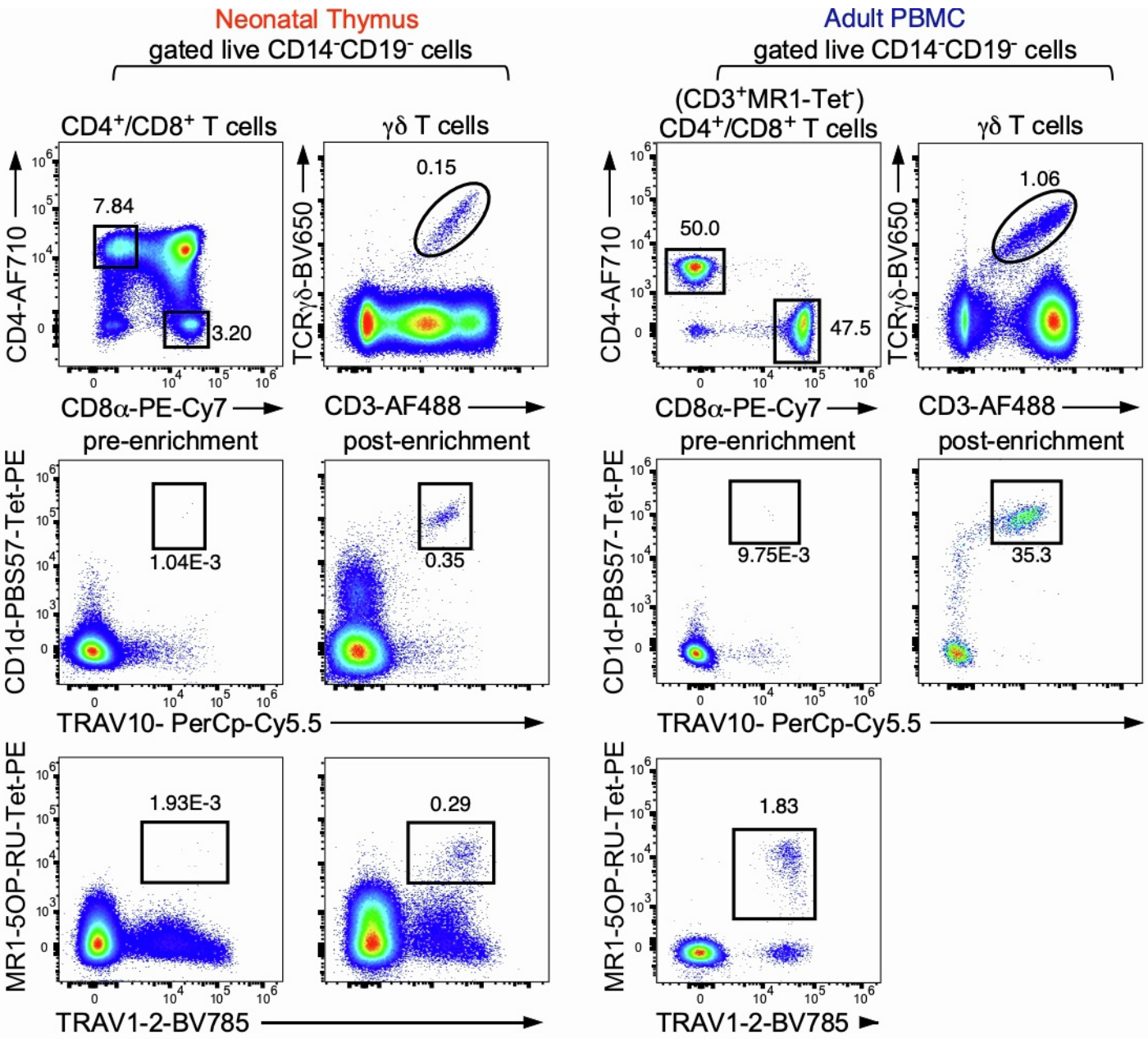
Cell Reports, Volume *43*

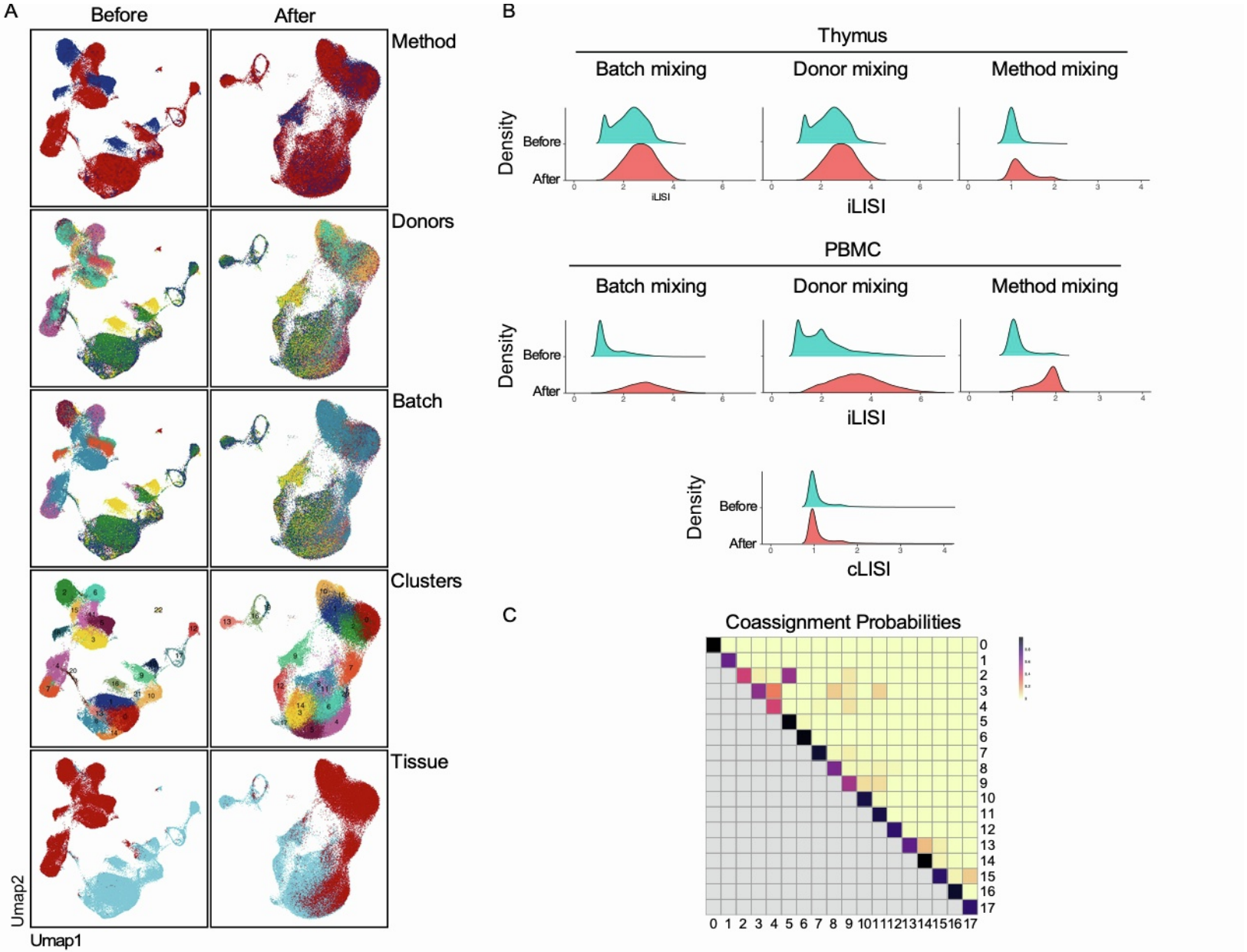
Supplemental information

Unraveling the phenotypic states of human innate-like T cells: Comparative insights with conventional T cells and mouse models

