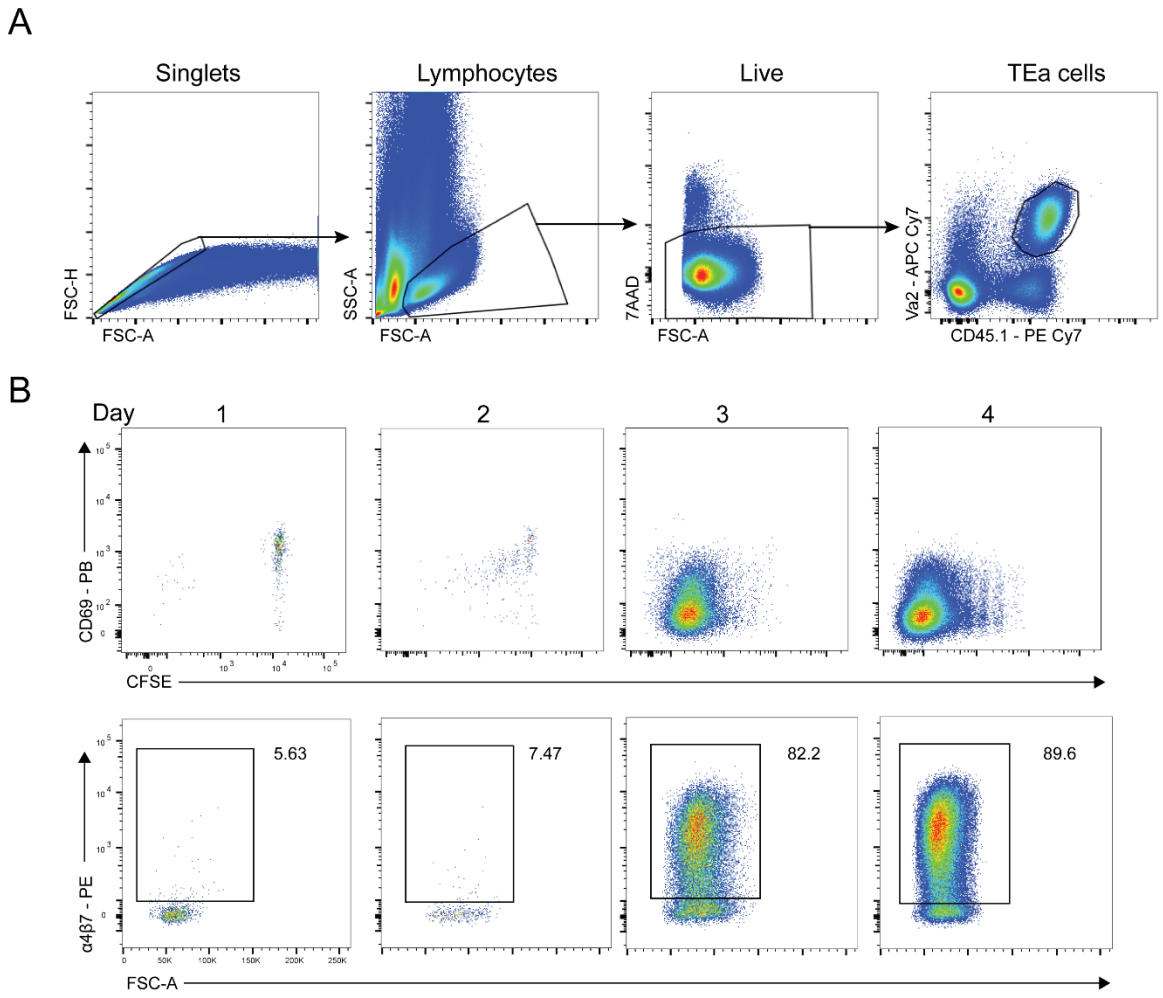
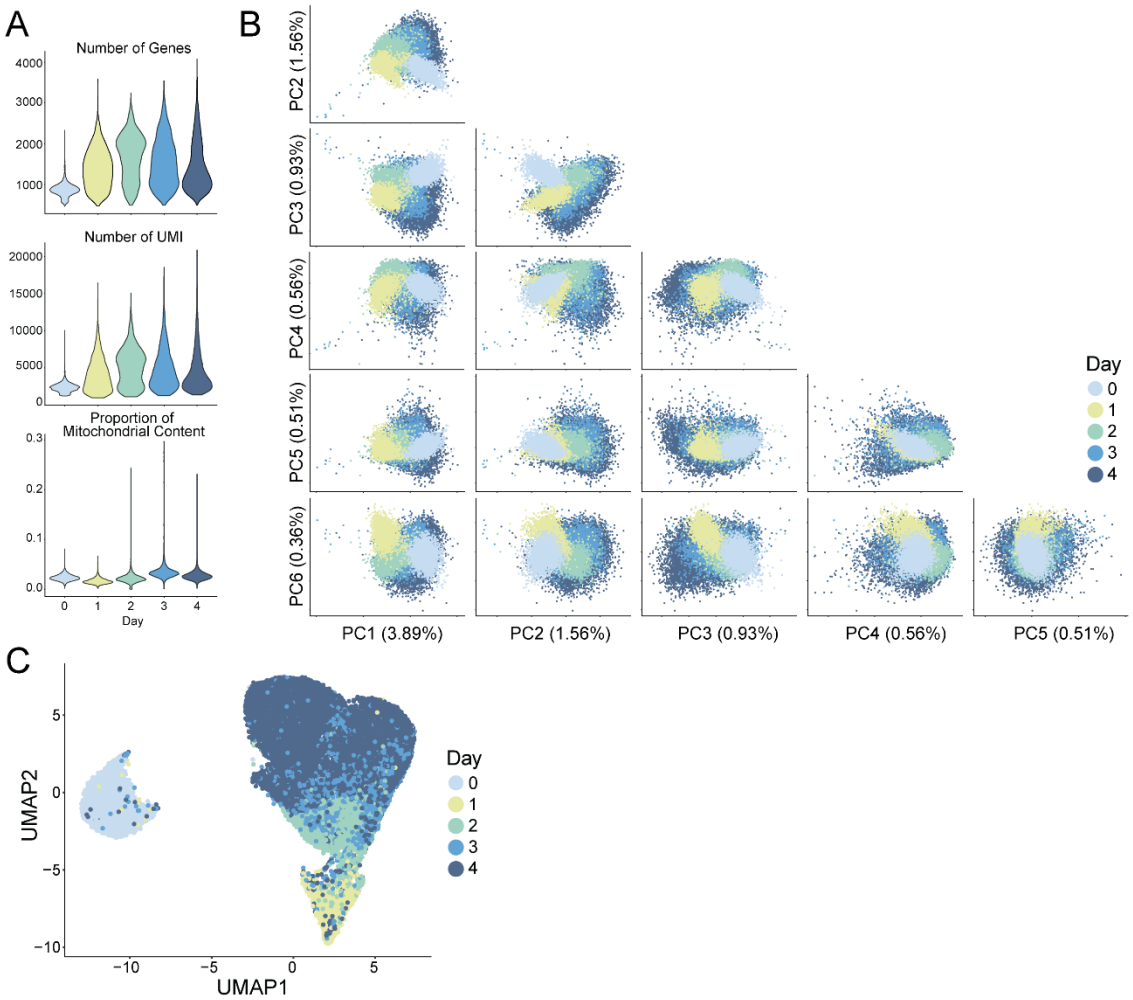
Supplemental data



