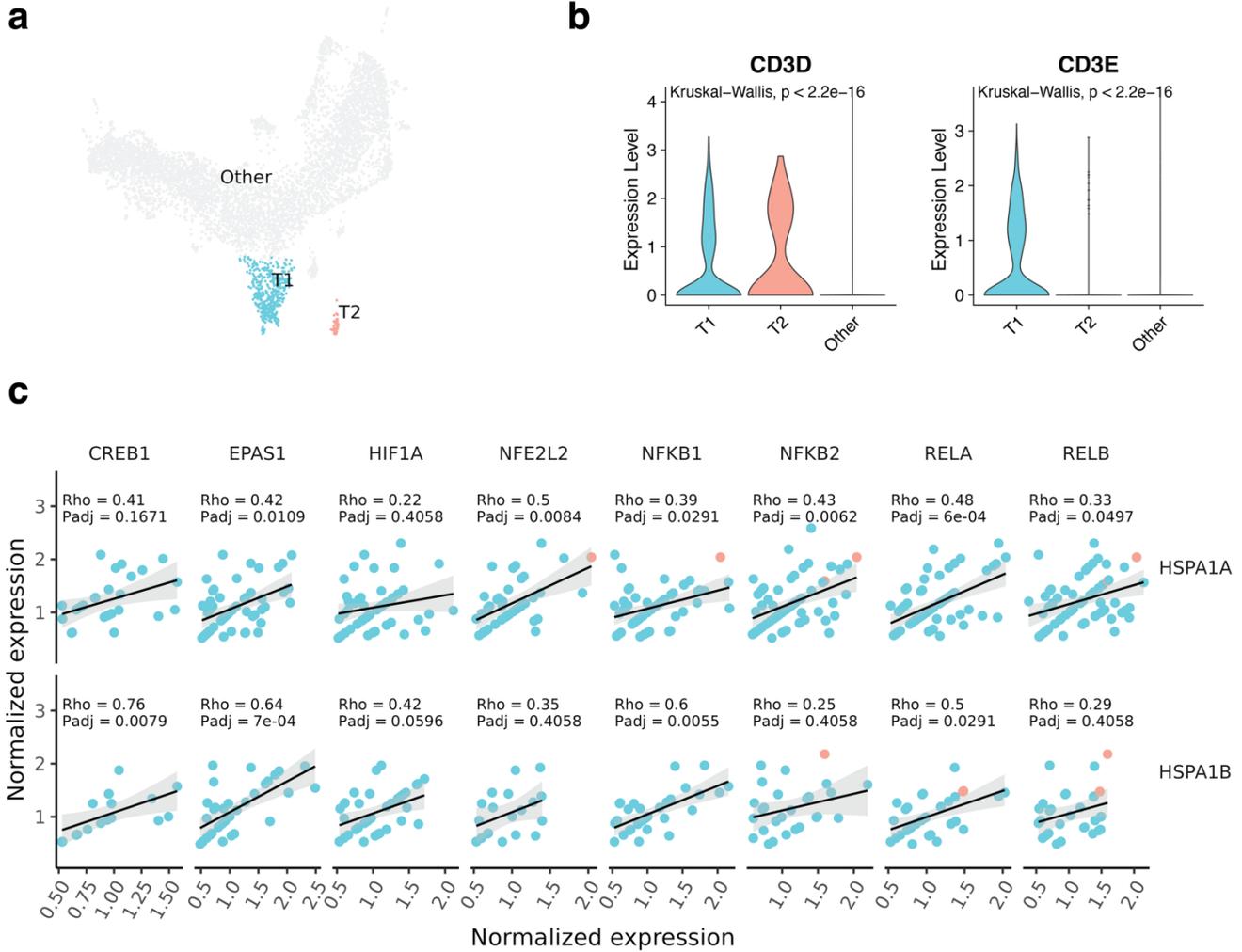
Supplementary Figure 11: Detection of T_{STR} cells in a NSCLC sample using spatial transcriptomics. (a) H&E stained tissue image with pathologic annotation. Zoom-in view of 4 representative regions (cyan squares) showing lymphoid aggregates (i, iii, iv) and diffuse lymphocytic infiltration (ii). **(b)** Mapping of T cells and **(c)** T_{STR} cells on the same histology image as in **(a)**, based on corresponding spatial transcriptomics data generated on the same tissue section (10x Genomics, Visium, the capture area is 6.5 x 6.5 mm). No replicates are available for this tissue section.