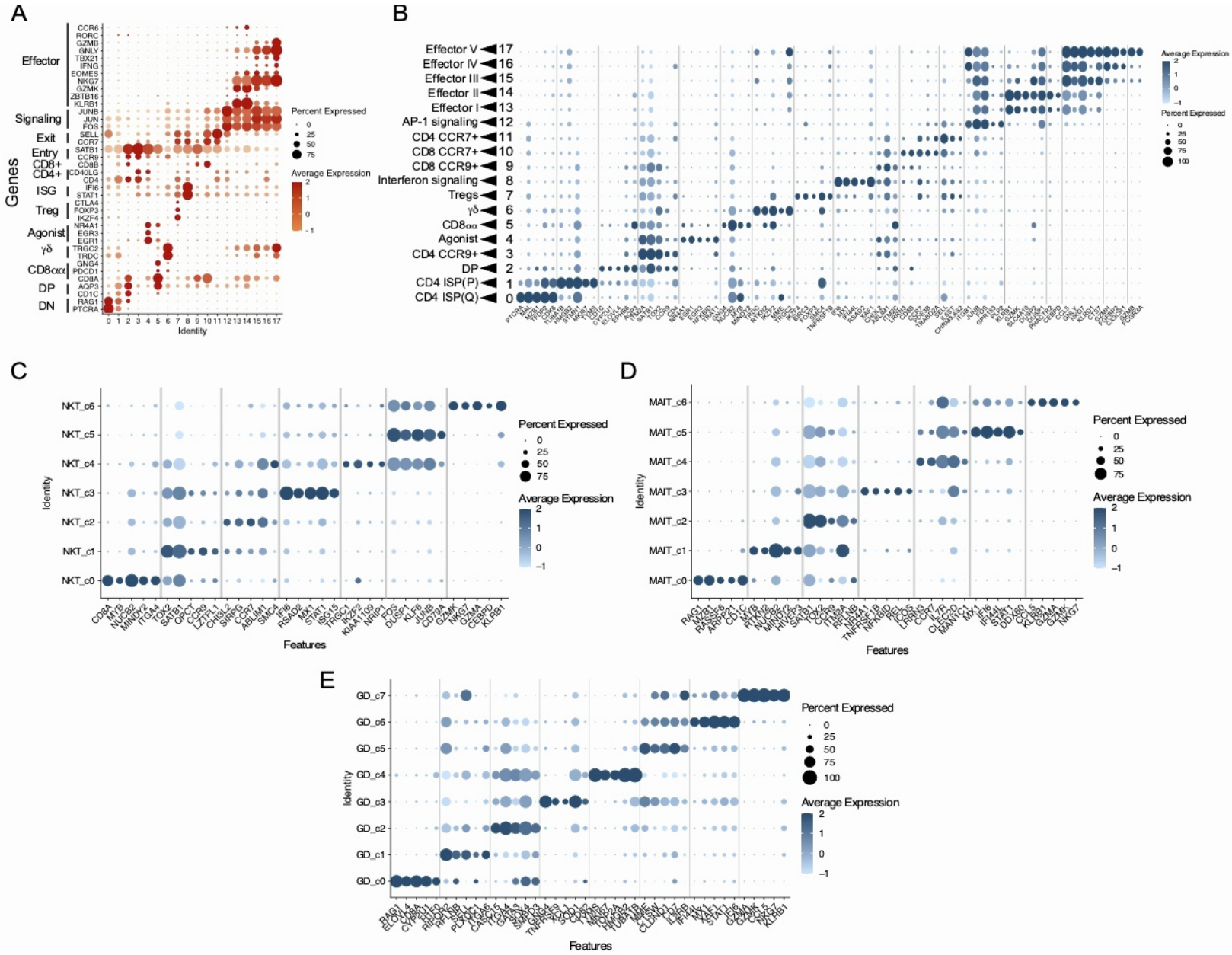
Liyen Loh, Salomé Carcy, Harsha S. Krovi, Joanne Domenico, Andrea Spengler, Yong Lin, Joshua Torres, Rishvanth K. Prabakar, William Palmer, Paul J. Norman, Matthew Stone, Tonya Brunetti, Hannah V. Meyer, and Laurent Gapin



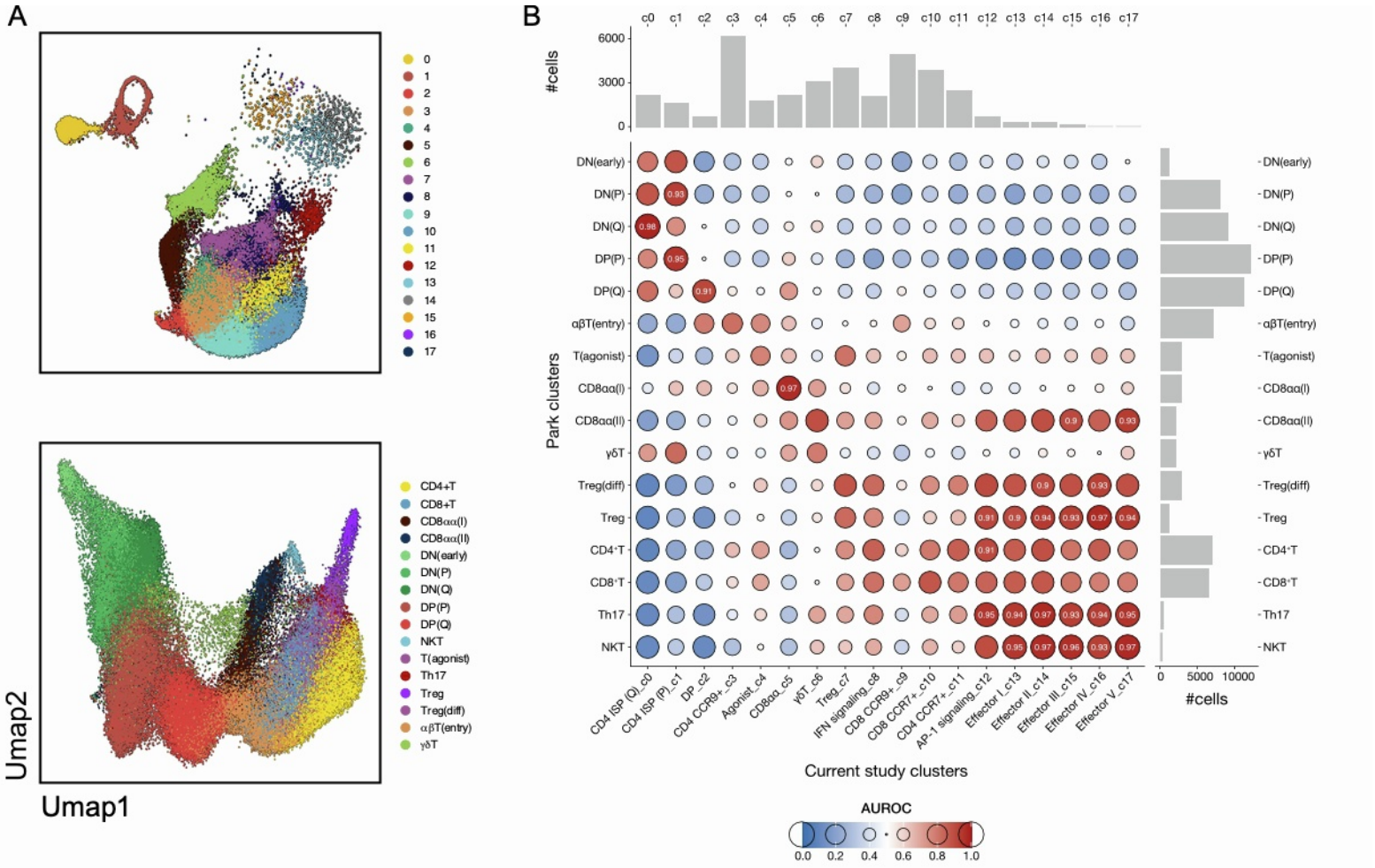
Supplementary Figure 1: Cell sorting strategy for single-cell sequencing. Non-myeloid ($CD14^-$), non-B-cell ($CD19^-$), live cells (viability dye efluor780) from both thymus and blood were sorted into CD4, CD8 and γδ T cells based on $CD4^+CD8^-$, $CD8^+CD4^-$ and $CD3^+TCR\gamma\delta^+$ marker expression, respectively. iNKT and MAIT cells were pre-enriched via CD1d-PBS57 and MR1-5OPRU magnetic beads and sorted based on binding to tetramer and TRAV10 or TRAV1-2 antibodies respectively.



Supplementary Figure 2: Batch integration and quality control. A. UMAP projection before and after integration with Harmony, colored by method (RNAseq, RNAseq+VDJseq), donor (1-13), batch (A-I), clusters (1-17) and tissue (thymus and blood). B. Degree of mixing during batch correction and dataset integration measured as the local inverse Simpson's index (LISI). The top and middle panels show the integration LISI (iLISI), which measures the effective number of datasets within a neighborhood, for thymic and PBMC-derived cells, respectively. Mixing was assessed on batch, donor and method used (as in depicted in A); the lower panel depicts the cell-type LISI (cLISI), to evaluate the accuracy of cell-type assignment. Blue curves indicate LISI before integration, red after integration. C. Cell co-assignment probabilities within (diagonal) and across clusters (off diagonal) assessed by cell bootstrapping and re-cluster. High co-assignment probabilities indicate cluster stability.