Supplementary Figure 1. Phenotypic assessment of TEa cells from the mLN.

(A) FACS gating strategy for isolation of TEa cells (Vα2^{+} CD45.1^{+}) from the mLN on day 1-day 4 post-transfer.

(B) Representative FACS plots showing the expression of CD69, α4β7 and CFSE tracking dye on TEa cells from the mLN on days 1-4 post-transfer.

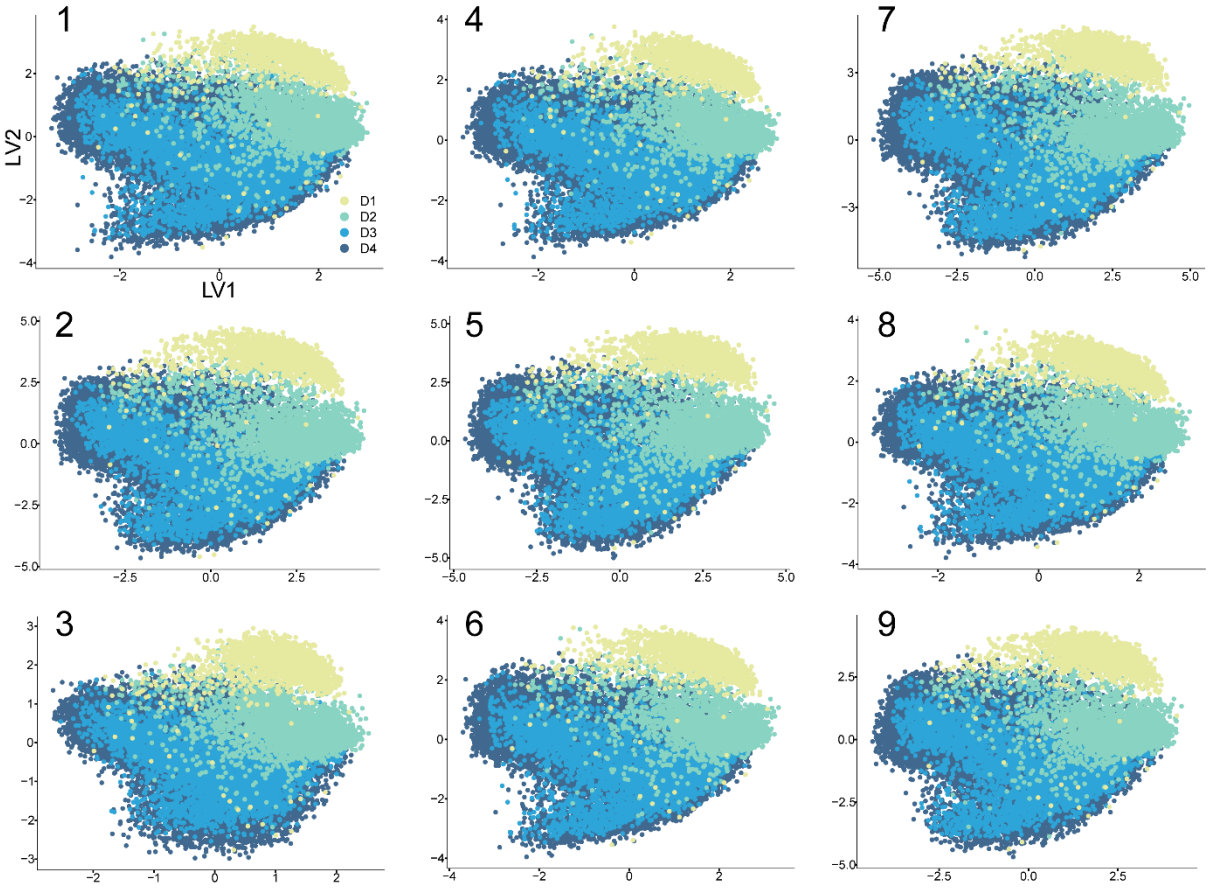


Supplementary Figure 2. Quality control metrics of scRNA-seq dataset from the mLN.

(A) Violin plots showing the distribution of TEa cells for number of expressed genes (cells filtered for 200<nGene<6000), number of UMIs and proportion of mitochondrial content (cells filtered for <0.35).

