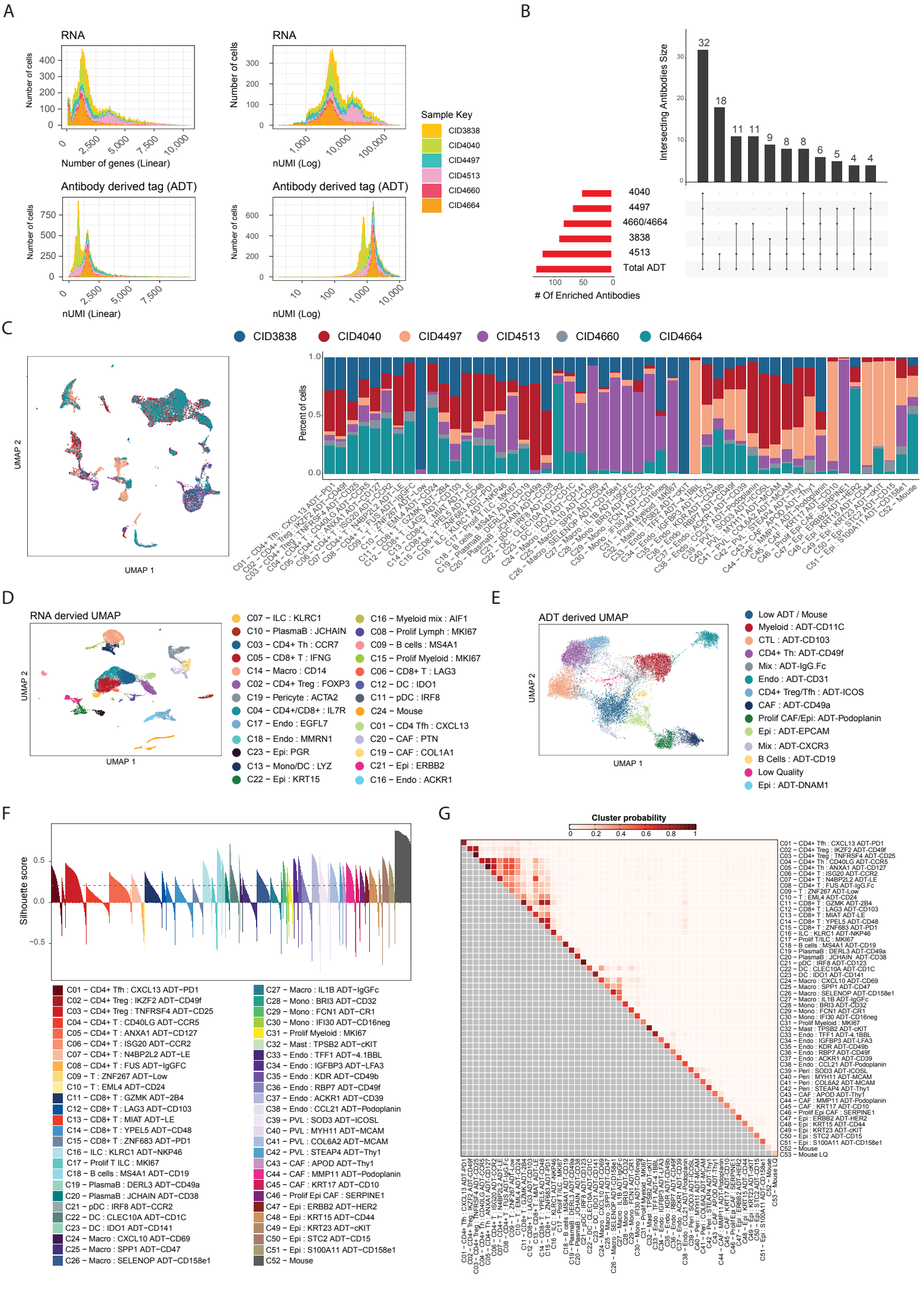
Extended Data Figure 1: Control metrics of CITE-seq data derived from 6 breast cancer samples. (A) Number of genes and UMI counts for both ADT and RNA assays, and mitochondrial content proportion for each sample, information of patient sample disease are available in Supplementary Table S2. **(B)** UpSet plot of CITE-seq antibodies found to be commonly enriched across samples. **(C)** UMAP of cells grouped by patient sample, and the proportion of each cluster derived from each patient. **(D)** UMAP of all cell clusters derived solely from RNA analysis. **(E)** UMAP of all cell clusters derived solely from ADT sequencing. **(F)** Silhouette score of each cluster for the evaluation of cluster stability. **(G)** Cluster stability matrix indicates the probability that a random cell from each paired cluster would be re-assigned to the same cluster following bootstrapping. High probability means paired clusters contain more related cells than not.

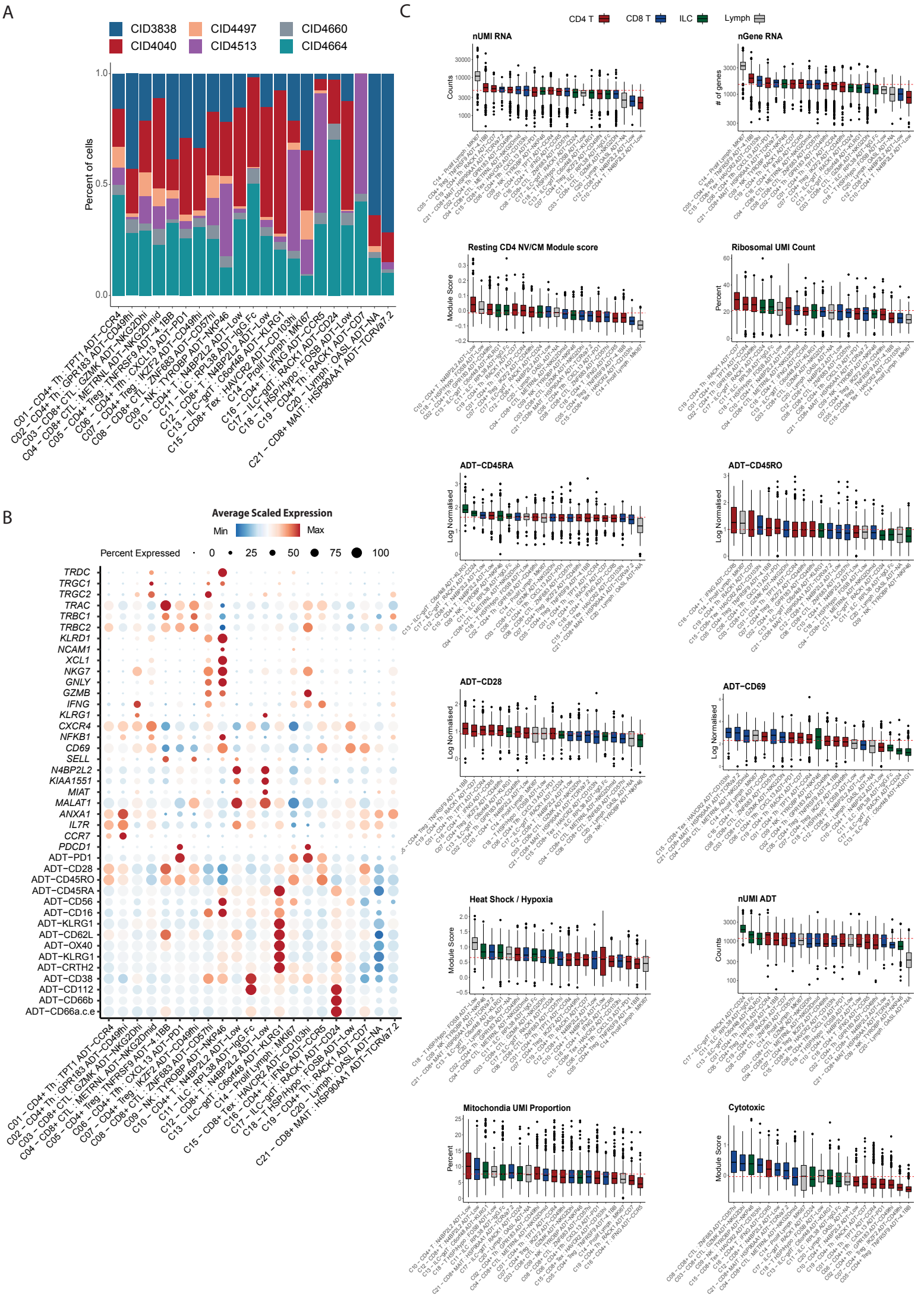
Extended Data Figure 1



Extended Data Figure 2: Integrated RNA + ADT WNN clustering analysis across T cell populations. T and ILC populations were stratified as described in Methods. **(A)** Proportion of patient cells from each sample belonging to each T/ILC cluster. **(B)** Dotplot visualisation of the expression of RNA and ADT markers of interest on a z-score normalised scale. **(C)** Log normalised expression of factors associated with lymphocyte activity and naïve/memory activation status. Bar plots are coloured by lymphocyte lineage. nUMI refers to the number of unique molecular identifiers derived from the RNA assay, and nGene refers to the number of genes identified. Heat shock/Hypoxia and Cytoxic enrichment scores were calculated using Seurats "*ModuleScore"* using genes listed in Table S4. **(D)** Radar plot projecting the gene signature module scores, scaled 0-1, for each respective cluster. Grey silhouette marks the module score profile when averaged across all clusters combined. **(E)** UMAP of the T cell/ILC dataset when clustered on RNA features alone. **(F)** Alluvial plot visualising the relationship of assigned cell clusters when derived from RNA alone versus RNA and ADT integration. **(G)** Cluster stability matrix shows the probability that a random cell from each paired cluster would be re-assigned to the same cluster following bootstrapping. High probability means paired clusters contain more related cells than not. **(H)** Silhouette score of each WNN-derived cluster calculated from RNA alone. Negative values indicate instability. **(I)** Violin plot showing normalised expression values of canonical markers of lymphocytes, myeloid, epithelial, and mesenchymal lineages. **(J)** Expression levels of isotype controls and mouse spike-in. (**K**) Log normalised expression of CD69 and CD103, sorted from high to low (left to right). Bar plots are coloured by broad lineage. **(L)** Pearson correlation coefficient values of select pairing of ADT markers stratified by cell cluster. The global Pearson coefficient value for each ADT pair is provided, please see Table S5 for Pearson coefficient values for each cluster. **(M)** AUCell enrichment score of signatures derived from Zemin et al.[36] pan cancer CD8+ T-cell exhausted (Tex) and Li et al. [53] dysfunctional CD8 T-cell, sorted from high to low (left to right). The top 50 differentially expressed genes were used towards calculating enrichment score. Red text indicates populations discussed in the text. **(N)** AUCell enrichment score of signatures derived from Savas et al. [6] human breast cancer CD8+ Tissue resident memory T cells (Trm). Boxplots are sorted from high to low (left to right) for each cluster. **(O)** Log normalised expression of ADT markers CD49a and CD39 across T-cell and ILC clusters, sorted from high to low (left to right). Box plots are coloured by lymphocyte lineage. **(P)** RNA & ADT co-expression heatmap of features (rows) known to be relevant to activation, migration, exhaustion and tissue residency of T cells. Red text highlights ADT markers and green text highlights RNA markers. Boxplots at the top of the heatmap show the protein expression of CD49f grouped by T cell activity (see Methods), sorted from high to low (left to right). Percent_ribo refers to the proportion of ribosomal counts relative to all other expressed genes. "CD4/CD8 rest" is a published signature score based on [41] that reflects enrichment of genes associated with inactive/resting CD4 & CD8 T cells.