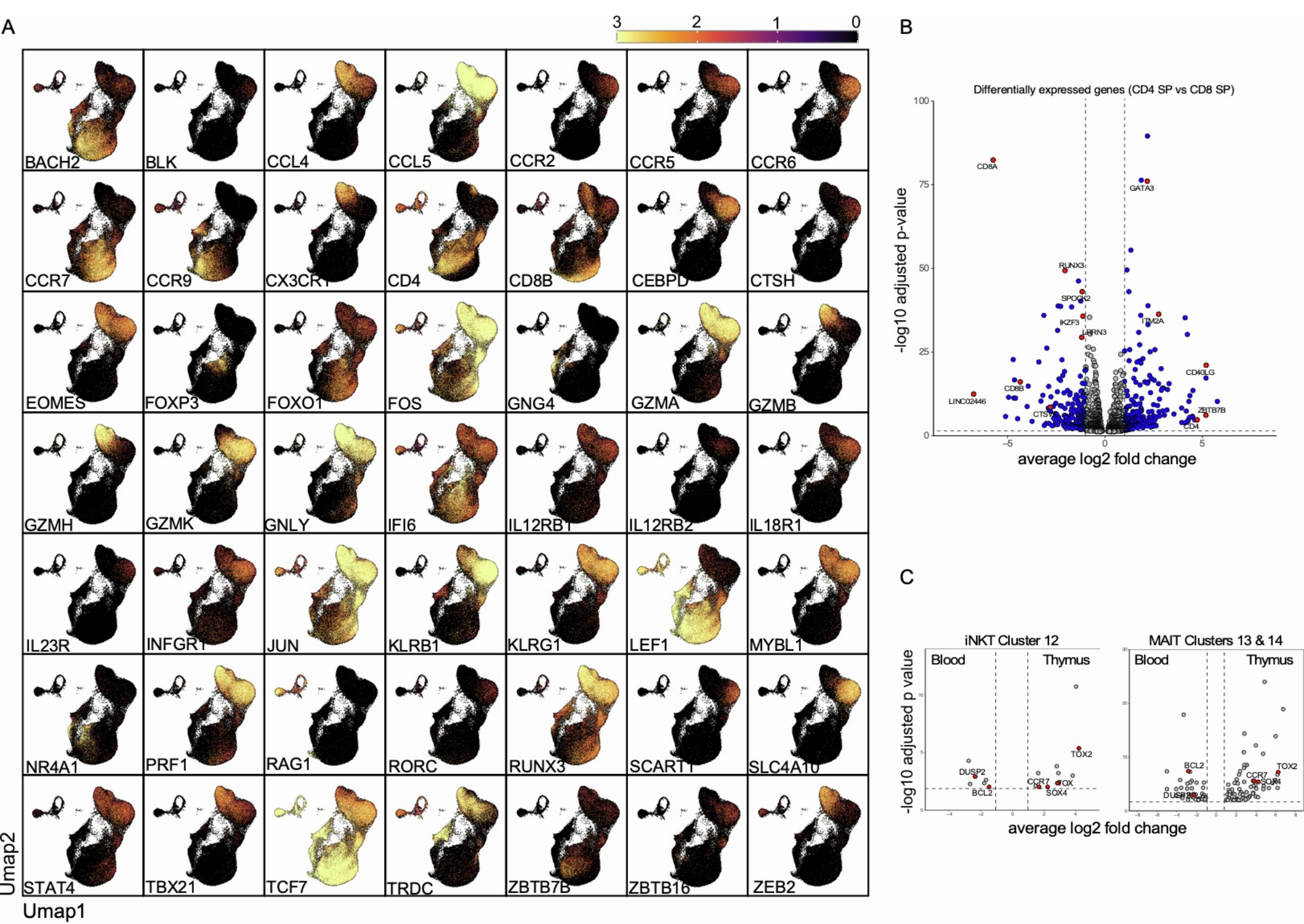


Supplementary Figure 3. Marker gene expression across cell clusters. A. Reference signature gene expression for clusters in Fig 1C and B. the top five genes that characterize these clusters in this dataset. Top five marker genes for C. iNKT, D. MAIT and E. γδ T cells corresponding to the clusters in Fig 2D.

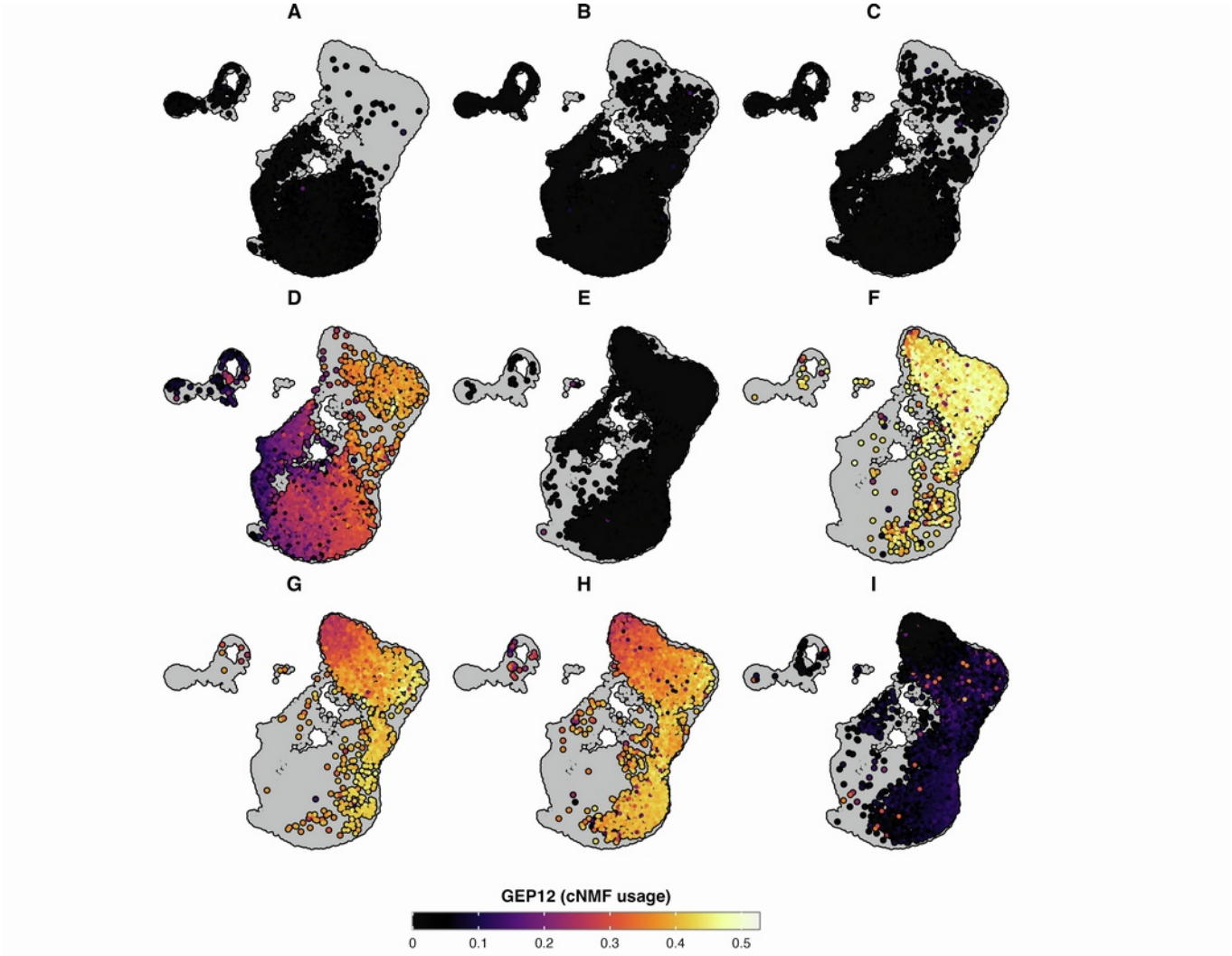


Supplementary Figure 4. Reproducibility of thymocyte data with human thymus atlas. A. UMAP representation of our integrated thymocyte data (top) and the Park *et al.* thymocyte data (bottom). Cells are colored by cluster.

B. Bubbleplot showing the MetaNeighbor AUROC score for pairwise similarities of our thymocyte clusters with the Park *et al.*[1] annotated thymocyte clusters. AUROC scores above 0.9 are written in white text.
Marginal bar plots represent the number of cells present in each cluster.

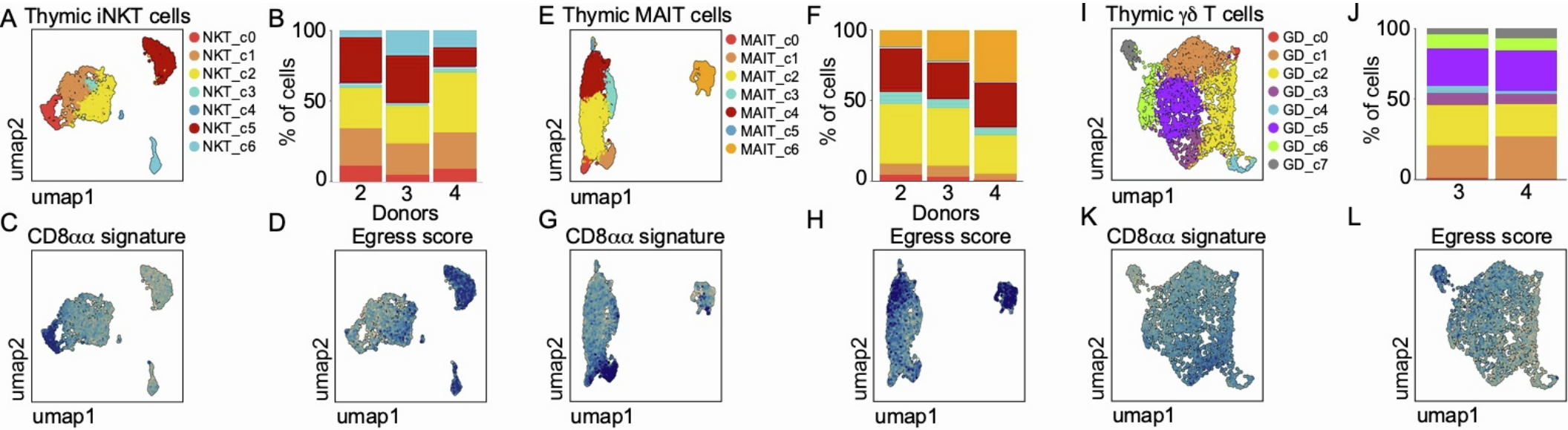


Supplementary Figure 5: Characteristics of gene expression on integrated dataset. A. Gene expression projection of signature genes.
B. Genes differentially expressed between thymic CD4 and CD8 SP T cells corresponding to clusters c3/c11 and c9/c10 in Fig 1C, respectively.
C. Genes differentially expressed between thymic and peripheral blood iNKT cells (left) and MAIT cells (right) in indicated clusters.