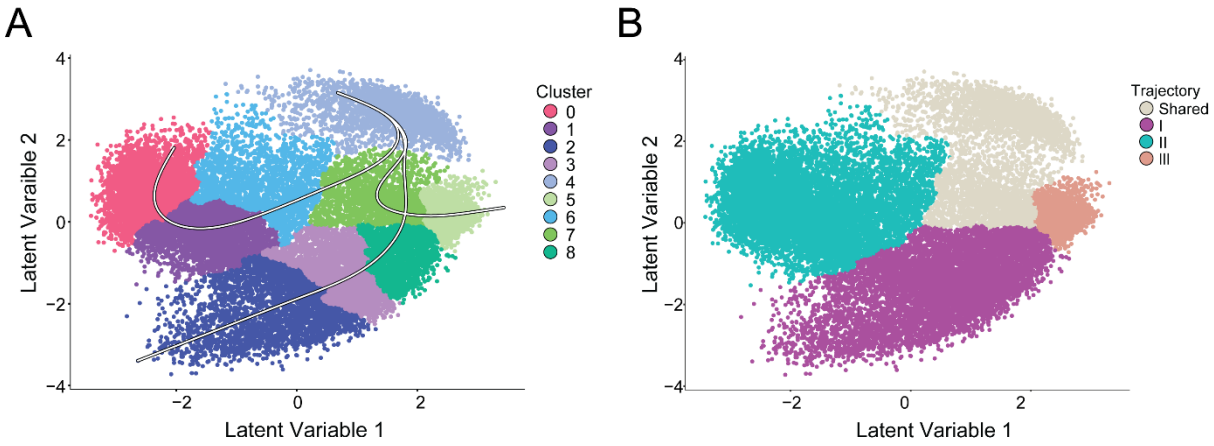
(B) Principal component analysis (PCA) plots showing TEa cells from the mLN at day 0- day 4 post-transfer.

(C) UMAP representation of TEa cells from the mLN at day 0- day 4 post-transfer.



Supplementary Figure 3. Nine replicates of bGPLVM on the scRNA-seq dataset from the mLN.

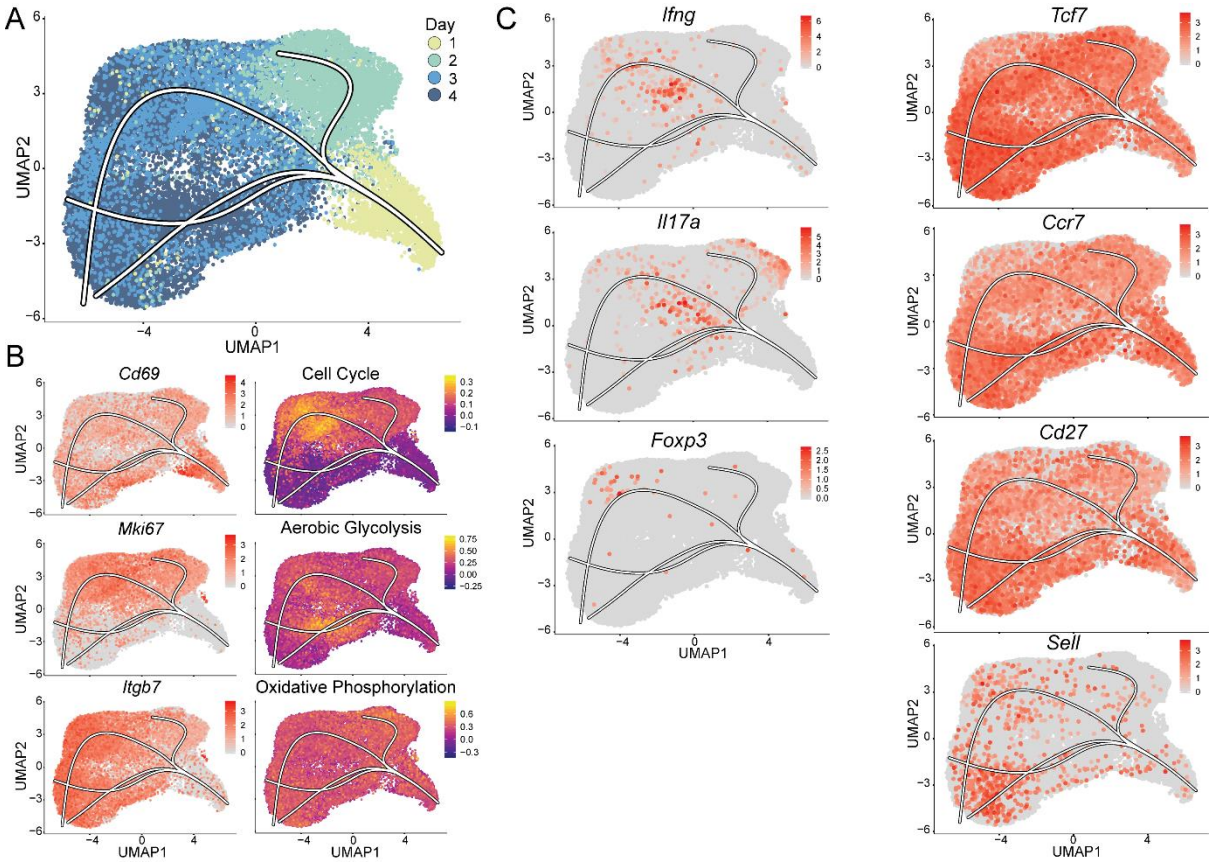
bGPLVM was run an additional nine times on the day 1-4 post-transfer dataset (mGVHD2). The first two latent variables from each run are displayed. Cells are coloured by timepoint.