In all boxplots (**J, K, M, N, O,P**) middle line marks the median value, the lower and upper hinges mark the 25% and 75% quantile, the whiskers correspond to the 1.5 times the interquartile range, the black dot's mark outliers. Red line indicates the median expression for each marker across clusters.

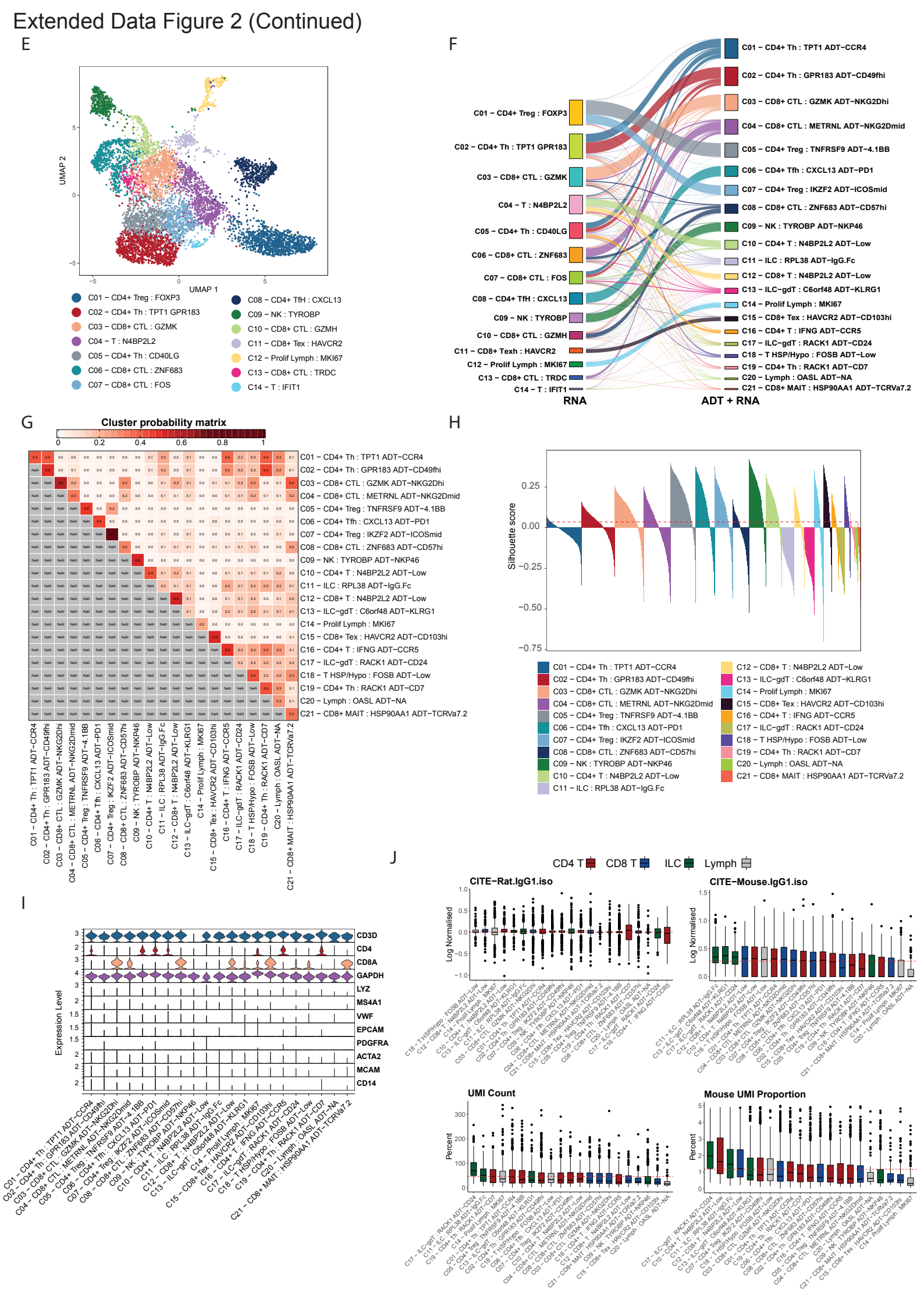
Extended Data Figure 2



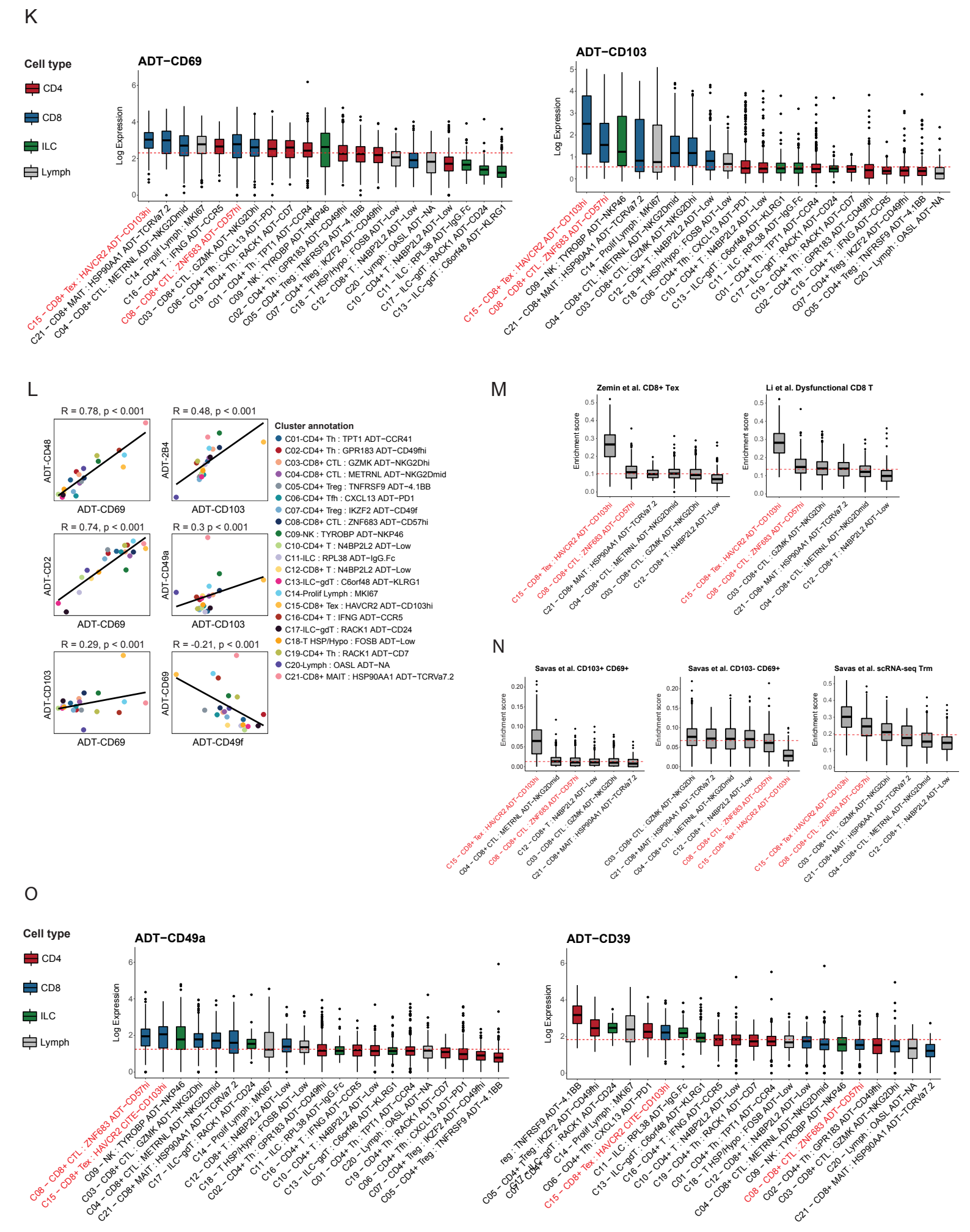
Extended Data Figure 2 (Continued)

D

Extended Data Figure 2 (Continued)