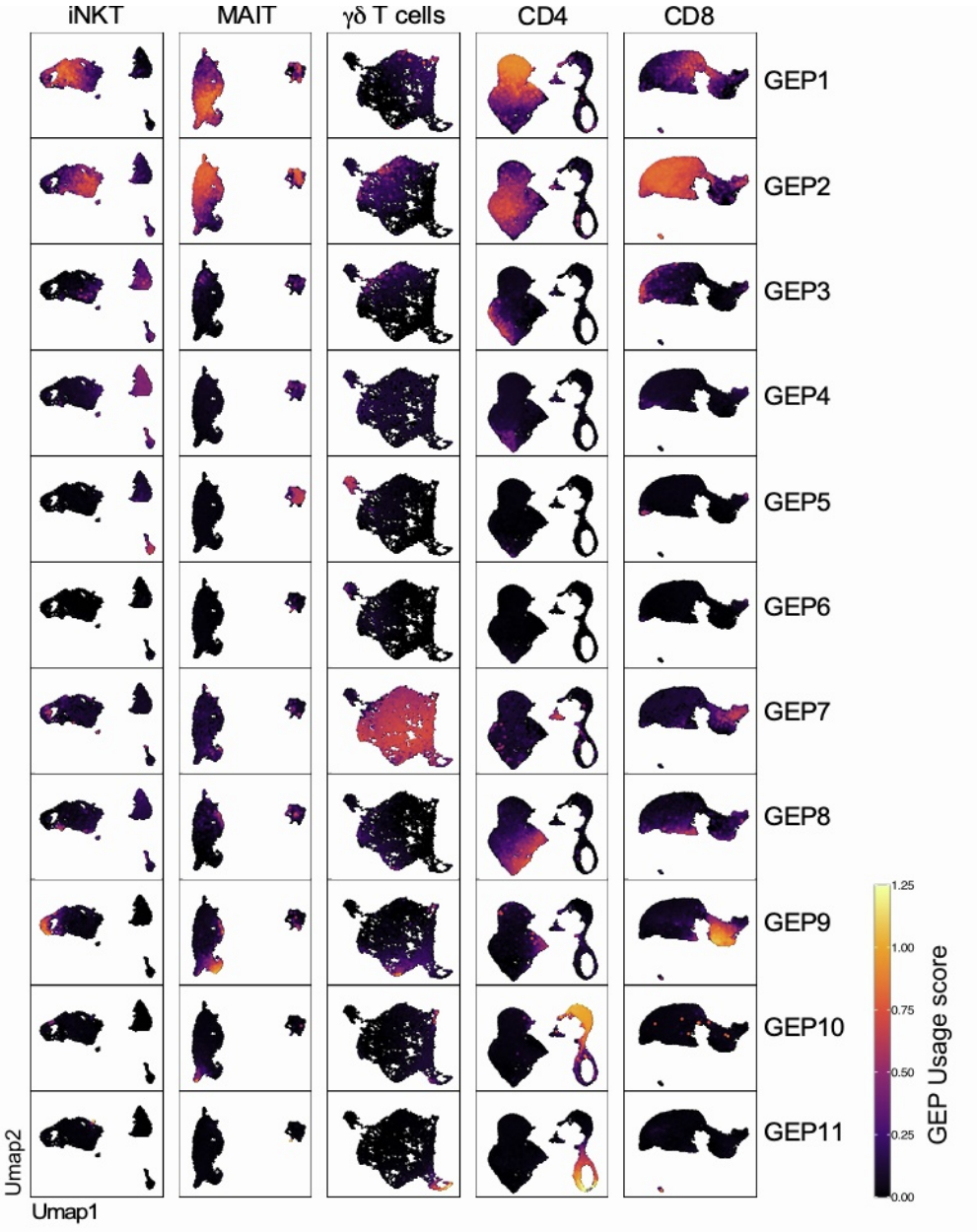


Supplementary Figure 6: Projection of GEP12 onto integrated T_{inn} and T_{conv} object.

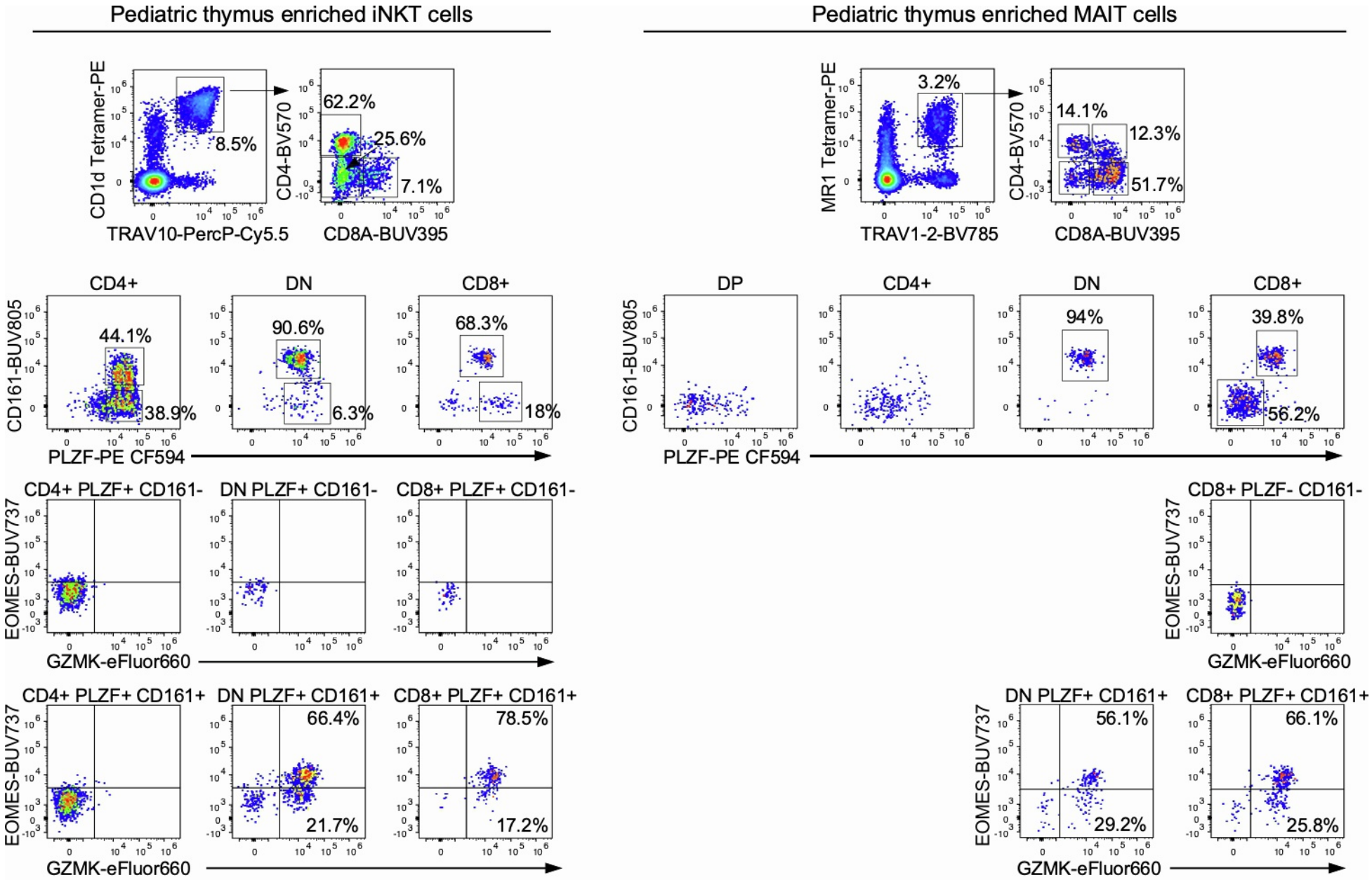
Each panel shows cells from a given sequencing batch, color-coded by the cNMF usage of GEP12. There is a clear separation of batches A-C, E, I and D, F-H, which align with the sequencing method used, RNAseq only or RNAseq+VDJseq, respectively (see Sup Table 1).