Supplementary Figure 4. Cluster assignment of the scRNA-seq dataset used for computational modelling with Slingshot.

(A) Cluster assignment of cells on bGPLVM visualisation used as input for Slingshot to identify the overlaid developmental trajectories.

(B) bGPLVM representation showing the assignment of cells to either Trajectory I, II, or III derived from Slingshot, or a 'shared' cluster which is common to all three trajectories prior to branching.

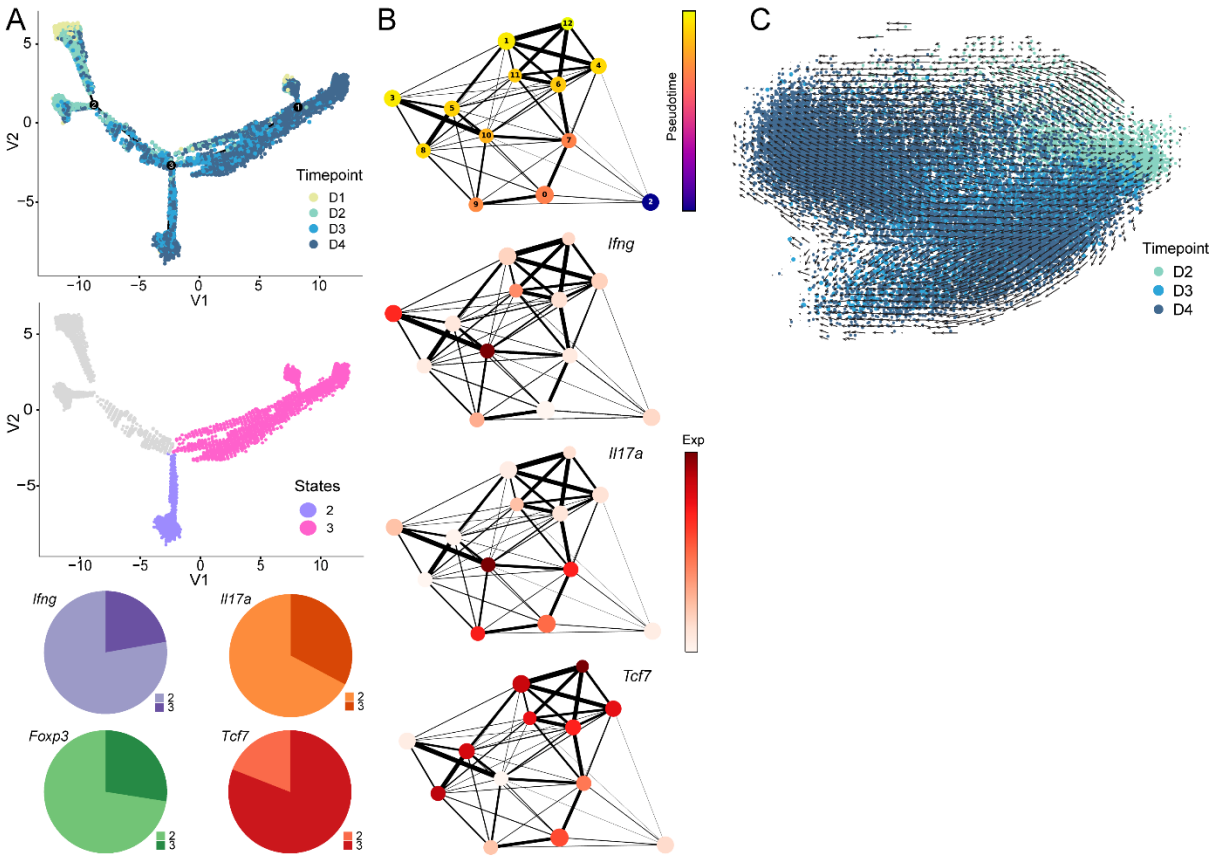


Supplementary Figure 5. An alternative dimensionality reduction, UMAP, coupled with Slingshot for identification of TEa developmental trajectories.