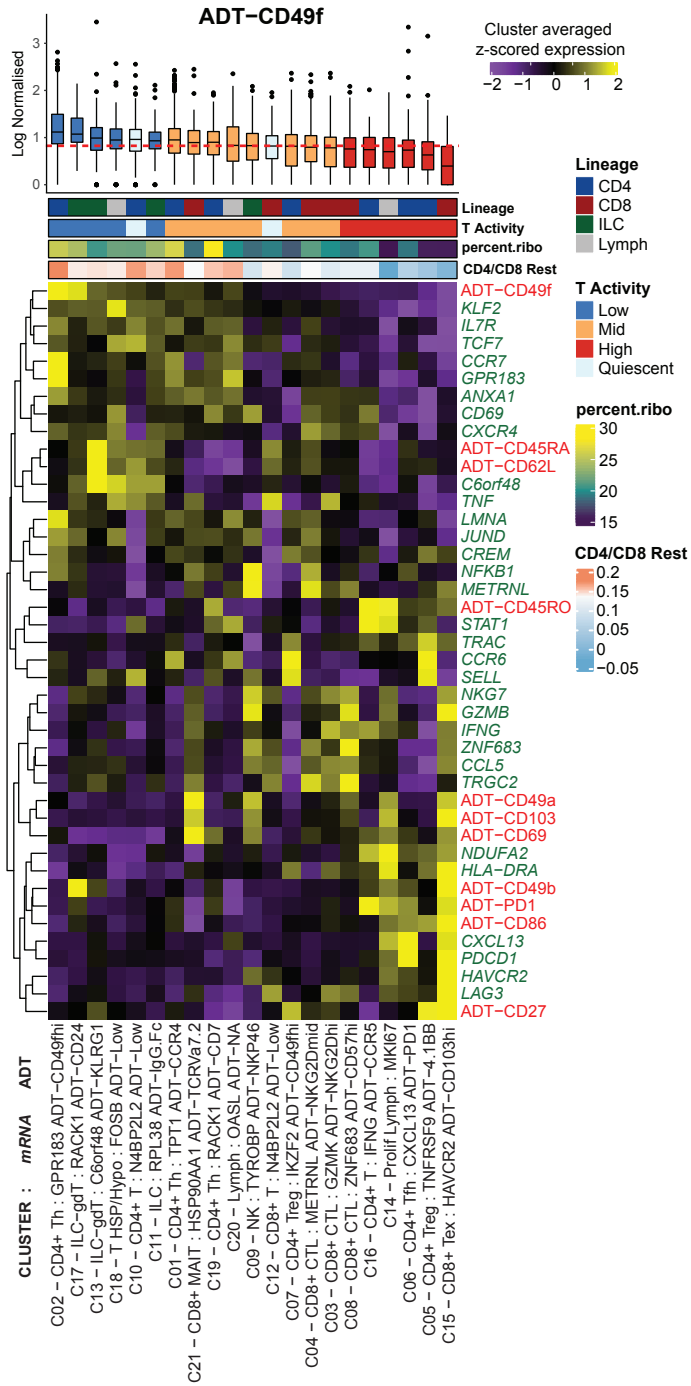


Extended Data Figure 2 (Continued)



Extended Data Figure 2 (cont.)

P



Extended Data Figure 3. The expression of protein and RNA features of TIL activation and tissue residency markers in breast cancers.

(A) Heatmap of top 5 differentially enriched GO biological process pathways for each cluster based on transcriptome.

(B) Dotplot of gene expression (subcluster average) of Tfh factors across the clusters

(C) Hierarchical clustering of differentially expressed genes between 3 identified Tfh states.

(D) Flow cytometric analysis of anti-CXCR5 antibody binding on CD4+ and CD19+ PBMC before and after being subjected to the standard tumor dissociation protocol (methods).

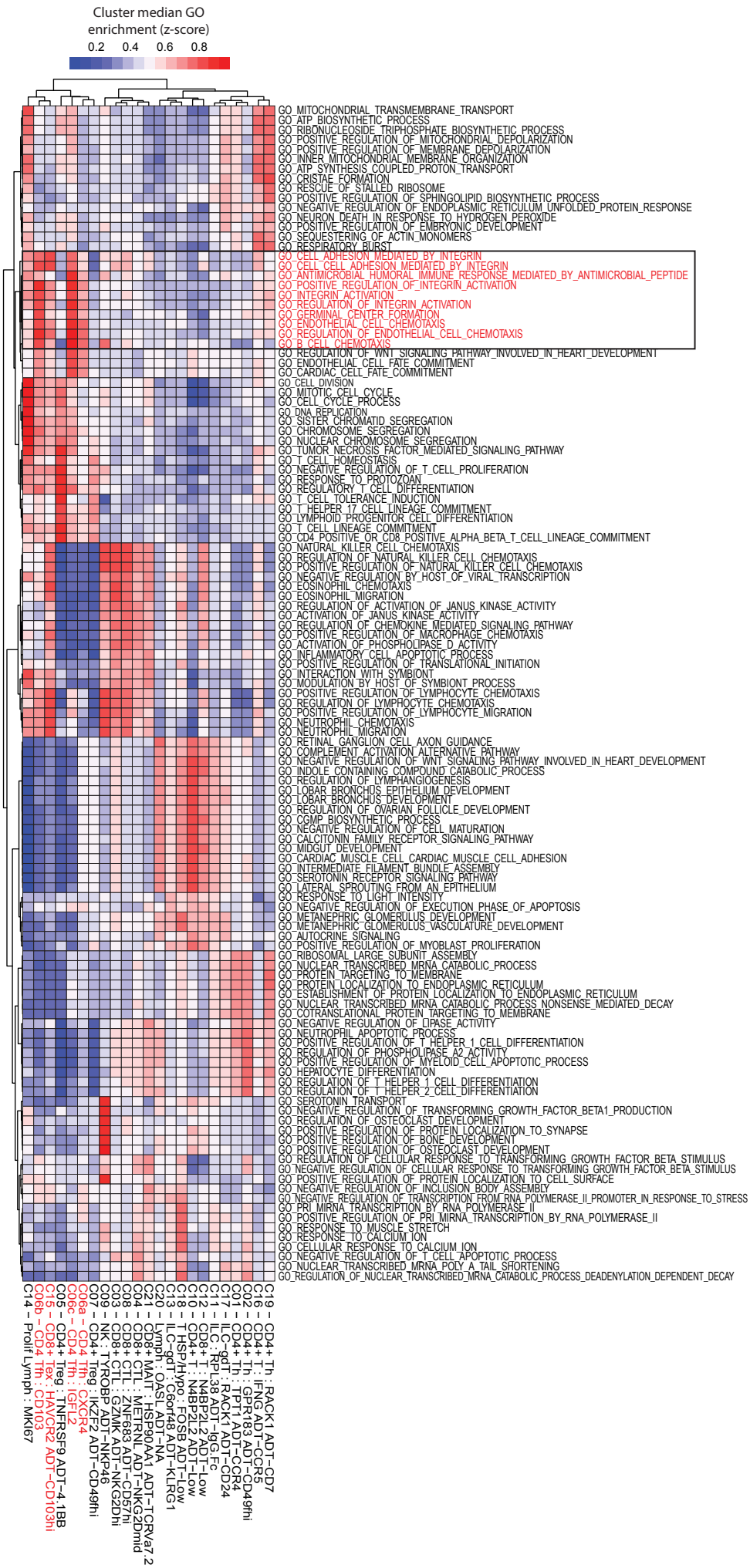
(E) Flow cytometric analysis of anti PD1, ICOS and CXCR5 antibody binding in various cell populations (as indicated) from undigested tonsil cells versus tumor cells subjected to the standard tumor dissociation protocol (methods).

(F) Undigested tonsil cells were used to define the bona fide Tfh-like gate. Two ICOS+PD1+ populations are identifiable, but (i) CXCR5+ cells fall in the PD1highICOSint gate and (ii) Tfh cells defined by either CXCR5/ICOS or CXCR5/PD1 also fall in the PD1highICOSint gate, this gate was henceforth used to identify Tfh-like cells in tumors where digestion excluded the use of CXCR5 as a marker.

(G) Flow cytometric identification of Tfh subsets based on markers identified in CITE-Seq. C06 a, b and c can be identified on the CD4+Tfh-like gate using CD103, CD49f, CD127 and HLA-DR. All CD4+CD103+ cells are Tfh like cells **(H)** Flow cytometric assessment of key markers of Tfh-like subclusters identified in CITE-Seq. The CD4+ gate for several tumor samples, tonsil and PBMC is shown, divided based on CD49f and CD103 expression