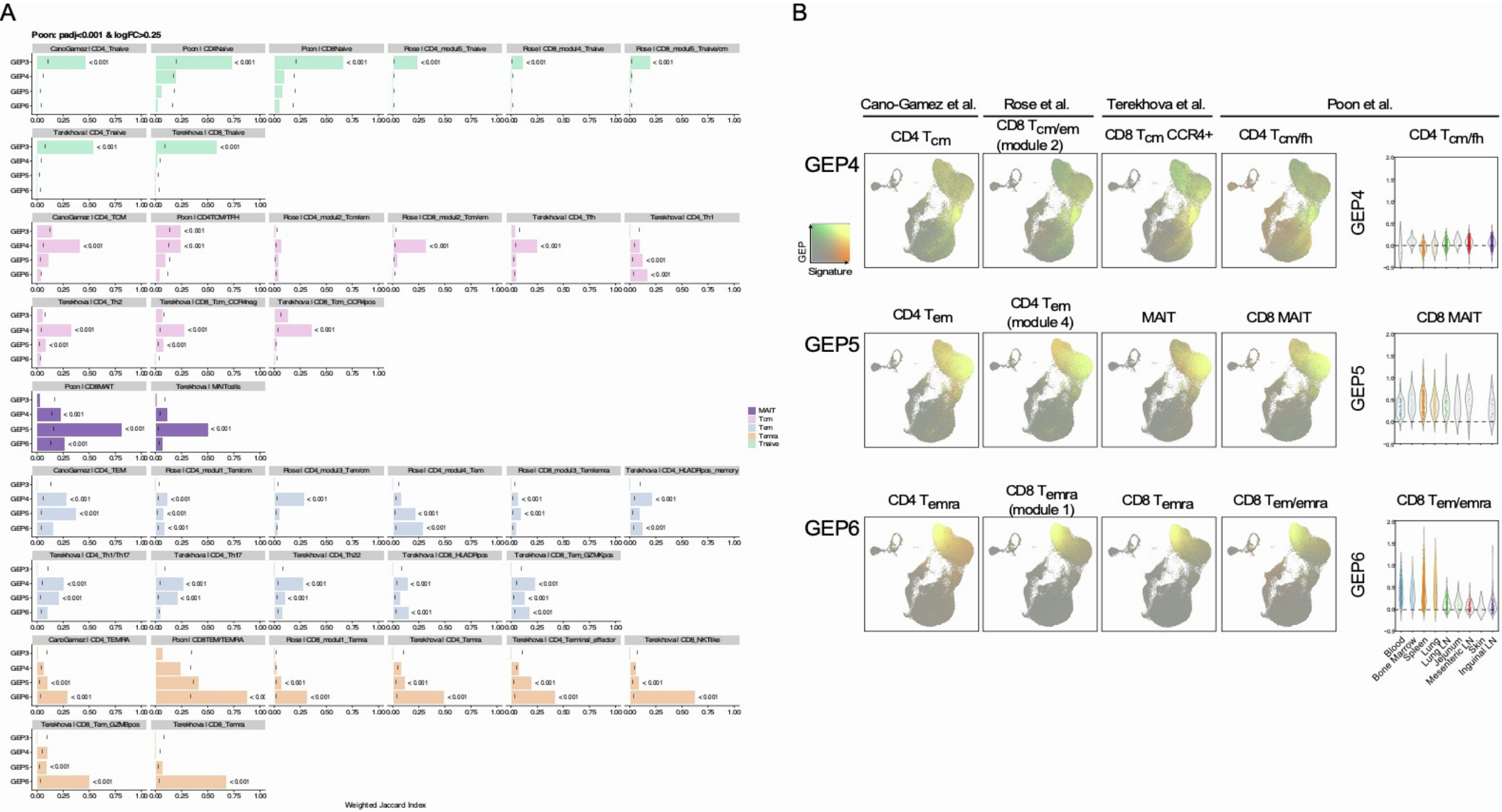
Supplementary Figure 7: Separate analysis of human innate T cell development. Clustering of hashtag-separated thymic iNKT (A), MAIT (E) and γδ T cells (I) and the respective proportion of cells per cluster and donor (B, F, J), the projection of the CD8αα signature (C, G, K) and egress score (D, H, L).



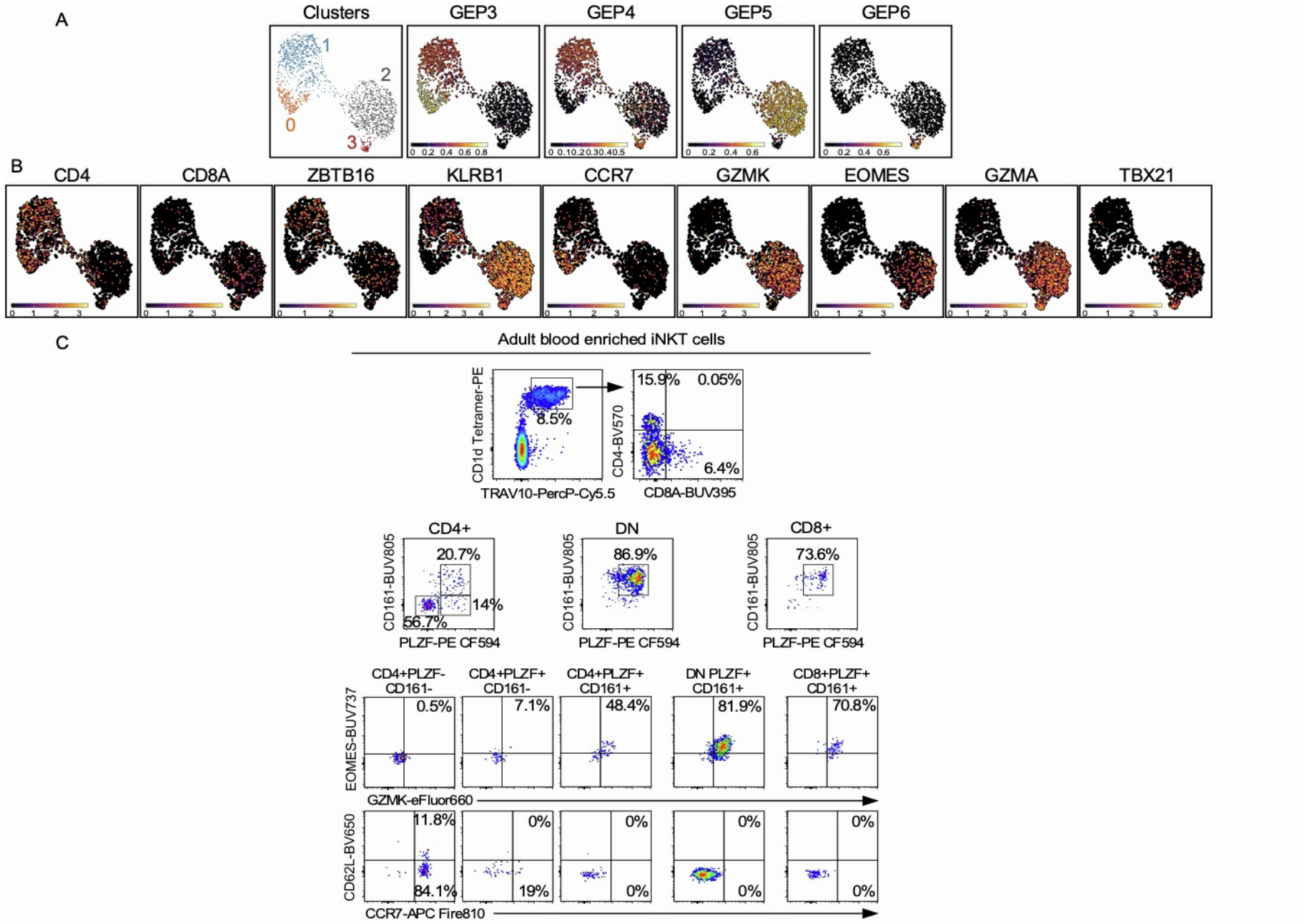
Supplementary Figure 8: Gene expression programs (GEP) in thymic T cell types. Cells are color-coded based on their respective GEP usage (rows) and cell types (columns). GEP usage derived from cNMF usage file.



Supplementary Figure 9: Effector phenotyping of thymic iNKT and MAIT cells by flow cytometry. Thymic iNKT (TRAV10^{+} CD1d-PBS57^{+}) and MAIT (TRAV1-2^{+} MR1-5OPRU^{+}) cells from postnatal thymus were analyzed by flow cytometry for the expression of co-receptors CD4 and CD8; transcription factor PLZF; and effector markers CD161, EOMES, GZMK. iNKT and MAIT cells were pre-enriched via CD1d-PBS57 and MR1-5OPRU magnetic beads.