Supplementary Figure 12

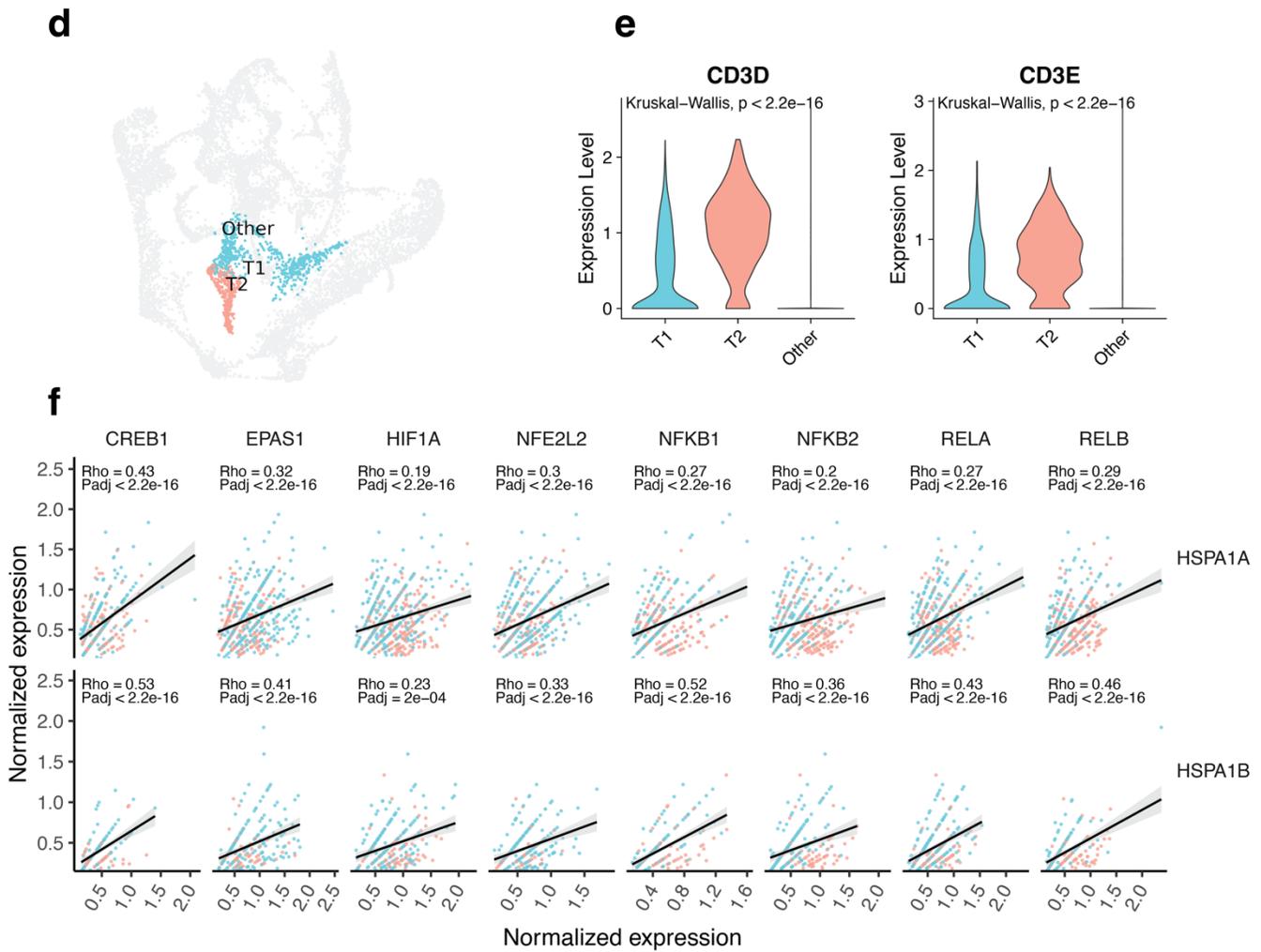


Supplementary Figure 12: Pan-cancer detection of TSTR cells by spatial transcriptomics (10x Genomics, Visium) and Co-mapping of T_{STR} cells and hypoxia-related gene expression using spatial transcriptomics. Extra representative tissue replicate sections (C2~6, D2~6, G1~2, H2~3) of BRCA are shown. **(Top row)** H&E stained tissue image. **(Second row)** Mapping of T cells and **(third row)** the TSTR cells on the same histology. BRCA, breast cancer; **(The rest of other rows)** spatial co-mapping of T_{STR} cells (in red) and hypoxia-related gene expression (in blue, the darker the color, the higher the level of gene expression) on the same regions as shown in the top row.

