SUPPLEMENTARY FIGURE LEGENDS

Supplement to:

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

Wen Zhou[1], Mary A. Yui[1], Brian A. Williams[1], Jina Yun[1], Barbara J. Wold[1], Long Cai[1], Ellen V. Rothenberg

Supplementary Figure S1, related to Figs. 1, 2, 3, 4, 5, 6, and 7. Summary schematics of biological questions addressed and analysis pipelines used.

a) Summary: logic flow of central biological questions in this study, how each step provides the rationale for the next, and breakdown of specific technologies and analyses used to address the specific questions. Questions are highlighted in red boxes, and results are shaded in gray boxes. Techniques and analysis used are described in italic text and colored in background with blue shading indicating analysis using single cell transcription profiling tools, purple shading indicating bulk RNA analysis, and orange shading indicating cell culturing assays. b) Summarizes relationships between methods, gene and cell filters being used, and data analytical pipelines used in this study. c) Sorting gates and logistics for purifying Kit^{hi} ETP-DN2a and DN3 cells for single cell analyses.

Supplementary Figure S2, related to Fig. 2. Highly sensitive seqFISH provides reproducible and robust RNA transcript quantitation for regulatory genes.

a) Scatterplot comparison between mean values of expression measured in a comparable population (Flt3+ ETP) in different seqFISH experiments with thymocytes from 4, 5, 8-week old animals. b) Scatterplot comparison between mean values of expression, measured in seqFISH and 10X Chromium, of genes listed in Table S2. Mean values taken from cells in comparable cell populations (Flt3+ ETPs). Patterns

were broadly correlated, but seqFISH detected approximately 10 molecules of RNA for each UMI count in the 10X analysis (seqFISH: n=1656, from 3 replicates. 10X: n=863, from replicate1). The result is consistent with the previously described 10% sampling rate of 10X Chromium v2 scRNAseq, at the sequencing depth being used (Islam et al., 2014; Kolodziejczyk et al., 2015) c) tSNE plots with combined 3 seqFISH replicates, including Kit positive and DN3 populations, colored in each panel to indicate the distribution of Kit positive cells from one of the individual replicates. The samples from different experiments and ages are interspersed without batch corrections. d) Detected transcript distribution comparison between seqFISH and 10X Chromium experiment on key regulatory genes in ETP-DN2 cells. (seqFISH: n=2524, 10X: n=4234.) SeqFISH detected the expression of *Notch1, Tcf7,* and *Runx1* in almost all ETP-DN2 cells, in agreement with their known functional roles whereas 10X had a high false negative rate. e) Scatterplots of antibody staining and RNA transcript count correlations, colored by the cell size estimation (area of image segmentation). f) Transcript and antibody distribution of cells at different stages (binned by *Il2ra, Bcl11b* transcripts). c-Kit and TCF1 agree well with *Kit* and *Tcf7* expression at all stages. Note that arrows indicating the PU.1 protein (encoded by *Spi1*) and *Spi1* RNA disagree at DN2b stage, as *Spi1* RNA drops in expression between DN2a-b stage while PU.1 protein appears to persist longer. This is likely a reflection of the extreme stability of PU.1 protein, as reported previously (Kueh et al., 2013). The antibody signals were plotted in arbitrary units on linear scales, with signal quantitation described in Methods. g) Developmentally ordered clusters of seqFISH transcript distribution. Clusters shown here were as presented in Figure 2f, excluding DN3b (cluster 7) and the ‘outlier’ myeloid cluster (cluster 8).

Supplementary Figure S3, related to Fig. 4. 10X Chromium scRNAseq replicates confirm the similar continuity and heterogeneity of cell states and lineage progression within the purified early T cell population.

a-b) Scatterplot comparison between mean values of expression measured in comparable population (early Flt3+ ETPs) in two 10X Chromium scRNAseq replicates

(n=863 cells replicate1, n=1442 cells replicate2). c) tSNE display of 10X replicate2 (7076 cells) colored by cluster. Clustering was performed with SLM algorithm, using PC 1 to 10. d) Heatmap of feature genes enriched in each sub-cluster analyzed in 10X replicate 2, ordered by approximate developmental state. Yellow=high expression, purple=low expression. Compare with similar clustering for replicate 1, shown in Fig. 4d. e) Alignment of seqFISH, 10X, and C1 datasets after CCA scaling, shown in principal components 1-2.

Supplementary Figure S4, related to Figs. 2-4. Discrete committed granulocyte precursor subset within the immunophenotypically-defined ETP population.

a) Experimental plan to test developmental potential. Purified Lin^-DN thymocyte subsets were sorted into wells (25 or 50 cells/well into 96 well plate) for co-culture with pre-plated OP9 stroma, with Notch ligands (OP9-DL1) or without, and then stained and FACs analyzed at indicated timepoints. b) $Flt3^+$, $Flt3^-$ ETPs, and DN2a cells cultured on OP9-DL1 for 4 days, then analyzed for developmental markers, CD44 and CD25, and Bcl11b-YFP. c) CD63+ Ly6c+, and CD63- Ly6c- ETP cells cultured on OP9-DL1 for 4d, then analyzed for markers of T-cell progression, CD25, and granulocytes (Gr1). d) Summary plot of percentages of CD25+ cells and Gr1+ cells after 4-5d culture with OP9-DL1 stromal cells. Thymocytes from Bcl2 transgenic mice were used to enhance cell survival. Also see Fig. S5.