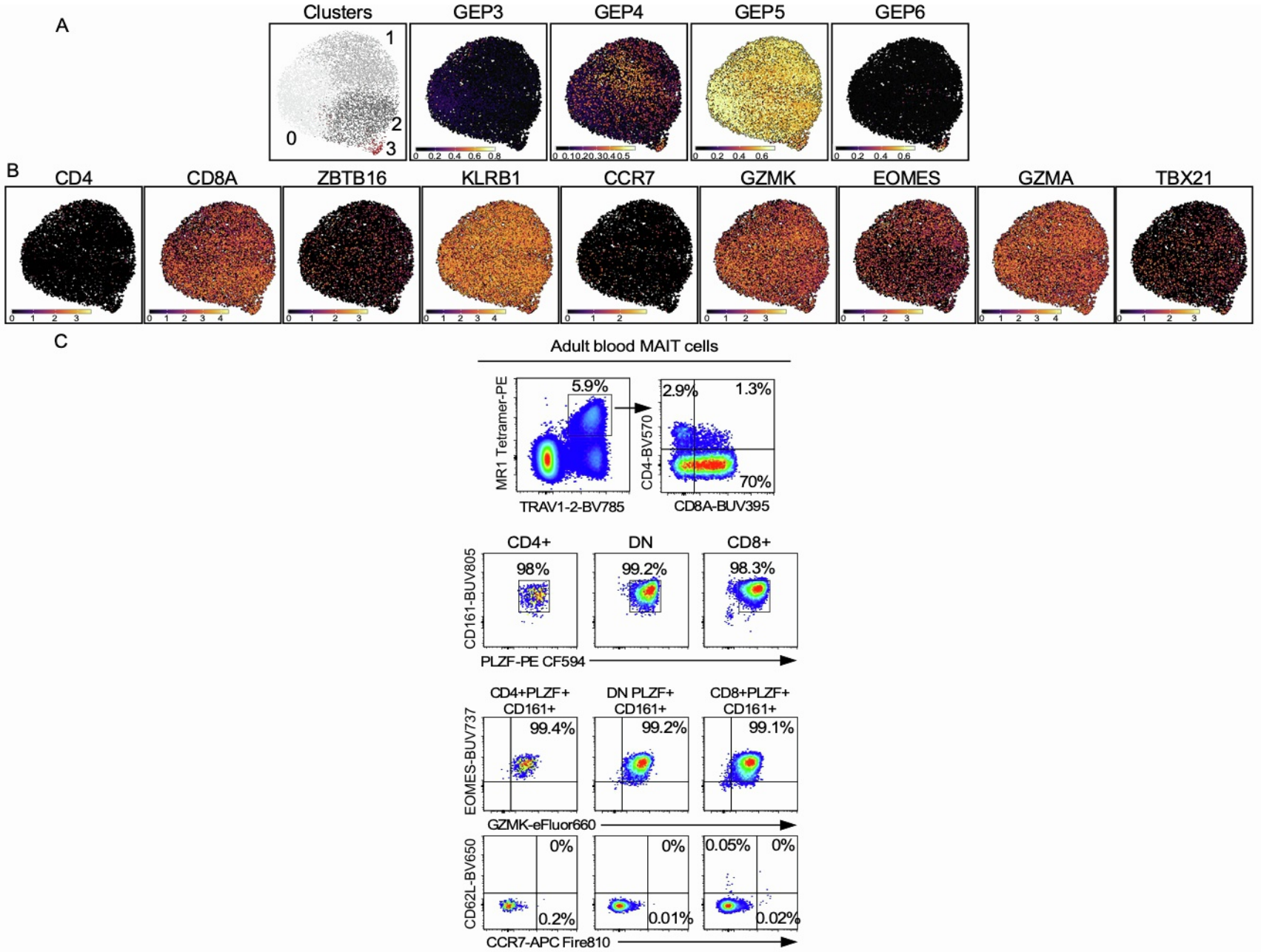


Supplementary Figure 10: Effector gene expression programs (GEPs) are consistent across datasets and human tissues. A. Proportion of genes in each peripheral GEP (3-6) corresponding to genes in public signature gene lists (Poon *et al.*[39], Rose et al.[38], Cano-Gamez *et al.*[37], Terekhova *et al.*[40]) measured by weighted Jaccard Index. For each GEP, the top gene lists with the highest overlap are shown. Tick marks represent the overlap expected from an empirical null distribution (see methods). B. Co-expression of effector GEPs (GEP4-6) and signature gene lists represented on integrated UMAP. For each GEP the co-expression with the gene list corresponding to the highest weighted Jaccard Index (from A) are shown. For the Poon dataset, violin plots on the right represent the effector GEPs scored in cells from the CD4 $T_{cm/fh}$, CD8 MAIT, or CD8 $T_{em/emra}$ clusters, across tissues; the horizontal dashed line is the median score across all clusters and all tissues from the Poon dataset.

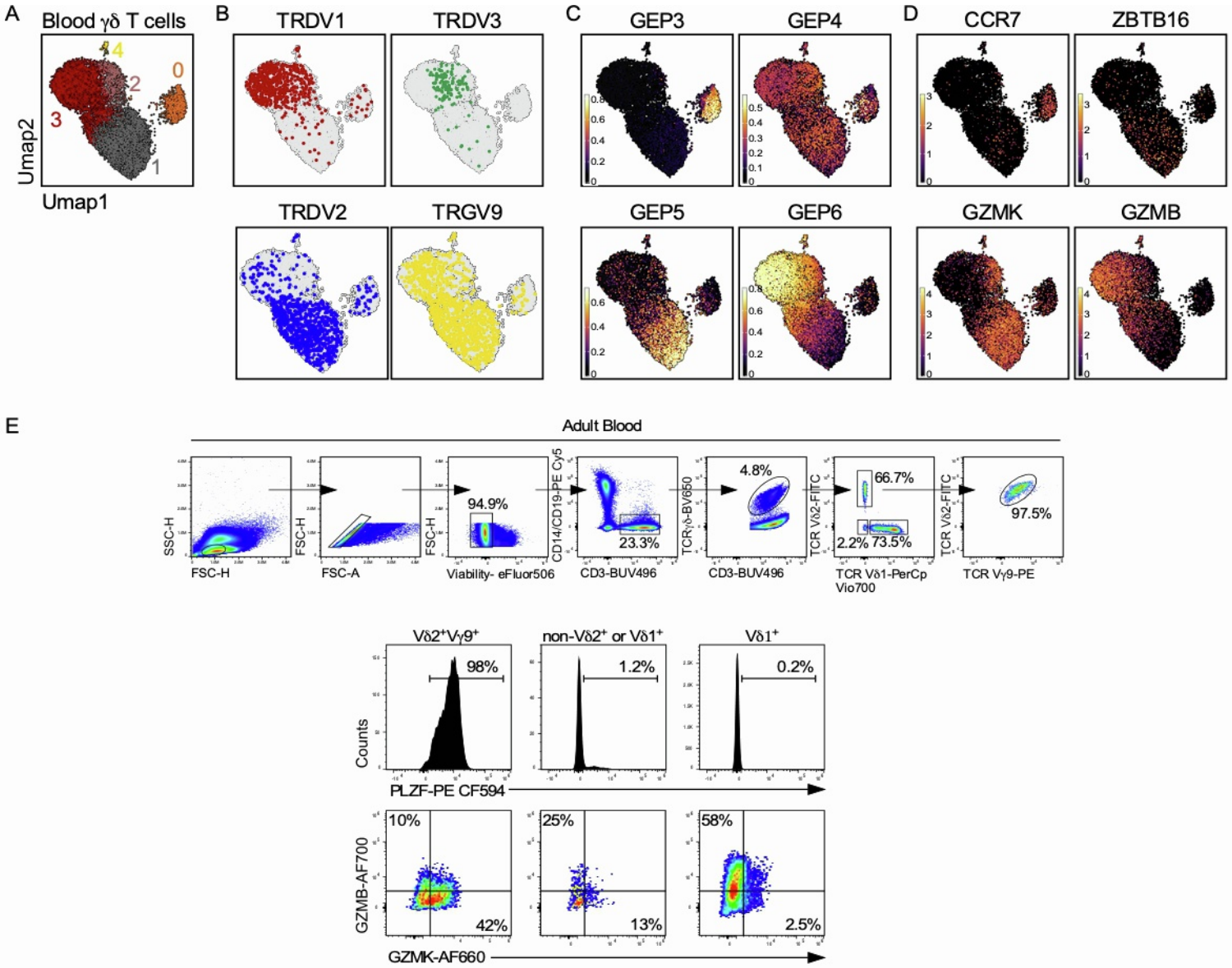


Supplementary Figure 11: Naïve and effector gene and protein expression of adult peripheral blood iNKT cells. A. Cluster assignment (as in Fig. 4A) and projection of naive-like (GEP3) and effector (GEP4-6) on adult peripheral blood iNKT cells (identified by cell hashtag). B. Gene expression projection of co-receptors (*CD4*, *CD8*), transcription factors *ZBTB16* (encoding PLZF) and *TBX21* (encoding TBET), naïve T cell marker CCR7 and effector markers *KLRB1* (encoding CD161), *EOMES*, and granzymes *GZMA*, *GZMK*;