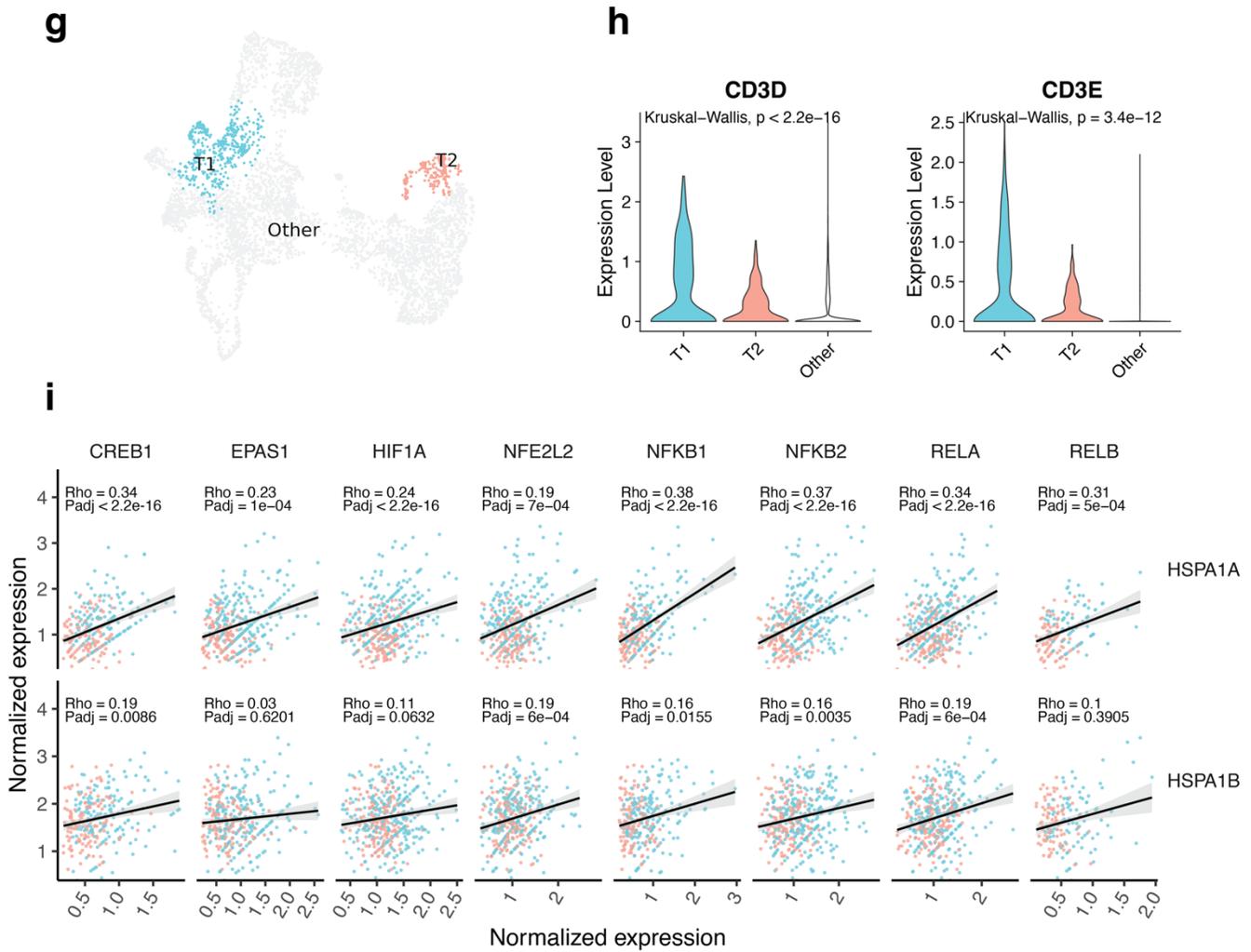
Supplementary Figure 13



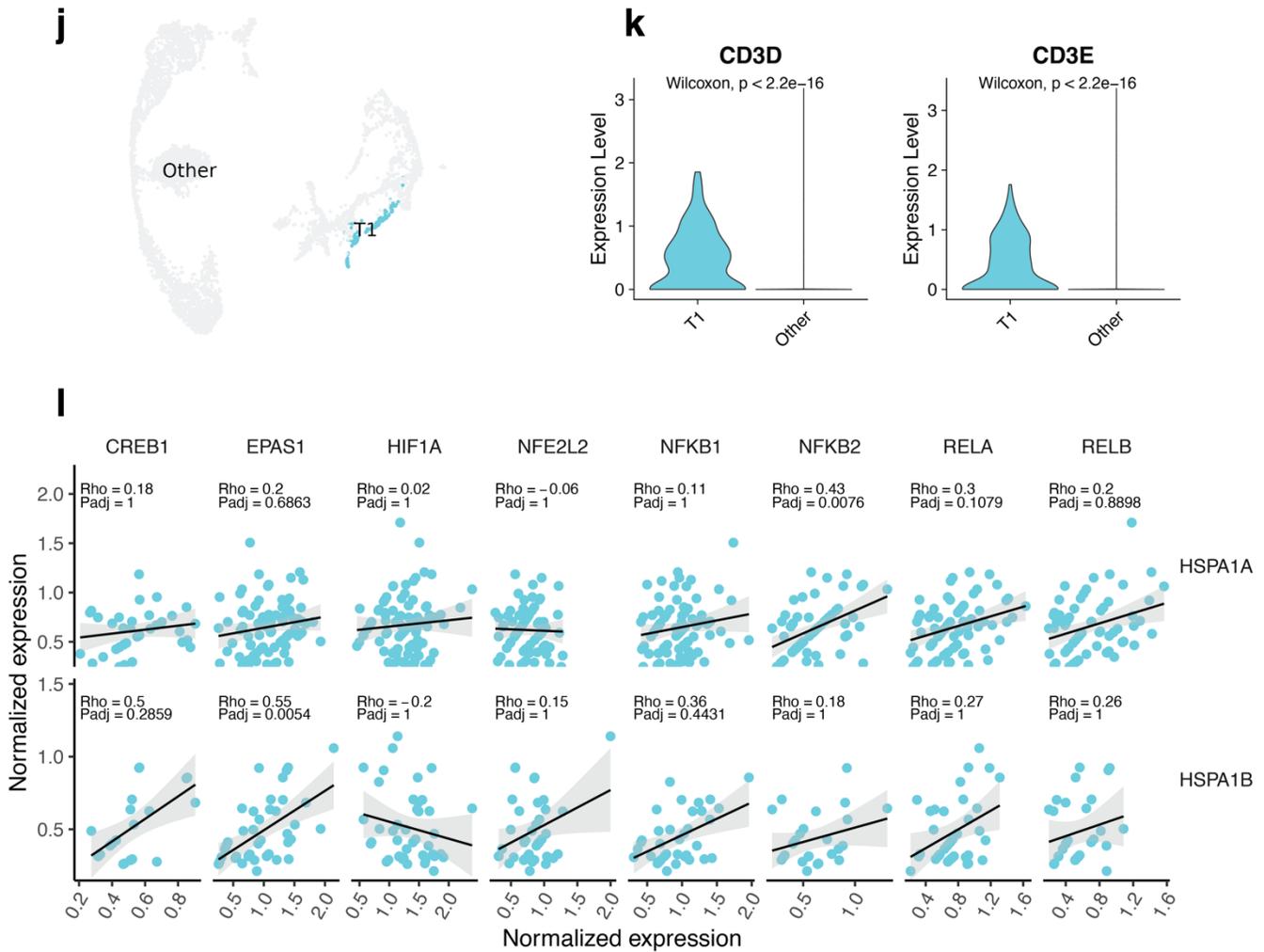
Correlation between stressed and hypoxia related gene expression in T cell enriched spots in BRCA spatial transcriptomics datasets (10x Genomics, Visium, Section C1~6, D1~6, G1~3, H1~3). (a) UMAP plot shows T cell enriched spot clusters T1 and T2. (b) Violin plots show expression level of T cell marker *CD3D* and *CD3E* across all spot clusters. (c) Scatter plots show the correlation between stressed and hypoxia related gene expression. Grey bands show 95% confidence level. Padj is the Holm-adjusted two-sided spearman correlation test p value.



