Supplement to: Single-cell analysis reveals regulatory gene expression dynamics leading to lineage commitment in early T cell development

By W. Zhou, M. A. Yui, B. A. Williams, J. Yun, B. J. Wold, L. Cai, and E. V. Rothenberg

SUPPLEMENTARY TABLE TITLES AND LEGENDS; SUPPLEMENTARY FIGURE LEGENDS; SUPPLEMENTARY FIGURES

SUPPLEMENTARY TABLE TITLES AND LEGENDS

Supplementary Table S1, related to Fig. 1. Bulk RNAseq data of genes differentially expressed between all ETP samples and Bcl11b-YFP^{+} DN2a samples. Genes with the average RPKM larger than 1, expression fold change larger than 2 either way and adjusted pval < 0.05 are shown. Values in RPKM.

Supplementary Table S2, related to Fig. 2 and Fig. 5. Curated regulatory and marker genes used in seqFISH analysis and supervised 10X Chromium analysis. Table indicates gene names and the combinations of criteria used for selecting each of these genes as particularly informative, based on their genetically defined functional importance or use as developmental state indicators (www.immgen.org)(Mingueneau et al., 2013)[reviewed in (Longabaugh et al., 2017; Rothenberg et al., 2016; Yui and Rothenberg, 2014)].

Supplementary Table S3, related to Fig. 2. SeqFISH raw transcript data and analysis of transcript distribution comparison between different stages of pro-T cells. Populations being compared are *Gata3*- (≤3 transcripts) **and** *Tcf7*- (≤5 transcripts) double negative ETPs, *Gata3*+ (>10 transcripts) **or** *Tcf7*+ (>20 transcripts) ETPs, and *Bcl11b*+ (> 5 transcripts) DN2s. Thresholds for binning were drawn to identify clear positives and negatives and avoid ambiguous intermediate levels of expression. Highlighted are p values $<10^{-6}$, two-tailed T test, unequal variances. Analysis performed from 4-week-animal dataset.

Supplementary Table S4, related to Fig. 3, Fig. 4, and Fig. S8. C1 and 10X marker genes identified in each sub-cluster in the analyses shown. Clustering based on SLM, markers identified with minimum fraction of 0.2 in the cluster and threshold of 0.2 using Wilcoxon rank sum test in Seurat 2. C1 supervised analysis was performed as shown in Fig.3. 10X unsupervised analysis was performed in Seurat 2 as shown in Fig. 4. 10X supervised analysis was performed as described in Fig. S8. pct.1, pct.2: weightings in principal components 1, 2 respectively.

Supplementary Table S5, related to Fig. 5. Differentially expressed genes identified by supervised pseudo-time analysis from 10X analysis (intersection of both independent 10X replicates, qval<1E-08). The genes are ordered and clustered based on the pseudotime expression pattern as shown in Fig. 5. The crosses mark the individual genes that overlap with perturbation assays, which were shown to be significantly regulated by PU.1 or Bcl11b, as described in Fig. 5f-g. Lists of genes regulated by PU.1 or by Bcl11b were from published data (Hosokawa et al., 2018a, 2018b; Ungerbäck et al., 2018) as described in STAR Methods.

Supplementary Table S6, related to Fig. 2 and STAR methods. Designed oligo probe template pools and sequential color barcode used for seqFISH experiments.