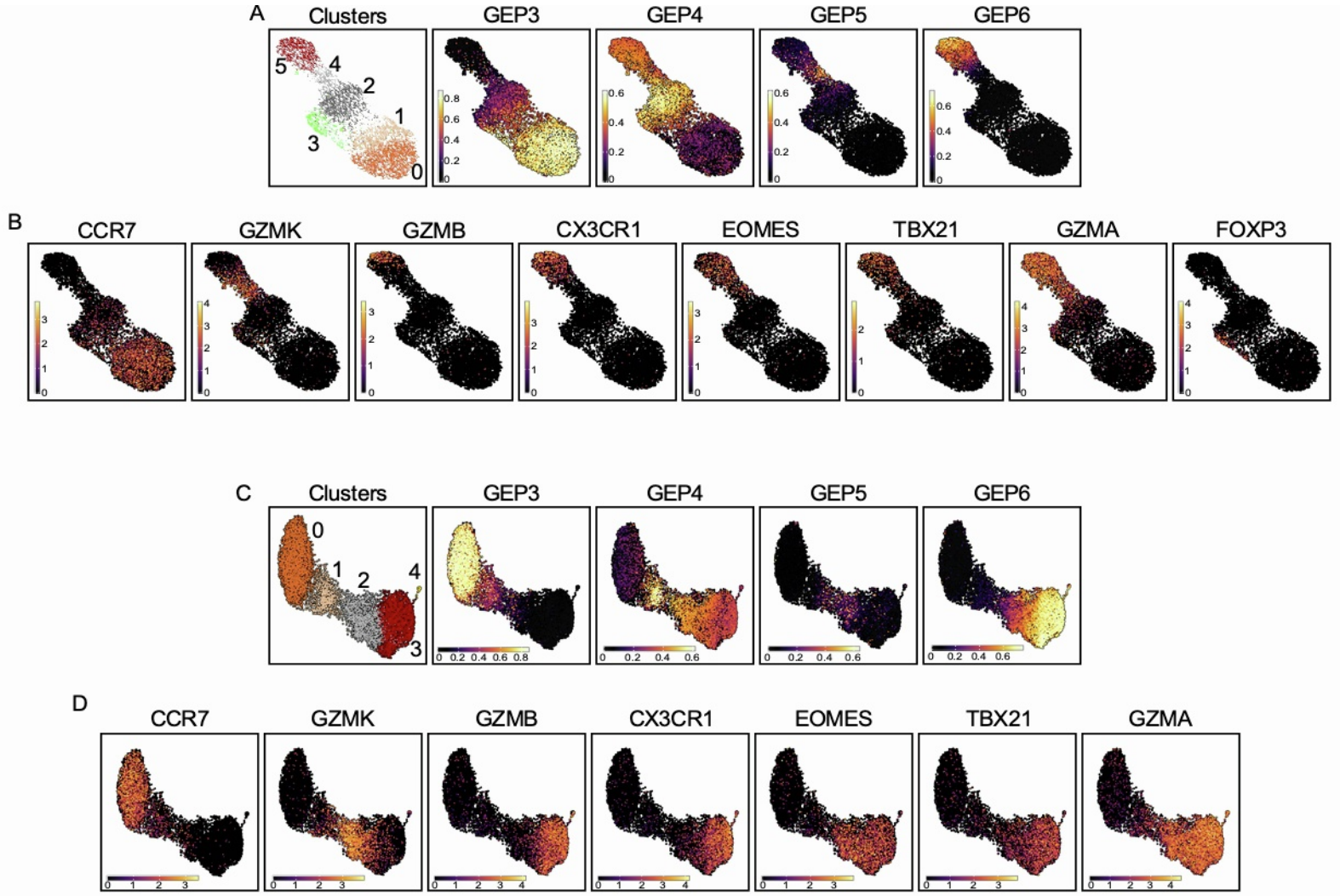
C: Flow cytometry of adult peripheral blood iNKT cells (TRAV10^{+} CD1d-PBS57^{+}) for a characteristic subset of markers in B.



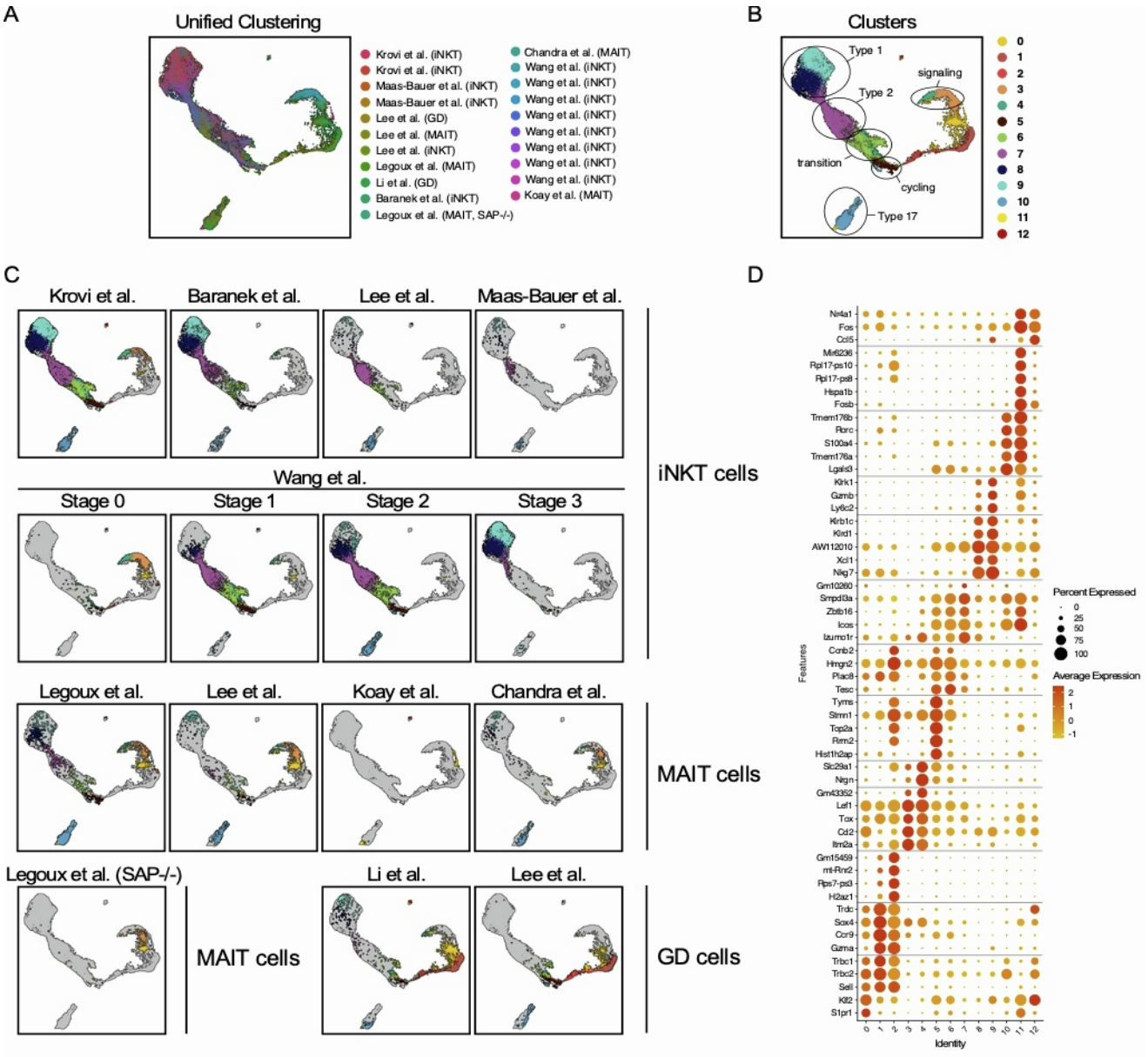
Supplementary Figure 12: Naïve and effector gene and protein expression of adult peripheral blood MAIT cells.
A. Cluster assignment (as in Fig. 4A) and projection of naive-like (GEP3) and effector (GEP4-6) on adult peripheral blood MAIT cells (identified by cell hashtag). B. Gene expression projection of co-receptors (*CD4*, *CD8*), transcription factors *ZBTB16* (encoding PLZF) and *TBX21* (encoding TBET), naïve T cell marker *CCR7* and effector markers *KLRB1* (encoding CD161), *EOMES*, and granzymes *GZMA*, *GZMK*; C: Flow cytometry of adult peripheral blood MAIT cells (TRAV1-2^{+} MR1-5OPRU^{+}) for a characteristic subset of markers in B.