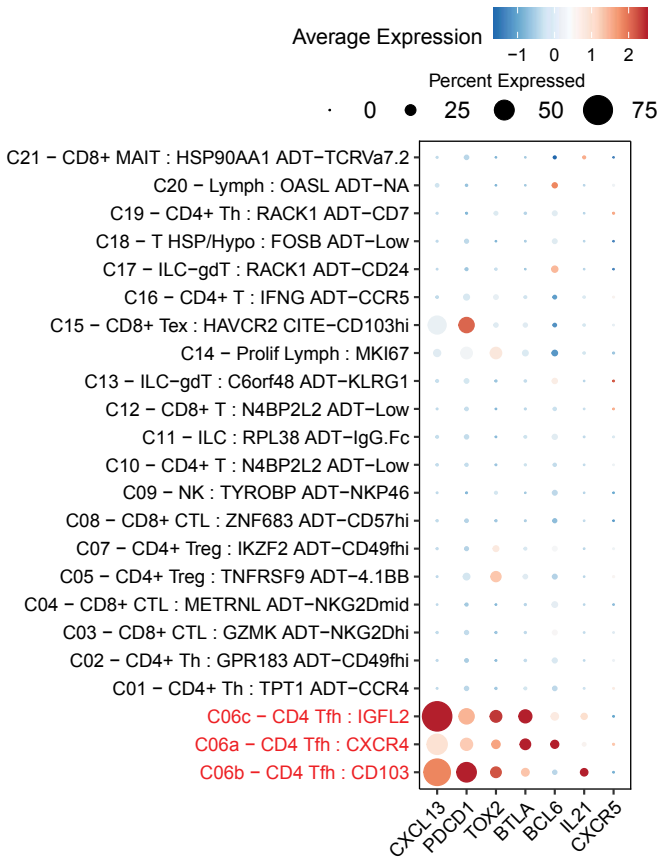
Extended Data Figure 3

A

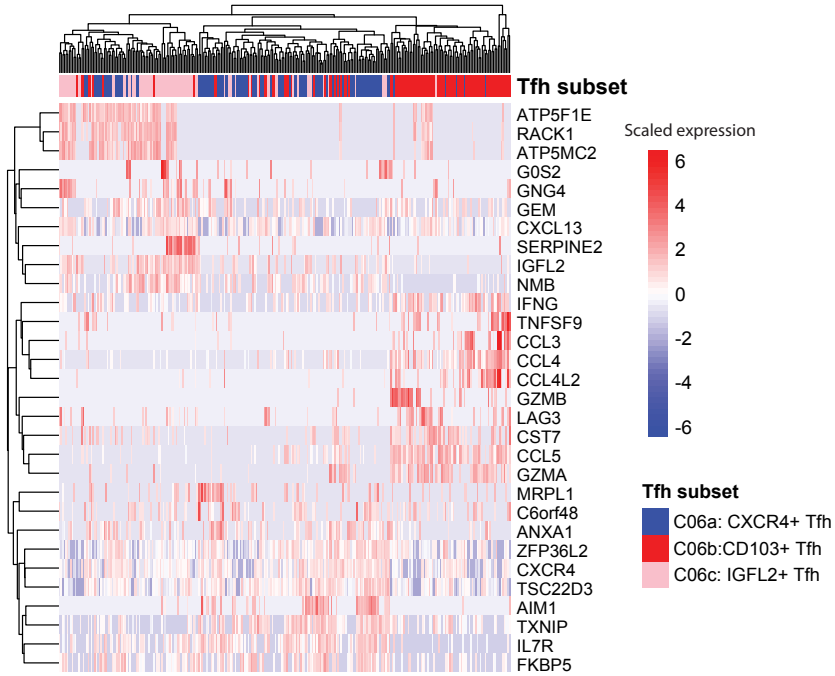


Extended Data Figure 3 (continued)

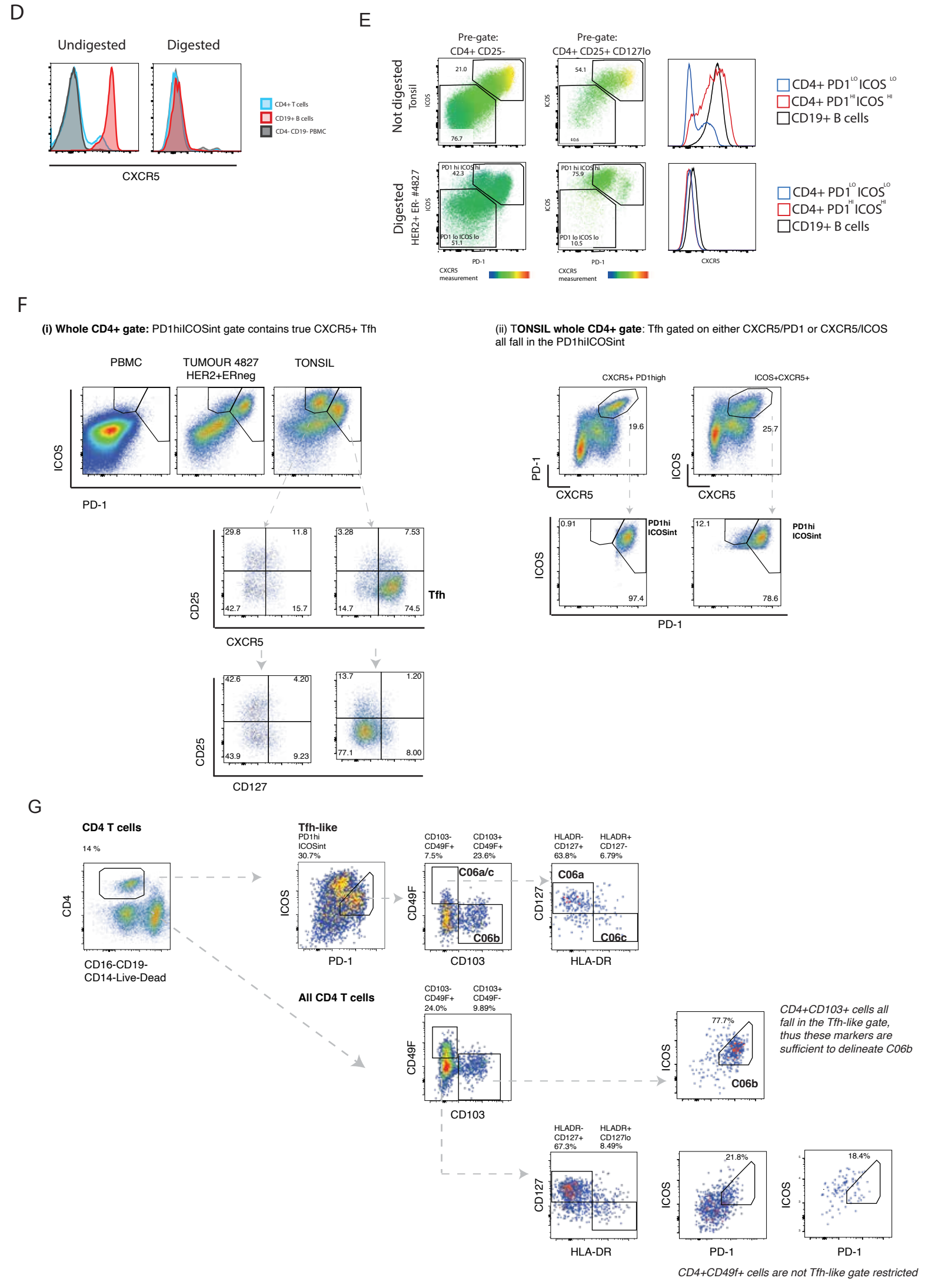
B



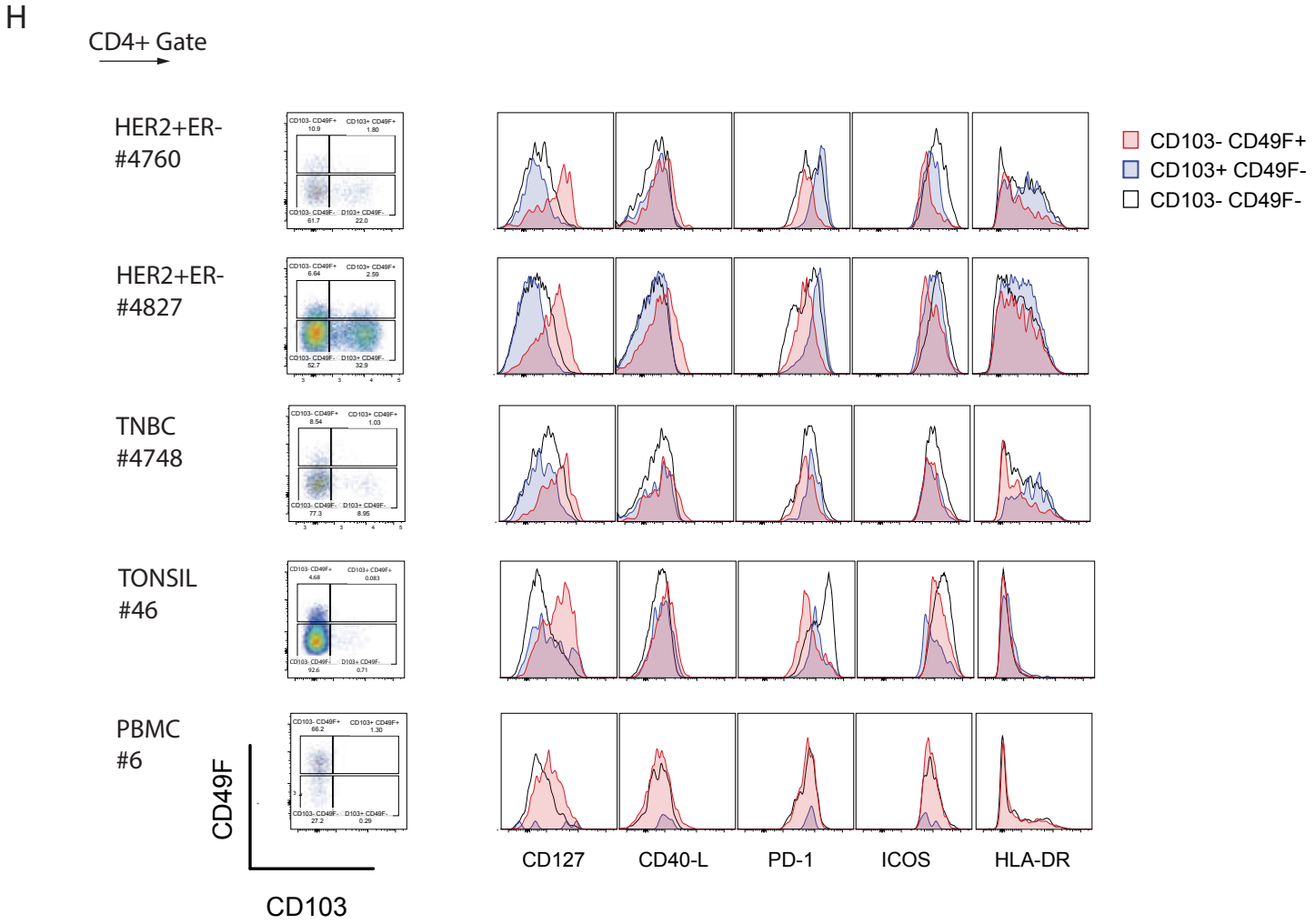
C



Extended Data Figure 3 (continued)