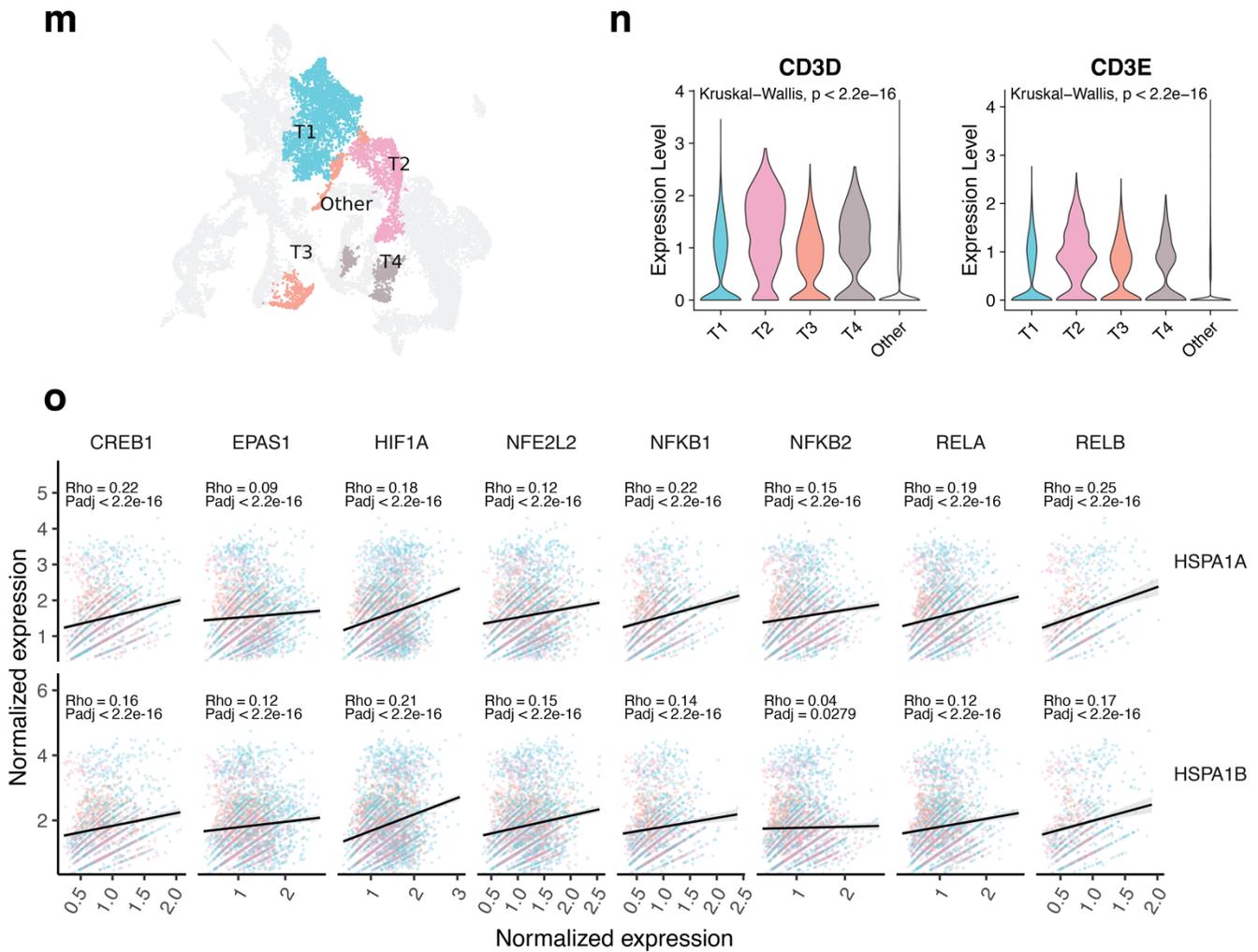
Correlation between stressed and hypoxia related gene expression in T cell enriched spots in GAC spatial transcriptomics datasets (10x Genomics, Visium, Section 14C and 14D). (d) UMAP plot shows T cell enriched spot cluster T1 and T2. **(e)** Violin plots show expression level of T cell marker *CD3D* and *CD3E* across all spot clusters. **(f)** Scatter plots show the correlation between stressed and hypoxia related gene expression. Grey bands show 95% confidence level. Padj is the Holm-adjusted two-sided spearman correlation test p value.