(A) UMAP representation of TEa cells on day 1- day 4 overlaid with the developmental trajectories identified by Slingshot.

(B) Expression of *Cd69*, *Mki67* and *Itgb7* or the cell cycle, aerobic glycolysis and oxidative phosphorylation gene signature scores on the UMAP representation overlaid with Slingshot trajectories.

(C) Expression of *Ifng*, *Il17a*, *Foxp3*, *Tcf7*, *Ccr7*, *Cd27* and *Sell* on UMAP representation overlaid with Slingshot trajectories.

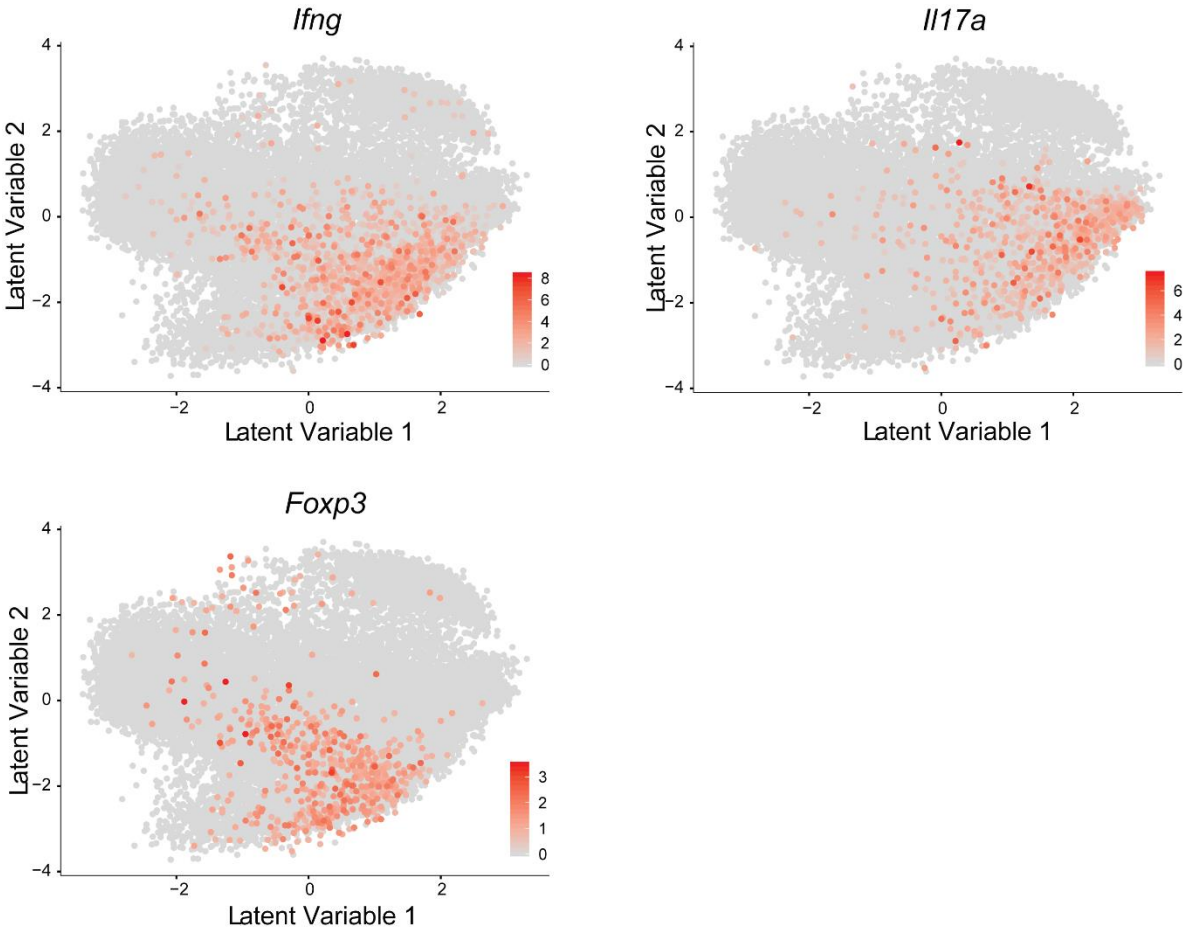


Supplementary Figure 6. Application of other trajectory inference methods resolves similar developmental bifurcation.

(A) Monocle DDRTree output with cells coloured by timepoint (*top*) and developmental states of interest (numbered) as their corresponding slingshot trajectory (*middle*). Pie charts (*bottom*) show the number of cells expressing *Ifng*, *Il17a*, *Foxp3*, and *Tcf7* within the states.