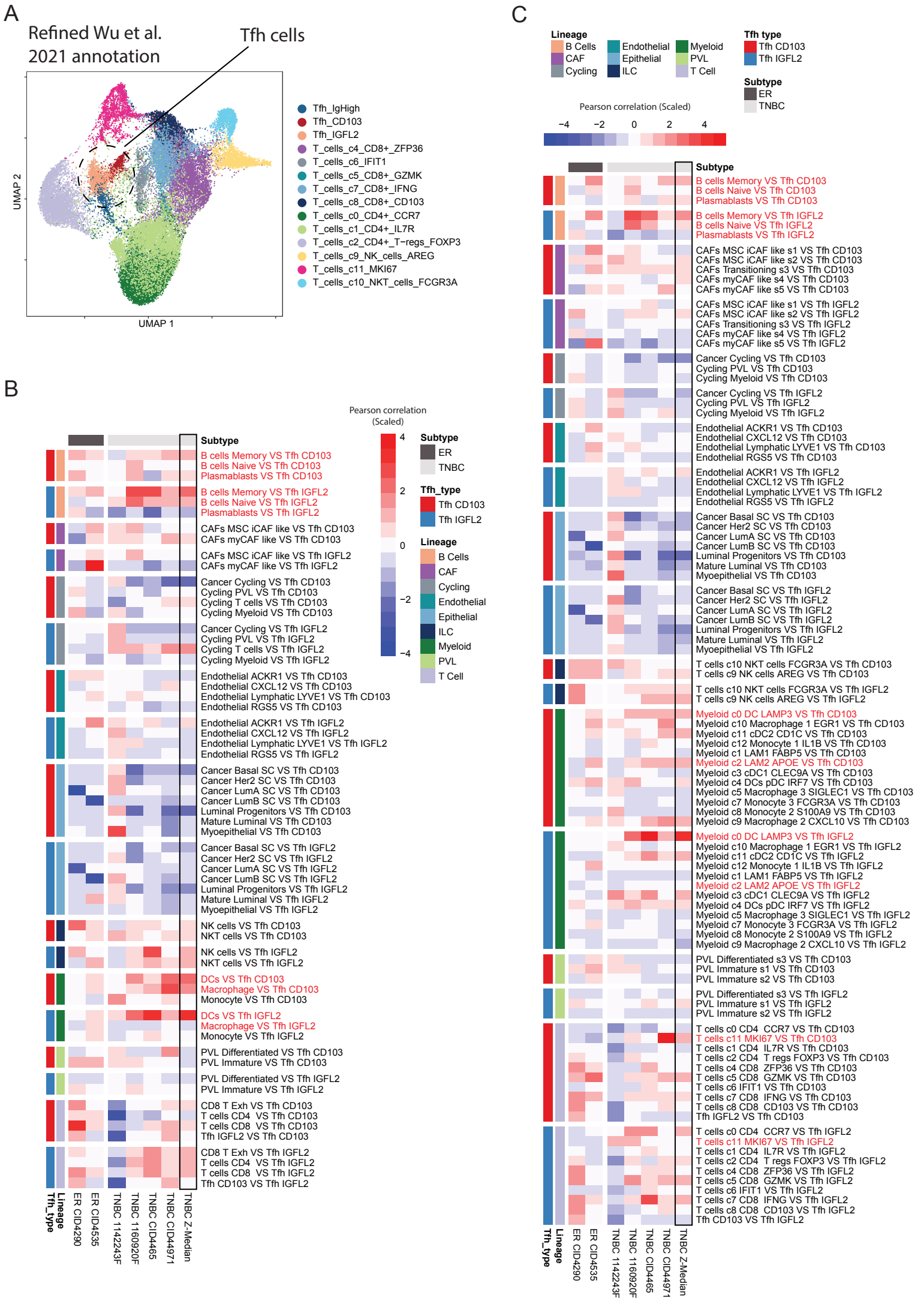


Extended Data Figure 3 (continued)



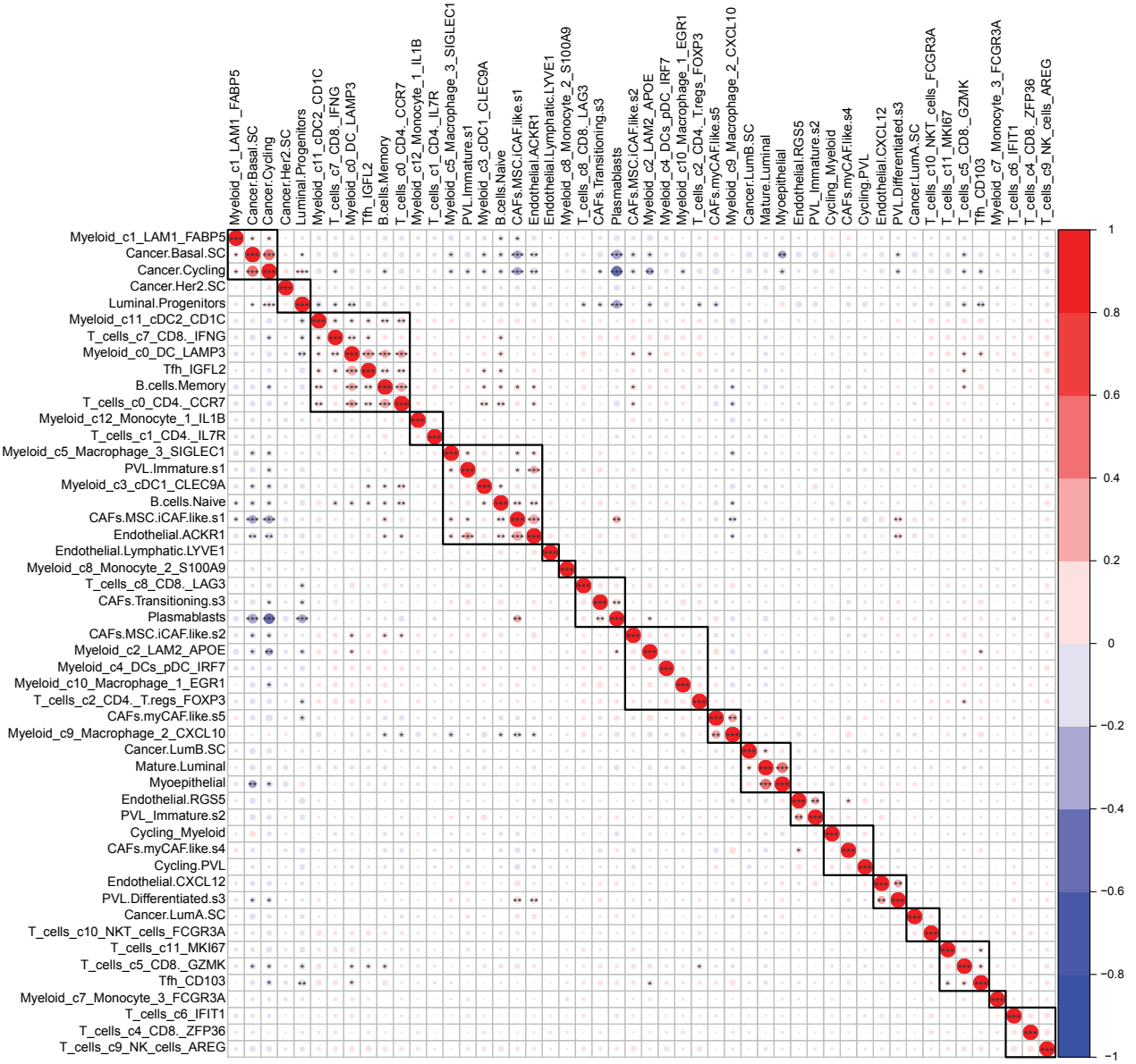
Extended Data Figure 4: Supplementary data for analysis of Tfh subset in Visium based spatial co-localisation and cell-cell signalling. (A) UMAP of T cells and ILCs from Wu et al.[7] with refined annotation of CD4 Tfh cells. **(B)** Pearson correlation heatmap of spatially deconvoluted cell pairs co-localised with CD4 Tfh subsets using Wu et al. data with refined annotations. **(C)** Pearson correlation heatmap of spatially deconvoluted cell pairs co-localised with CD4 Tfh subsets at a higher clustering resolution. **(D)** Dotplot of Pearson correlation values for spatially deconvoluted cells from (C). **(E)** Gene set enrichment analysis boxplots of GO biological process pathways
“GO_MACROPHAGE_ACTIVATION_INVOLVED_IN_IMMUNE_RESPONSE”and
“GO_CHRONIC_INFLAMMATORY_RESPONSE” across T cells and ILCs, sorted from high to low (left to right). Red line indicates the median expression. Box plots are coloured by lymphocyte lineage. Red line indicates the median expression for each marker across clusters. Boxplots middle line marks the median value, the lower and upper hinges mark the 25% and 75% quantile, the whiskers correspond to the 1.5 times the interquartile range, the black dot’s mark outliers. **(F)** Relative information flow contribution of signalling pathways (the sum of total R-L communication probability found with each signalling pathway) to be differentially enriched in CD4 Tfh cell subsets. Bar plots visualise the weighted variance in contribution when comparing CD103+ Tfh to IGFL2+ Tfh when scaled 0-1. Pathways found to be significant ($p < 0.05$) are coloured either blue if enriched in IGFL2 Tfh cells, or red if enriched in CD103 Tfh. **(G)** CCL, CSF and BTLA signalling pathway network information flow across all clusters. ‘Sender’ and ‘Receiver’ reflects direct expression of ligands and receptors (agonistic or antagonistic), ‘Mediator’ and ‘Influencer’ quantifies clusters' potential role in controlling receptor-ligand expression flow of the pathway within the system (TME). **(H)** Dotplot visualising receptor-ligand signalling information differentially increased or decreased in expression by CD103 Tfh and IGFL2 Tfh clusters. Red text highlights signalling pathways of interest.