Correlation between stressed and hypoxia related gene expression in T cell enriched spots in LUAD spatial transcriptomics datasets (10x Genomics, Visium, Section 14C and 14D). (g) UMAP plot shows T cell enriched spot cluster T1 and T2. (h) Violin plots show expression level of T cell marker *CD3D* and *CD3E* across all spot clusters. (i) Scatter plots show the correlation between stressed and hypoxia related gene expression. Grey bands show 95% confidence level. Padj is the Holm-adjusted two-sided spearman correlation test p value.

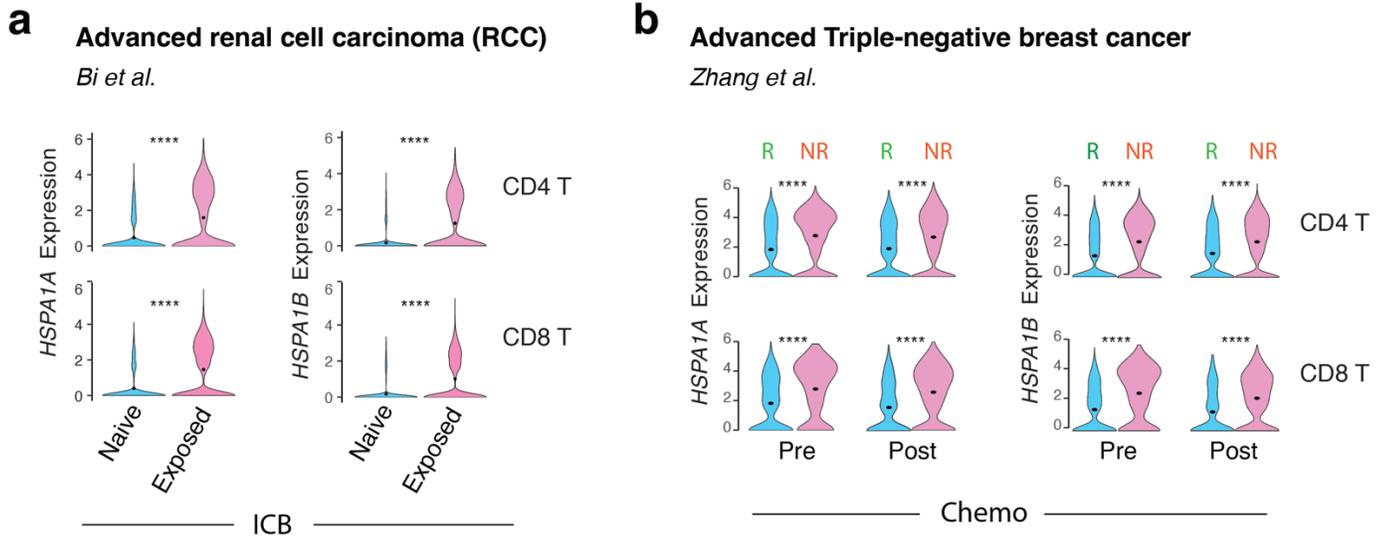


Correlation between stressed and hypoxia related gene expression in T cell enriched spots in CSCC spatial transcriptomics datasets (10x Genomics, Visium, Section P2_ST_rep1~3). **(j)** UMAP plot shows T cell enriched spot cluster T1. **(k)** Violin plots show expression level of T cell marker *CD3D* and *CD3E* across all spot clusters. **(l)** Scatter plots show the correlation between stressed and hypoxia related gene expression. Grey bands show 95% confidence level. Padj is the Holm-adjusted two-sided spearman correlation test p value.