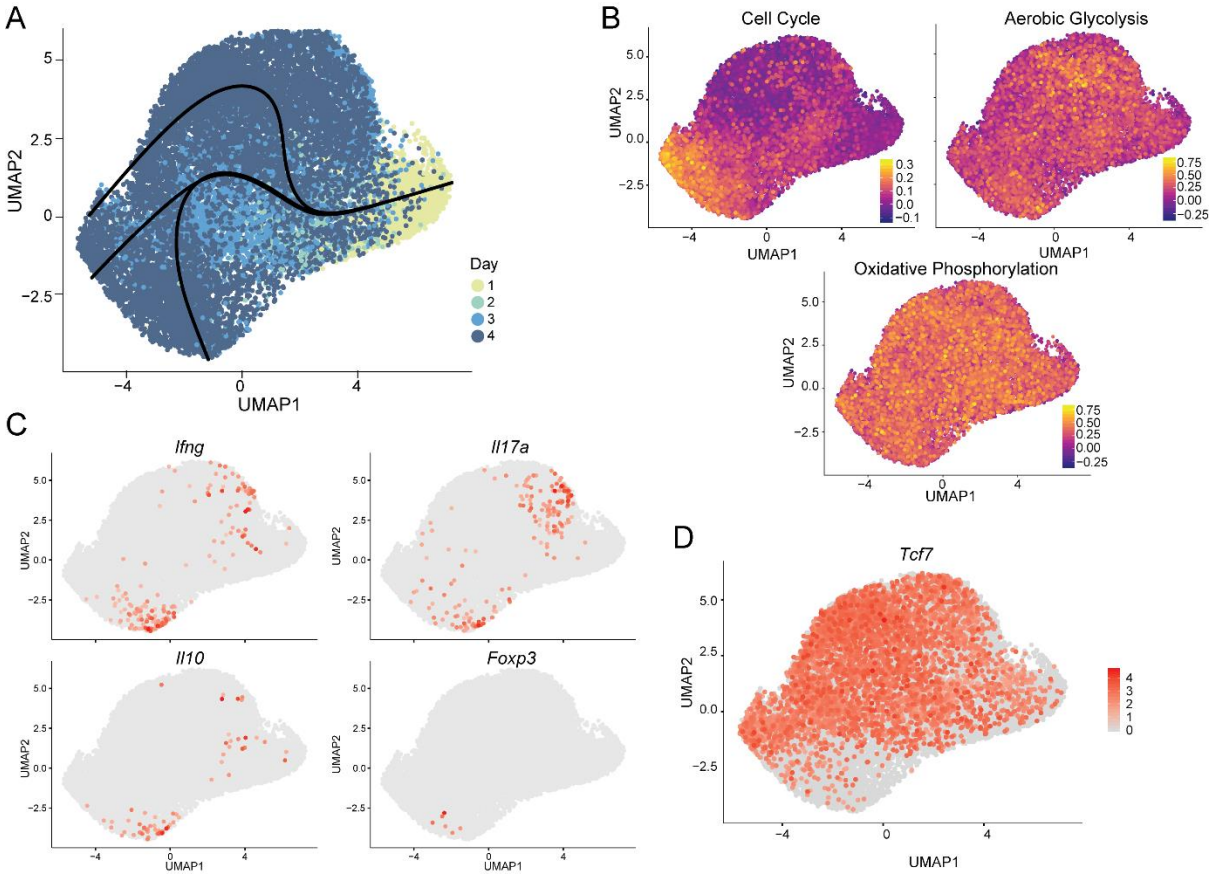
(B) PAGA graph output where each node represents a group of cells characterized by unsupervised clustering and edge weights quantify the connectivity between groups. (*Top*) Cells are shaded by pseudotime and each cluster labelled with a number. Cluster 2 approximately corresponds with day 1 cells. (*Bottom*) Average *Ifng*, *Il17a*, and *Tcf7* expression in each cluster.

(C) Output from scVelo superimposed on bGPLVM latent variables 1 and 2. Cells shaded by timepoint. Arrows indicate inferred RNA velocity direction and arrow length corresponds to the magnitude of the calculated velocity.



Supplementary Figure 7. Imputation of cytokine expression does not reveal distinct trajectories among *Tcf7*^{lo} TEa cells.

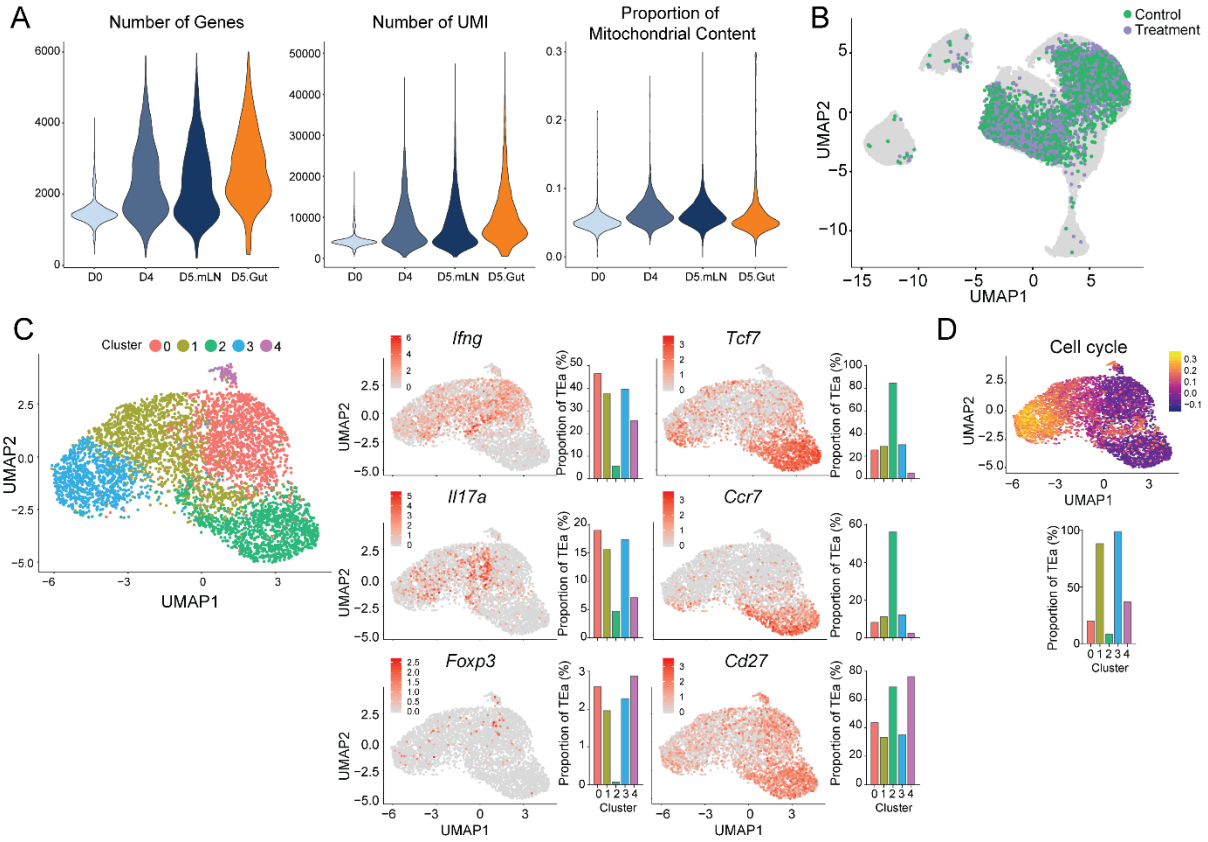
Imputed expression of *Ifng*, *Il17a*, and *Foxp3* by ALRA on bGPLVM latent variables 1 and 2.