Extended Data Figure 4

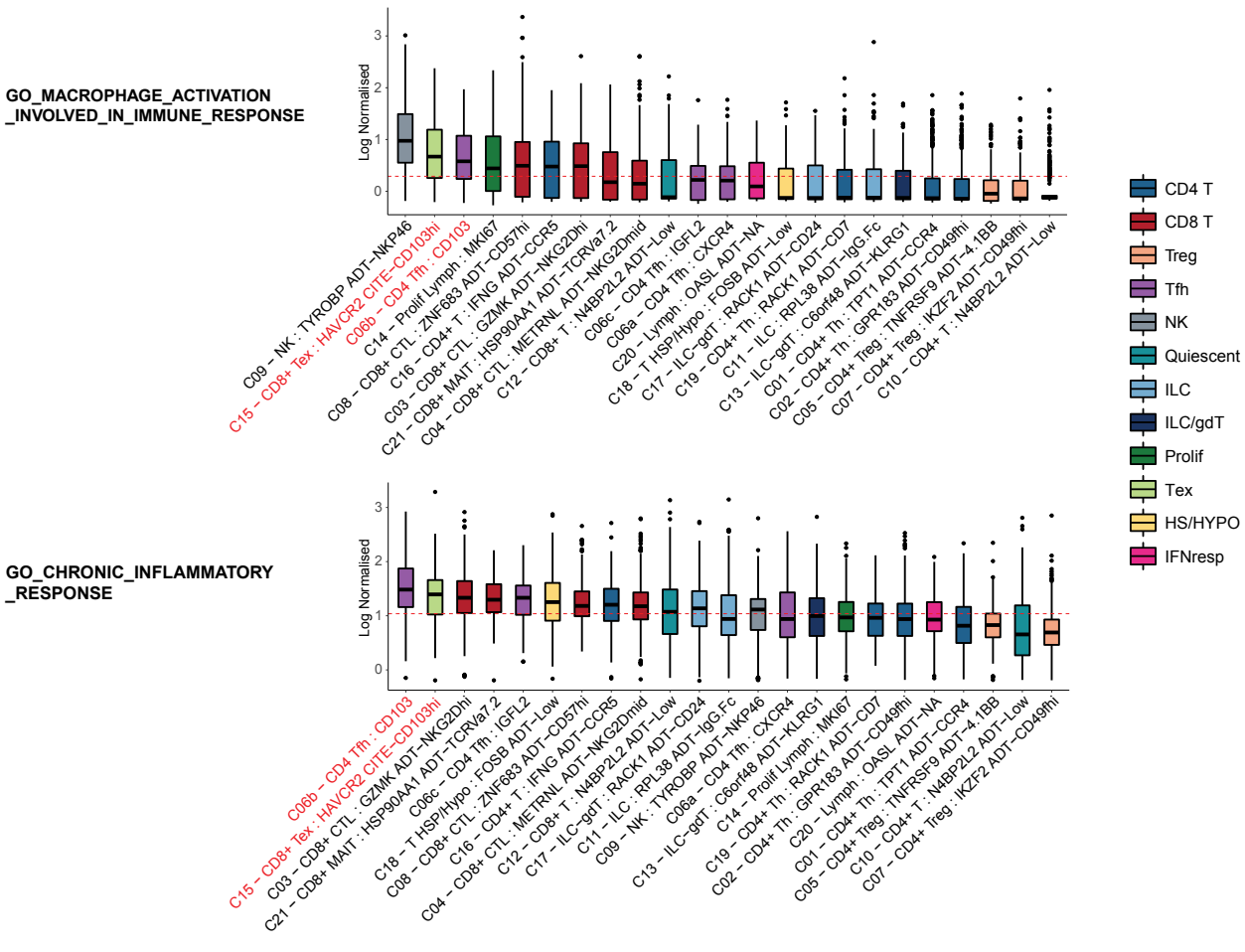


Extended Data Figure 4 (continued)

D

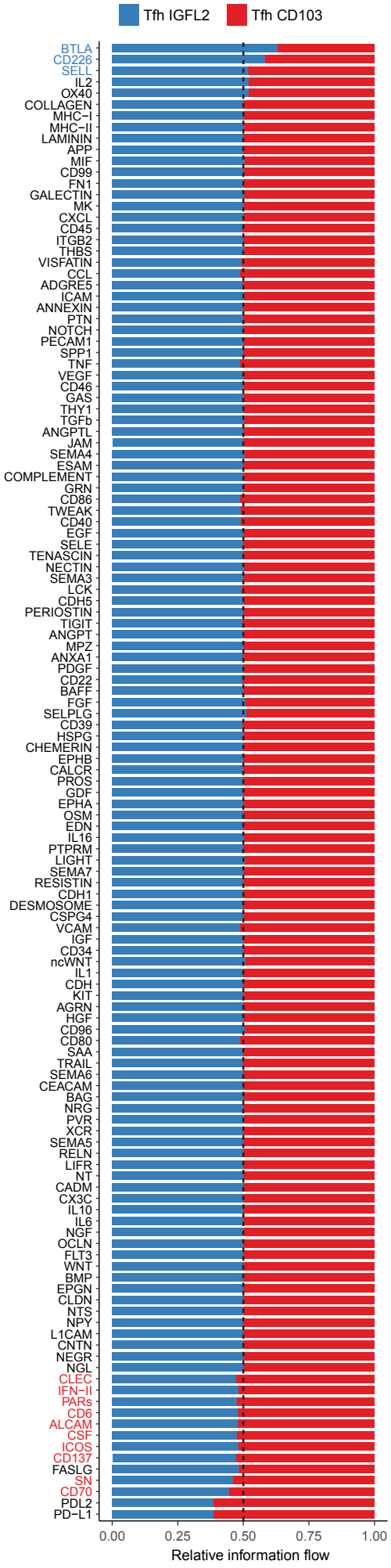


E

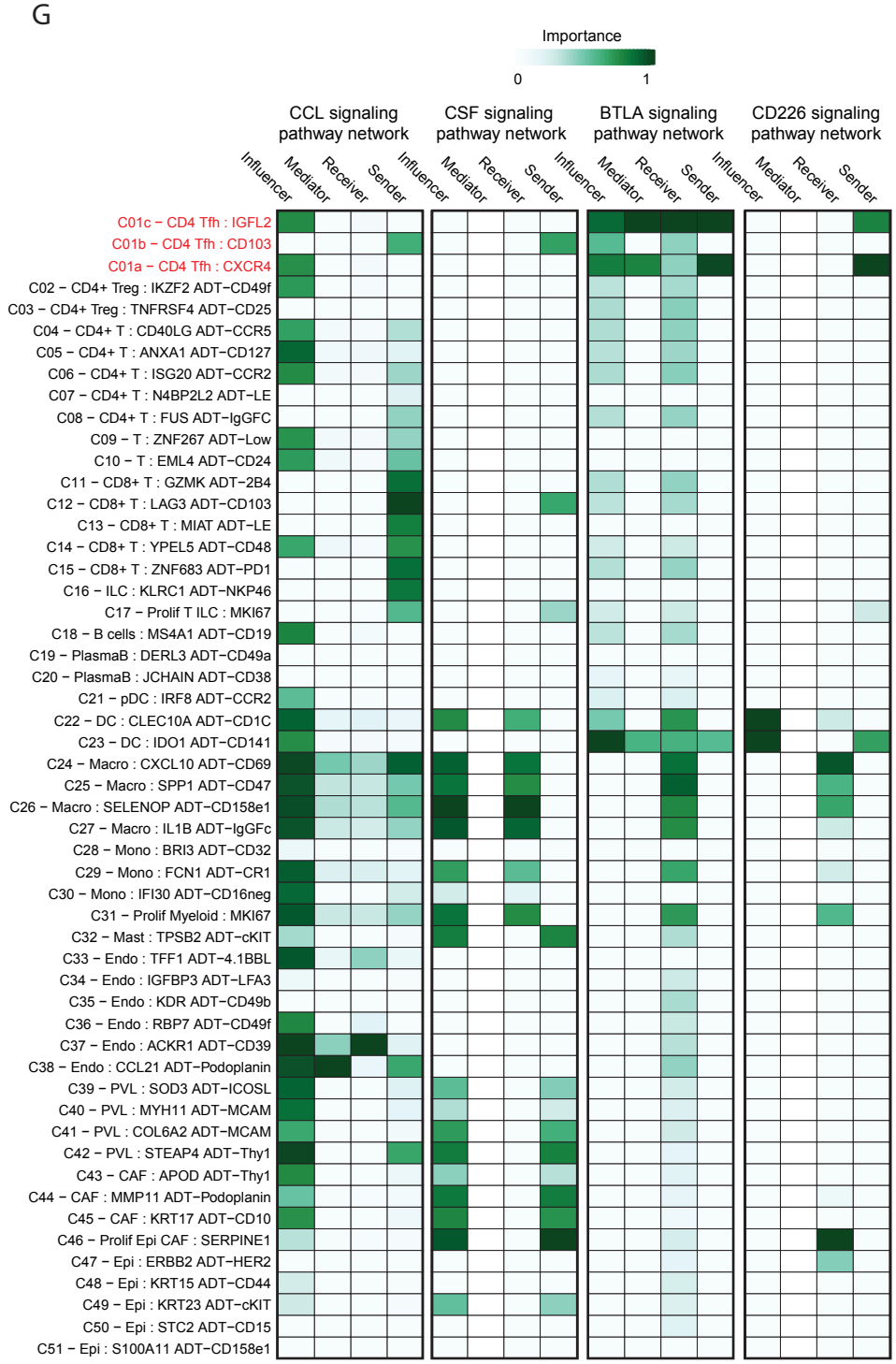


Extended Data Figure 4 (continued)