Correlation between stressed and hypoxia related gene expression in T cell enriched spots in ccRCC spatial transcriptomics datasets (10x Genomics, Visium, Section ffpe_c_2, ffpe_c_7, ffpe_c_34, ffpe_c_45 and ffpe_c_51). (m) UMAP plot shows T cell enriched spot cluster T1, T2, T3 and T4. **(n)** Violin plots show expression level of T cell marker *CD3D* and *CD3E* across all spot clusters. **(o)** Scatter plots show the correlation between stressed and hypoxia related gene expression. Grey bands show 95% confidence level. Padj is the Holm-adjusted two-sided spearman correlation test p value.

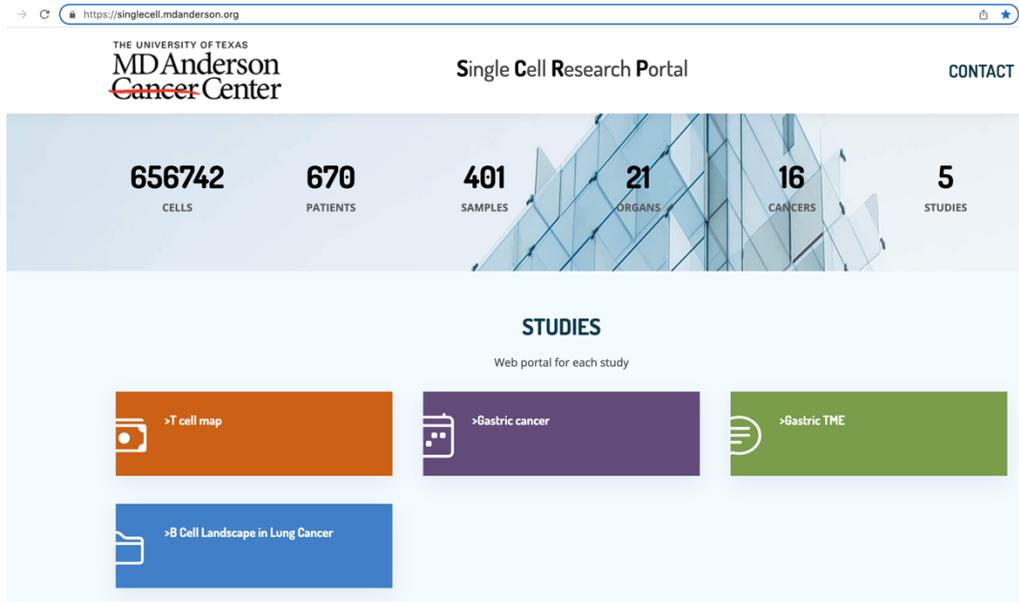
Supplementary Figure 14



Supplementary Figure 14: Expression of *HSPA1A* and *HSPA1B* in tumor-infiltrating T cells from two scRNA-seq cohorts. (a) Increased expression of *HSPA1A* (left) and *HSPA1B* (right) in tumor-infiltrating CD4 and CD8 T cells from patients with advanced renal cell carcinoma (RCC) who exposed to immune checkpoint blockade (ICB) compared to those who were ICB naïve. The scRNA-seq data were downloaded from the study by Bi et al. (*Bi et al. Cancer Cell. 2021. PMID: 33711272*). **(b)** Increased expression of *HSPA1A* (left) and *HSPA1B* (right) in tumor-infiltrating CD4 and CD8 T cells from patients with advanced triple-negative breast cancer (TNBC) who were non-responders (had stable diseases or progressive diseases) post chemotherapy compared to those who were partial responders. The scRNA-seq data were downloaded from the study by Zhang et al. (*Zhang et al. Cancer Cell. 2021. PMID: 34653365*). ****FDR adjusted p value <0.0001. P values were calculated using the two-sided Welch’s t-test. All p values in (a) are less than 2.2e-16. In (b), for p values of CD4 from left to right, $p_{\text{Pre_R-vs-NR}} < 2.2\text{e-}16$, $p_{\text{Post_R-vs-NR}} = 3.66\text{e-}10$, for p values of CD8 from left to right, $p_{\text{Pre_R-vs-NR}} < 2.2\text{e-}16$, $p_{\text{Post_R-vs-NR}} = 3.66\text{e-}10$, $p_{\text{Pre_R-vs-NR}} = 3.66\text{e-}10$, $p_{\text{Post_R-vs-NR}} = 1.83\text{e-}7$.