Supplementary Figure 8. Independent replicate of scRNA-seq from the mLN with UMAP and Slingshot.

(A) UMAP representation of day 1- day 4 cells in an independent replicate (mGVHD1). Superimposed trajectories generated via Slingshot.

(B) UMAP representation of day 1- day 4 cells in an independent replicate with *Ifng*, *Il17a*, *Il10*, and *Foxp3* expression superimposed.



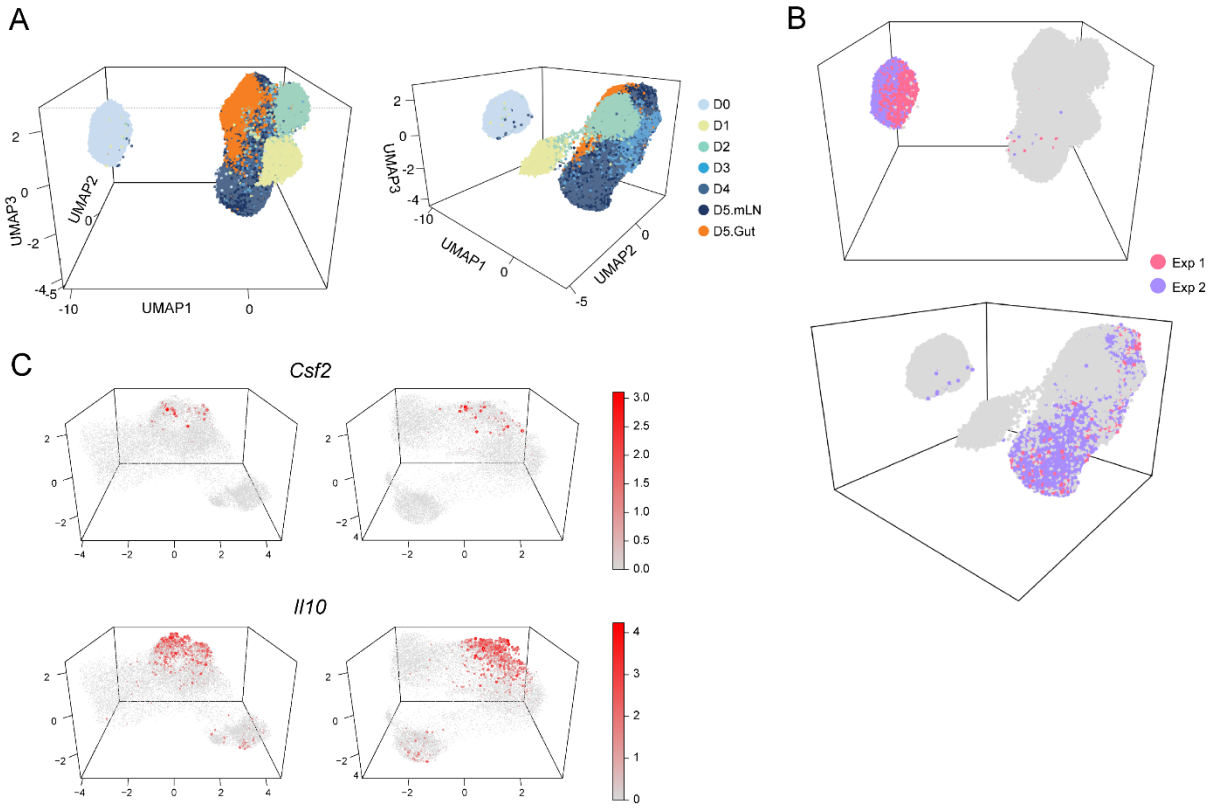
Supplementary Figure 9. Quality control and preliminary analysis of scRNA-seq dataset from the mLN and gut.