F

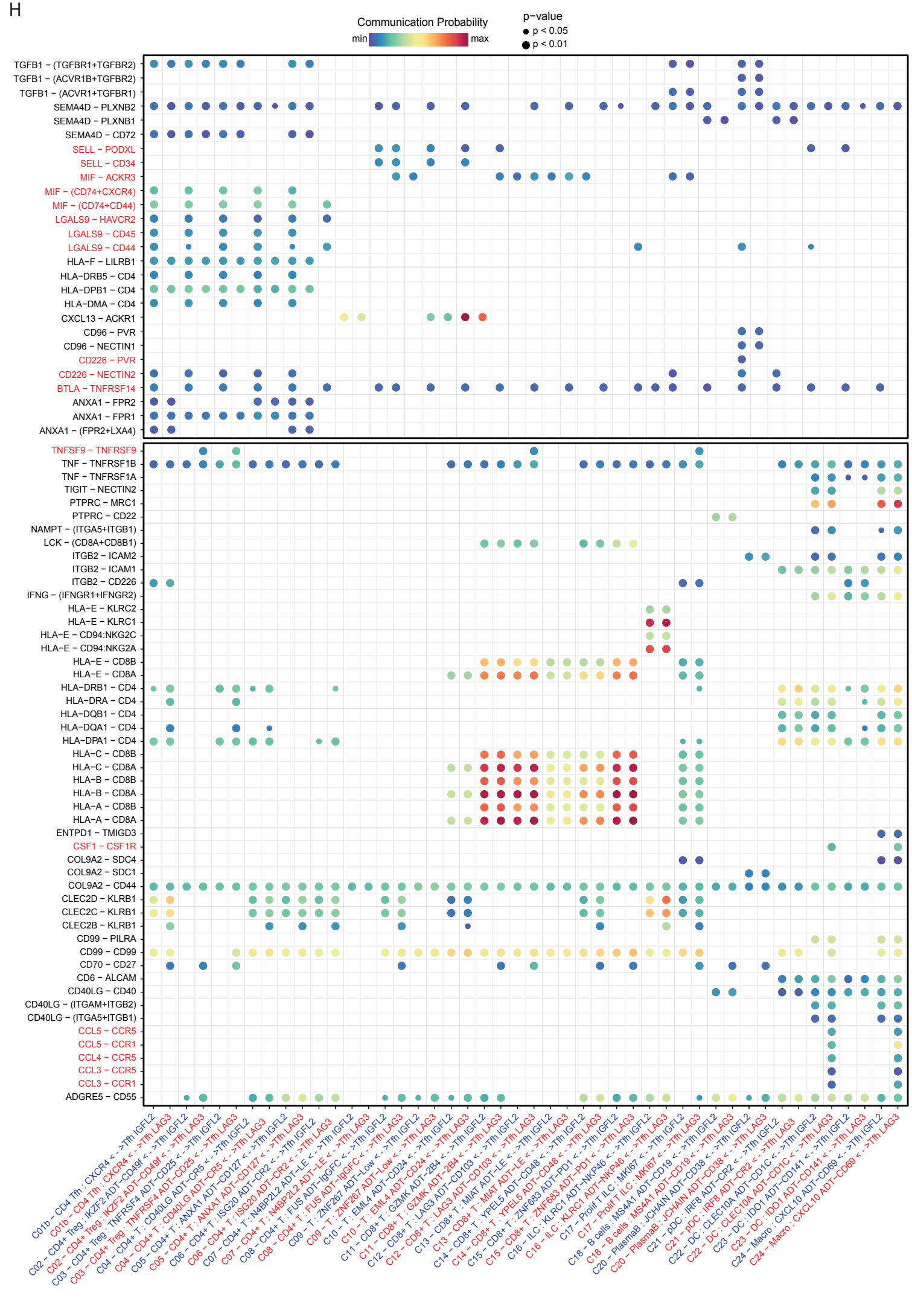


G



Extended Data Figure 4 (continued)

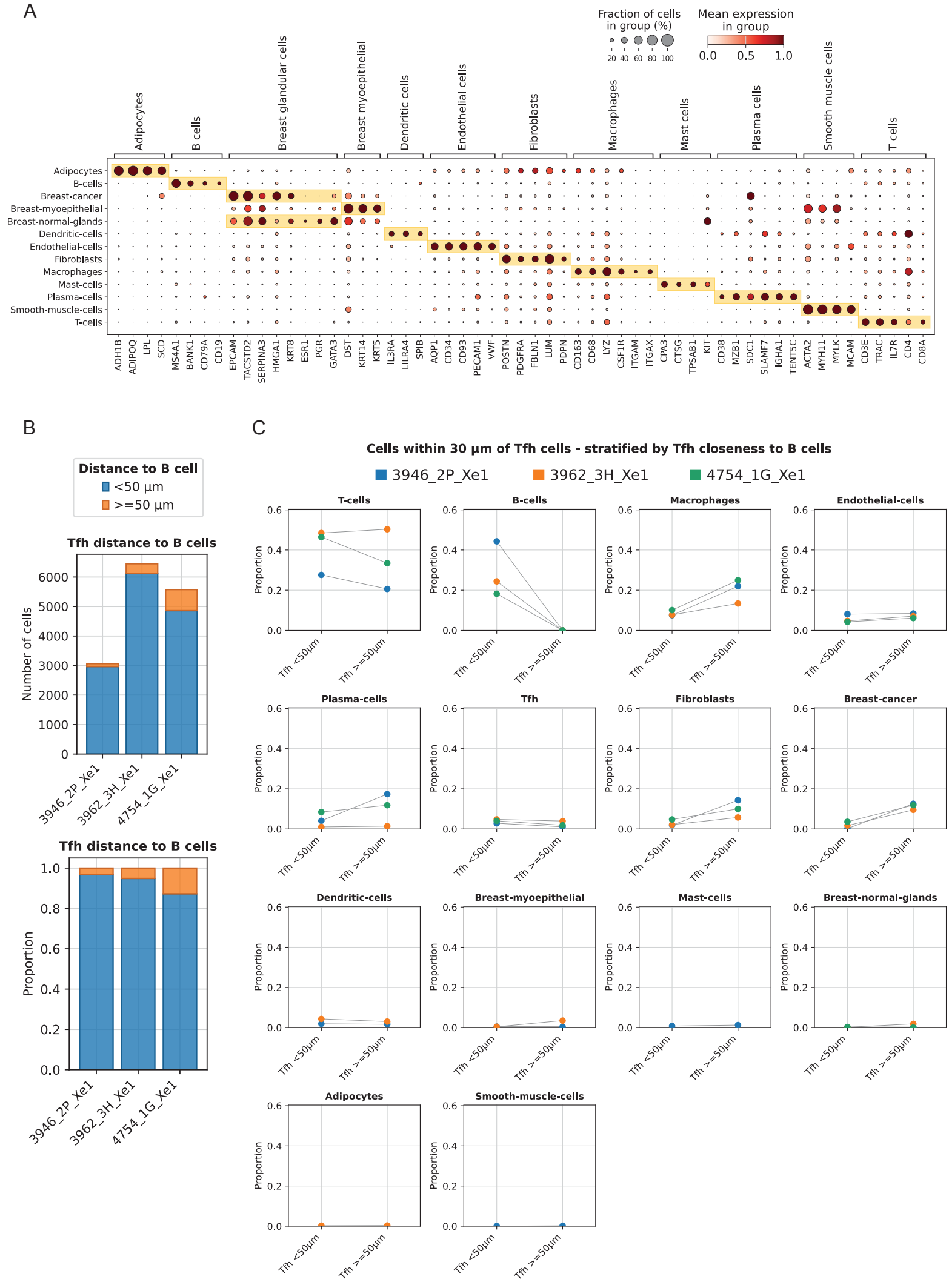
H



Extended Figure 5. Supplementary data for analysis of cell localisation in Xenium based spatial transcriptomics

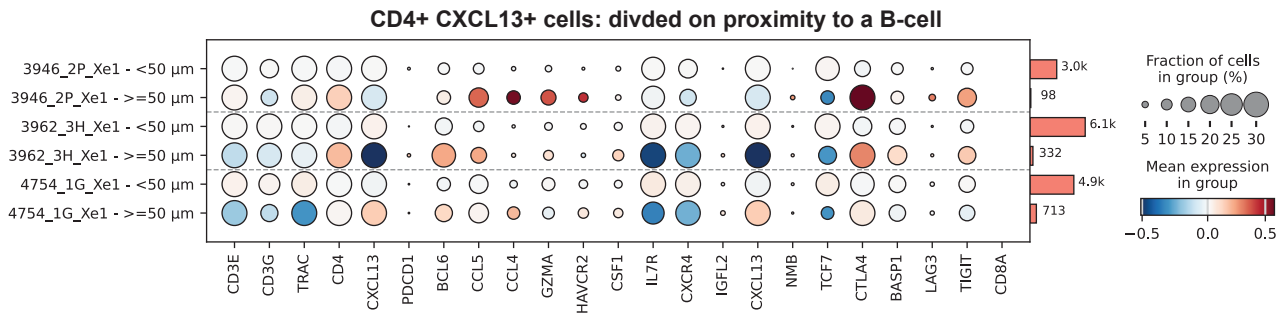
Xenium spatial transcriptomic analysis with the Human Breast 280 plus 100 custom gene panel (Supplementary Table S6) was applied to three TNBC samples **(A)** Clusters were assigned cell type labels based on marker genes defined in the Human Breast panel. Epithelial clusters were manually refined using marker expression and morphology to distinguish normal from cancer. Clusters were assigned cell type labels based on marker genes defined in the Human Breast panel. Epithelial clusters were manually refined using marker expression and morphology to distinguish normal from cancer. Dot plots showing expression of selected selected genes that define that define major cell types in Xenium analysis of TNBC samples. **(B)** Absolute cell number and proportion of B cell proximal and distant (+/- 50uM) in each TNBC sample **(C)** Each panel displays a major cell population and indicates the proportion of the cells within 30uM of *B cell proximal Tfh* (<50uM) versus *B cell distant Tfh* (>50uM). **(D-G)** Dots plots show gene expression divided by individual samples. The title of each panel indicates the cell type being assessed and labels on the left indicate patient ID and whether the cell populations being examined are close or distant. Fraction of cells (circle size) and relative expression (colour) are indicated

Extended Data Figure 5

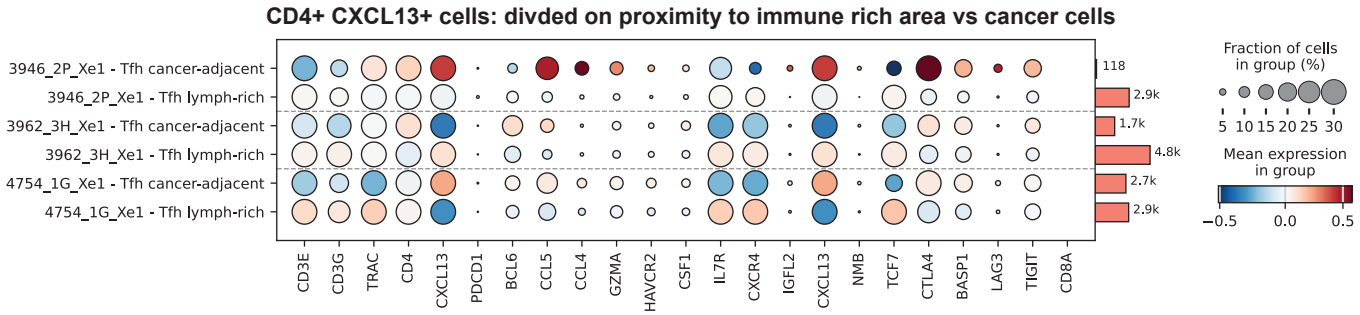


Extended Data Figure 5 (continued)