Supplementary Figure 15

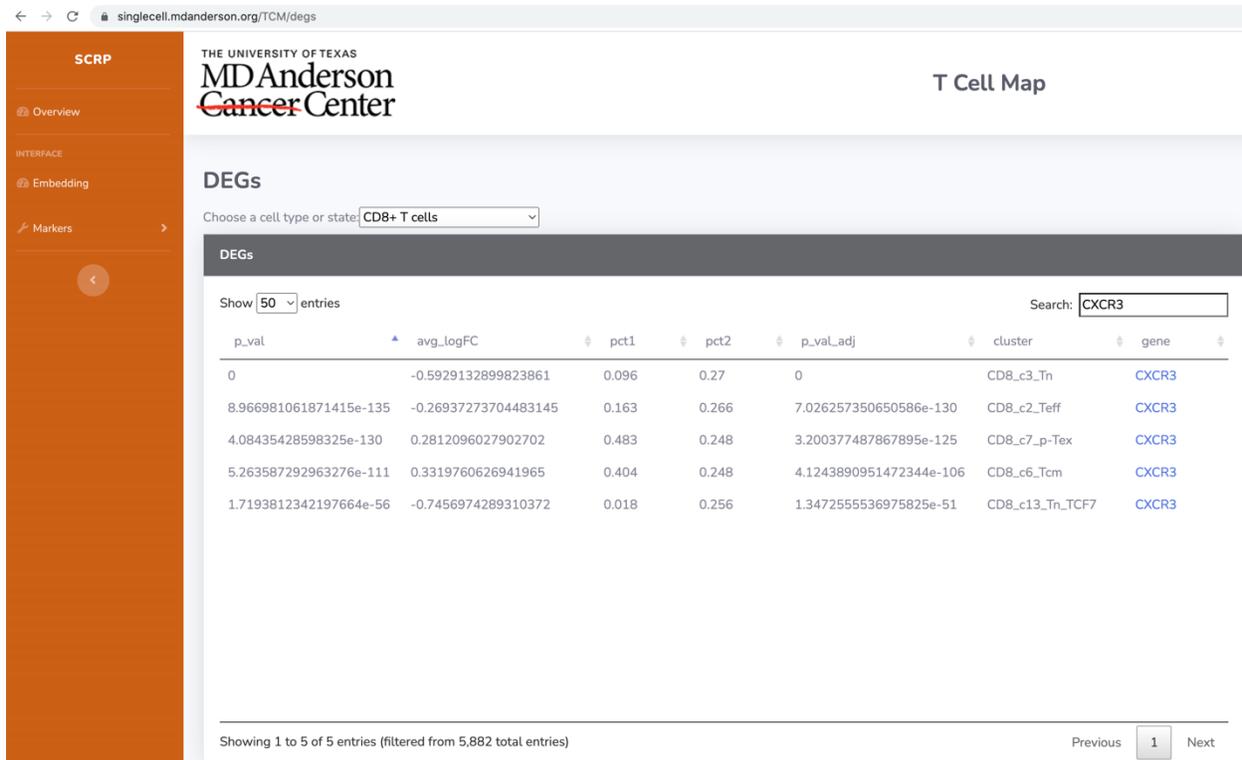
SCRP main page (<https://singlecell.mdanderson.org/>).



T Cell Map (<https://singlecell.mdanderson.org/TCM/>) provides an overview of the study, such as number of cells, samples, patients, cancer types, and by sample cell number distribution, cell type composition, tissue composition and cell composition, etc., which are customized for each project based on their metadata.



DEGs allows to search genes of interest (<https://singlecell.mdanderson.org/TCM/degs>), check the average gene expression level and fraction of cells expressing a gene in a T cell cluster as compared to the rest of other T cell clusters for a selected cell type.



View gene expression on UMAP (<https://singlecell.mdanderson.org/TCM/degs>), enables to generate UMAP plots to visualize the expression distribution of genes of interest.