(A) Violin plots showing the distribution of TEa cells for number of expressed genes (cells filtered for 200<nGene<6000), number of UMIs and proportion of mitochondrial content (cells filtered for <0.3 or <0.15).

(B) UMAP representation of integrated replicates of scRNA-seq datasets generated using PCA (Seurat) without batch correction. Day 5 post-transplant mLN TEa cells are shaded according to whether they underwent the standard extraction protocol (control) or the extraction protocol for gut (IEL) TEa (treatment).

(C) UMAP representation using PCA (Seurat) of day 5 post-transplant gut (IEL) TEa cells, shaded by cluster from unsupervised clustering (Seurat) (*left*). Expression of *Ifng*, *Il17a*, *Foxp3*, *Tcf7*, *Ccr7*, and *Cd27* on UMAP representation (*right*). Bar graphs represent the proportion of cells within each cluster that express each gene.

(D) Same dataset and embeddings as (C) with the cell cycle gene signature score.

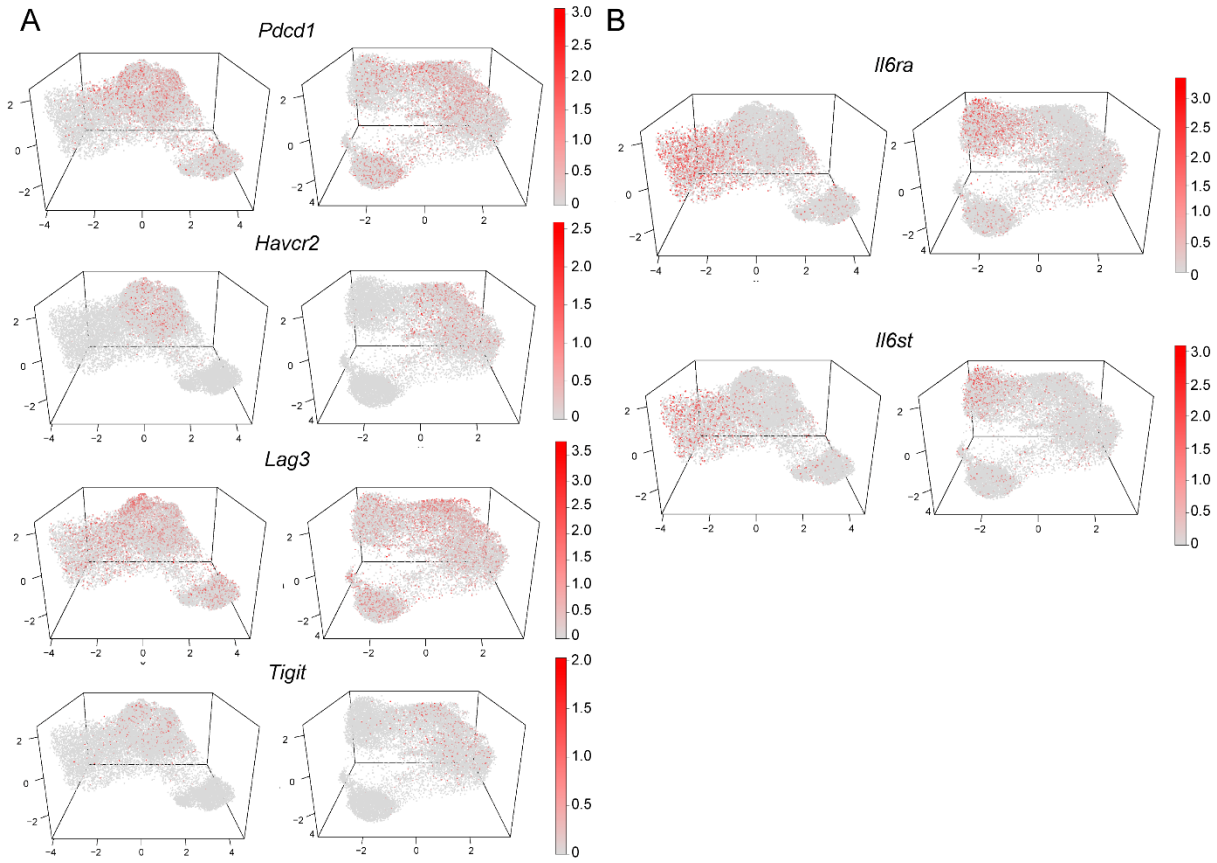


Supplementary Figure 10. Integration of mLN and gut scRNA-seq experiments.

(A) Two angles of 3-dimensional UMAP representation produced from 30 scVI latent variables of day 0 – day 5 TEa cells.

(B) Same UMAP representation as (A) with cells shaded by their experiment of origin, highlighting cells from day 0 (*top*) or from day 4 (*bottom*).

(C) Two angles of 3-dimensional UMAP representation showing TEa cells from day 2 mLN, day 5 mLN, and day 5 gut, with expression of *Csf2* and *Il10*.



Supplementary Figure 11. Analysis of expression of genes associated with T cell exhaustion and IL-6 signalling.

(A) Expression of *Pdcd1*, *Havcr2*, *Lag3* and *Tigit* on 3-Dimensional UMAP visualisation.

(B) Expression of *Il6ra* and *Il6st* on 3-Dimensional UMAP visualisation.