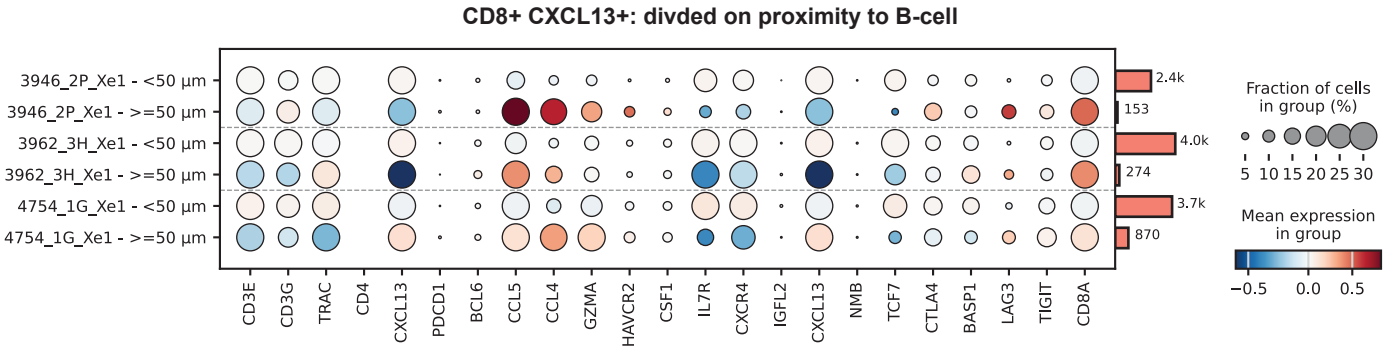
D



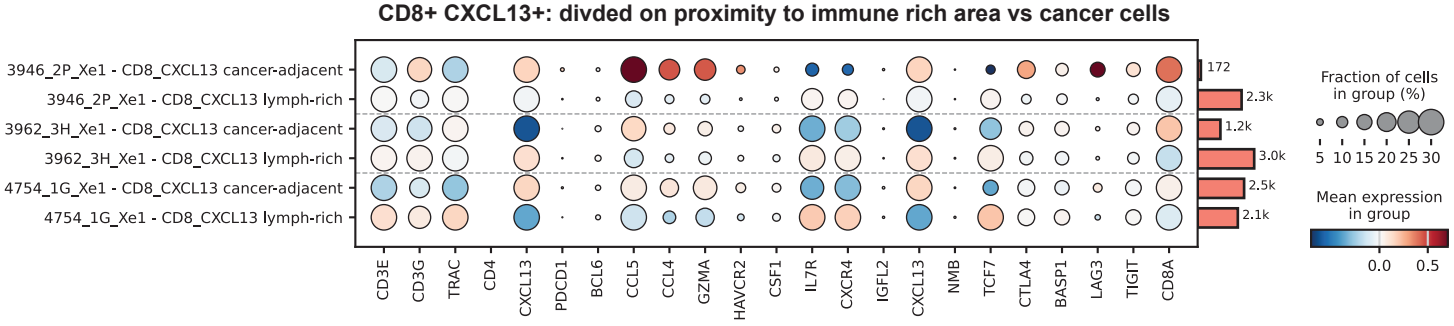
E



F



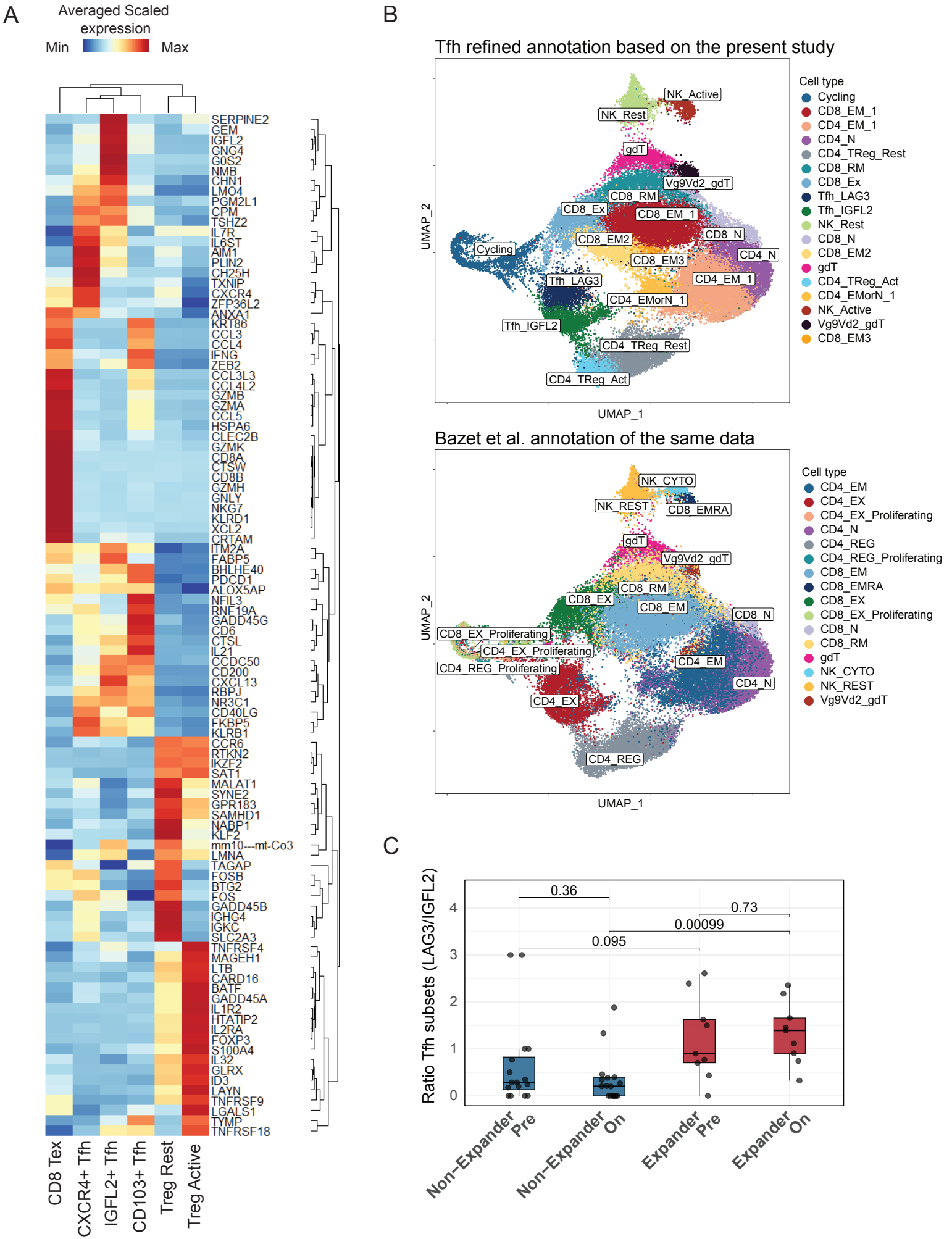
G



Extended Data Figure 6.

(A) Heatmap of top differentiating RNA features from each identified Tfh state in our dataset along with CD8+ Exhausted T cells and Tregs (MAST test; $p < 0.05$). **(B)** UMAP of Bassez et al.[4] breast cancer anti-PD-1 cohort T cell dataset. The top plot shows the Tfh redefined annotation used for all downstream analysis in our study. The bottom plot is a UMAP using the Bassez et al. original annotations [4] . **(C)** Comparative analysis of the Tfh_LAG3 (C06b CD103)/Tfh_IGFL2 (C06c) ratio across timepoints and response groups, baseline (Pre) to on-treatment (On). Box plots indicate median and interquartile range (IQR); overlaid points represent individual patients. P-values indicate significance of pairwise comparisons (Wilcoxon test). **(D)** Slope charts visualizing the mean change in relative abundance for each T-cell subset from baseline (Pre) to on-treatment (On). Lines connect the group mean frequencies; color denotes clinical response (Red = Expander; Blue = Non-Expander). **(E)** Paired volcano plots displaying the fold change of mean frequency (On vs. Pre) for T/NK-cell subsets within the expander (left) and non-expander (right) cohorts. The x-axis denotes the log2 fold change of the population mean; the y-axis denotes significance (-log10 P-value). **(F)** Volcano-style scatter plot identifying cell types with significantly divergent proportion change rates between expansion groups. The X-axis displays the "Difference in Expansion; calculated as the mean patient-matched log2 fold change in Expanders minus the mean in Non-Expanders. A positive value indicates the cell type expanded more in responders than in non-responders. The Y-axis shows statistical significance (-log10 P-value, Wilcoxon rank-sum test). Point size represents the magnitude of this effect (the absolute difference), with larger points indicating a greater disparity in growth behavior between the expansion groups regardless of direction. **(G)** Barplot represents the total count of differentially expressed genes (DEGs) identified by comparing On-treatment versus Pre-treatment samples for each T-cell cluster. Barplots are ordered by the magnitude of the Expander response (high to low). **(H)** T-cell subsets ranked by the total number of significant DEGs in Expanders divided by those in Non-Expanders. Red bars (Ratio greater than 1) identify T-cell subsets with increased DEGs in patients classified as expanders, while blue bars (Ratio less than 1) indicate subsets where transcriptional changes are dominated by the non-expansion cohort.

Extended Data Figure 6



Extended Data Figure 6 (Continued)