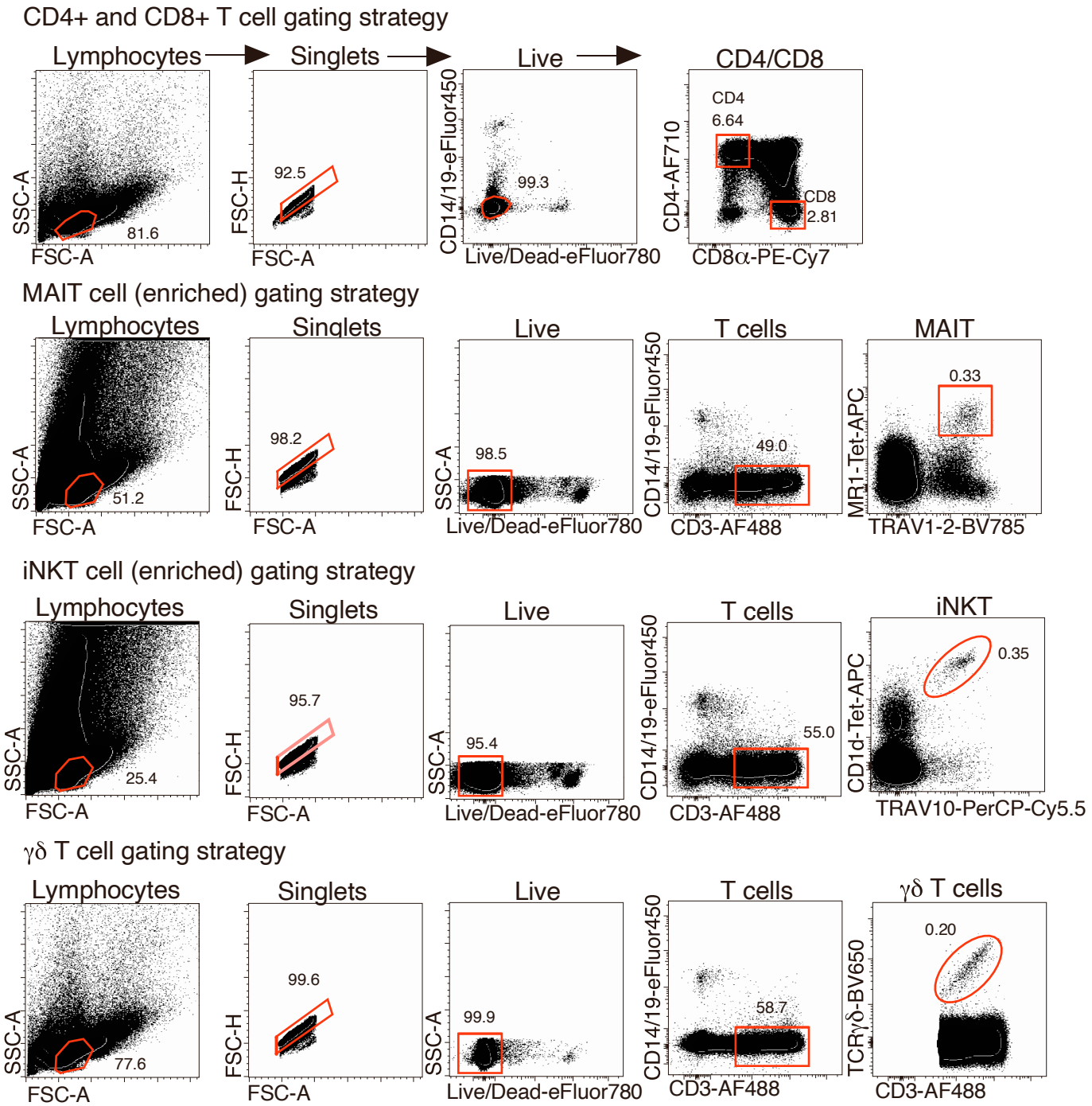
Supplementary Figure 13: Gene and protein expression of adult peripheral blood γδ T cells. A. Cluster assignment (as in Fig. 4A). B. γ and δ variable segment usage (D-V1-3, G-V9), and C. projection of naive-like (GEP3) and effector (GEP4-6) on adult peripheral blood γδ T cells (identified by cell hashtag). D. Gene expression projection of transcription factors *ZBTB16* (encoding PLZF), naïve T cell marker *CCR7* and granzymes *GZMB*, *GZMK*; E: Flow cytometry of adult peripheral blood γδ T cells. γδ T cells were separated by γ and δ chain usage, either as $Vδ2^+Vγ9^+$, $Vδ1^+$, or non-$Vδ1^+$ non-$Vδ2^+$ cells and analyzed for their expression of the granzymes in D.



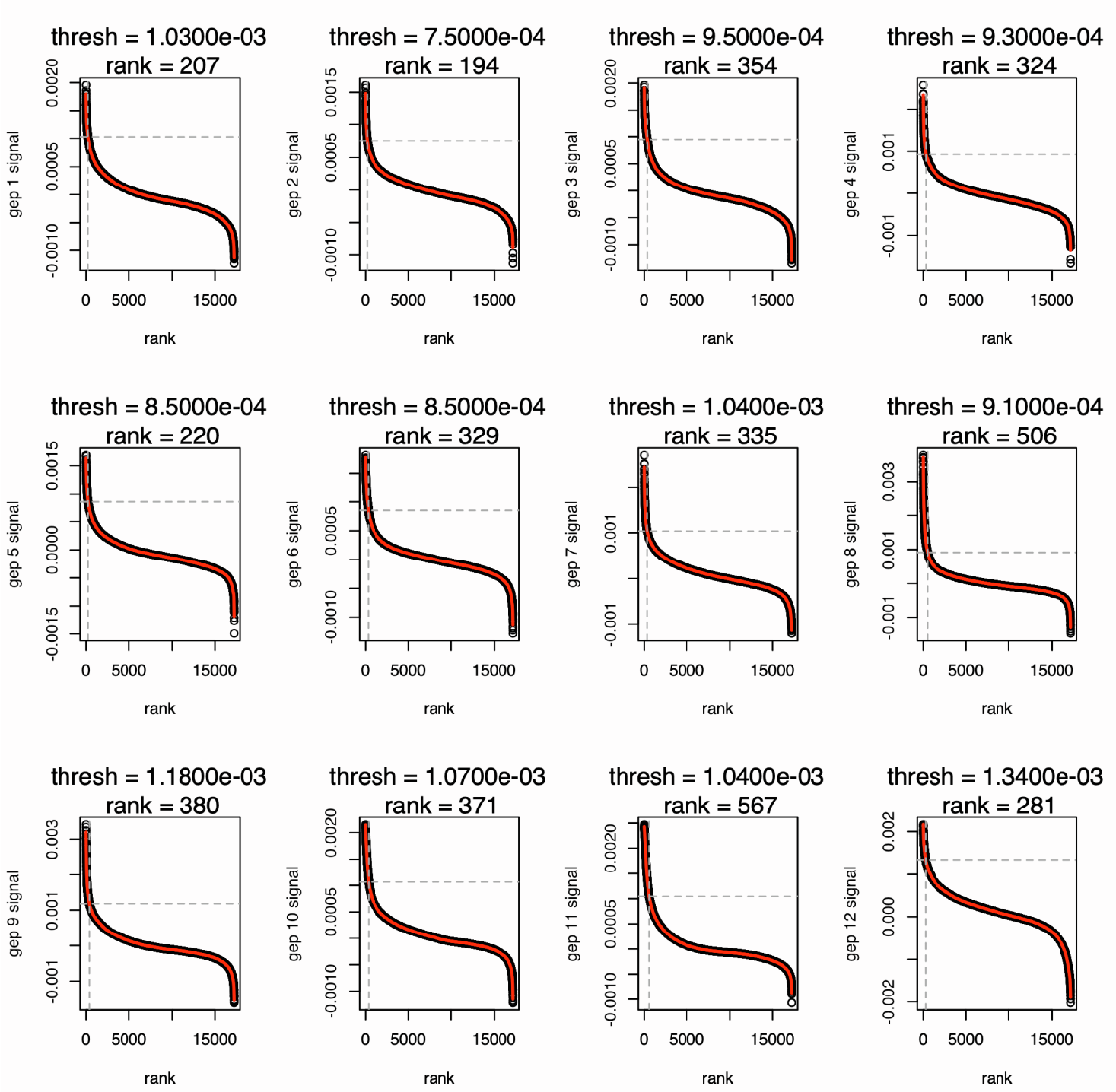
Supplementary Figure 14. Characteristic gene and protein expression of adult peripheral CD4 and CD8 T cells. A./C. Cluster assignment (as in Fig. 4A) and projection of naive-like (GEP3) and effector (GEP4-6) on adult peripheral blood CD4 and CD8 T cells (identified by cell hashtag), respectively. B./D. Gene expression projection of transcription factors *TBX21* (encoding TBET), *FOXP3*, naïve T cell marker *CCR7* and effector marker *EOMES*, chemokine receptor *CX3CR1* and granzymes *GZMA*, *GZMB*, *GZMK*.



Supplementary Figure 15. Reference mouse T_{inn} dataset. A. Integration of single-cell RNAseq data from flow-sorted mouse iNKT, MAIT, or γδ T cells combined from nine independent studies (Refs 52-58) and B. their annotation into 13 clusters, C. spanning across studies and cell lineages. D. Bubble plot of key genes characterizing the 13 clusters.



Supplementary Figure 16. Gating strategies implemented to identify the various T cell populations for analyses and sorting. The target (red gate) cell population in each panel is indicated above each panel. iNKT and thymic MAIT cells were pre-enriched by CD1d-PBS57 and MR1-5OPRU tetramers and magnetic beads, respectively.



Supplementary Figure 17. Determining genes associated with cNMF derived Gene Expression Programs (GEPs). Gene ranks (sorted most to least associated, x-axis) are displayed against their gene_spectra_score output from the cNMF analysis (y-axis) as black dots. The slope at the first elbow point in the fitted sigmoid curve (red line) was calculated as the minimum threshold for genes to be retained in the given GEP. The same slope (grey dashed line) was applied to every GEP to prevent bias in ranked gene selection, as the gene ranking between GEPs are not comparable and relative to each GEP.