Supplementary Figure S5, related to Figs. 2-4. Commitment assays under conditions lacking Notch signaling to test ETP subsets for alternative lineage potentials.

Subsets of ETPs and DN2a cells were sorted according to surface marker expression patterns indicated and tested for developmental potential under conditions favoring myeloid development. Assays were performed with cells isolated from Bcl2-tg mice to promote survival. 25 or 50 cells/well were plated into 96 well plate with pre-seeded OP9-control stroma cells, cultured for 7 days under myeloid conditions, and analyzed by

flow cytometry as shown. One representative culture is shown from each subset except Flt3$^-$ ETPs, which are represented by two different cultures. Here, subsets are shown with data from the most advanced T-lineage precursors, Bcl11b-YFP$^+$ DN2a cells, which are already T-lineage committed. a) Gating strategies and representative flow cytometry analysis for alternative lineage assays at day 7 of culture. Cells were gated on FSC and SSC, 7AAD negative and CD45 positive for live lymphocytes (top two rows), and then separated by anti-NK1.1 + anti-Dx5 and anti-Gr1 for NK cells (NK) and Granulocytes (Gr1$^+$ cells) respectively (third row). The non-NK and non-Granulocyte population (lower left of panels in third row) was further separated using anti-CD11b and anti-CD11c for Macrophage (MP) and Dendritic cells (DC), respectively. The cells that were negative for all alternative-lineage markers in the staining panels were categorized as 'unknown'. The cell numbers generated from individual categories were divided by the input cell number and displayed in stacked bar graphs in b) and also in Fig. 6e. b) Summary graphs for results of alternative lineage potential assays of ETP subsets distinguished by Flt3, CD63 and Ly6c expression, compared with DN2a cells separated into Bcl11b-YFP$^-$ and Bcl11b-YFP$^+$. Isolated cells were cultured on OP9-Control stroma, under myeloid lineage supporting cytokine conditions for 7 days (see STAR Methods). The stacked bar-graphs represent the developmental potentials of each ETP subset to generate cells of non-T lineages under these permissive conditions. Top panel shows results from Ly6c$^-$CD63$^-$ ETP cells subdivided by Flt3, and DN2a cells subdivided by Bcl11b-YFP. The bottom panel shows results from Ly6c and CD63 single and double positive ETP populations. Under these conditions, CD63$^-$ Ly6c$^-$ ETPs generated multiple types of cells of alternative non-T lineages, including Gr1$^+$ granulocytes (magenta) (top). However, Ly6c$^+$ CD63$^+$ double positive cells gave rise exclusively to Gr1$^+$ granulocytes, while CD63$^+$ or Ly6d$^+$ single positive cells generated Gr1$^+$ cells as well as other lineages (bottom). Individual replicates are presented in separate bars.

Supplementary Figure S6, related to Fig. 4. Resolution of component developmental processes: developmental progression coupled with orthogonal spread of cell cycle signatures in SPRING analysis.

a-b) SPRING display of expression topology of ETP-DN2-DN3 cells. (Performed with PC1-20, and k=5 on raw 10x dataset with cells filtered by minimum 2500 UMI counts, but not by mitochondrial content, and genes filtered by 60th percentiles for variability) a) Expression levels of key genes are highlighted in green on relative scales. The key genes were categorized as early developmental genes (*Flt3, Lmo2*, and *Mef2c*, early ETPs; *Spi1*, all ETP and DN2a cells) and later genes (*Il2ra*, ETP to DN2a transition marker; *Fgf3*, DN2a-specific; *Bcl11b*, commitment marker; *Rag1*, upregulated in DN3a) for marking the developmental direction. The second orthogonal axis was represented by proliferative and cell cycle state markers, with G2/M-active genes *Birc5* and *Mki67* (similarly with *Top2a* and *Cenpa*, not shown*)* expressed by cells at the lower right with the highest UMI counts, with G1-to-S phase cyclin *Ccne2* immediately adjacent, and G1-expressed gene *Samhd1* concentrated at the other end). The committed granulocyte precursor population appeared as a spur (upper right) away from the main distribution (*Elane*). b) Developmental stages and axes annotated based on overall marker expression patterns, with total UMI counts displayed on red-yellow scale as shown.

Supplementary Figure S7, related to Fig. 5. Analysis of RNA velocity using Velocyto: supporting analysis and minimal interaction of cell cycle with population flux.

a) Fraction of 10X Chromium reads mapped to different genomic regions: “spliced” represents exonic reads, “unspliced” represents intronic reads. b) Mean and variable filter of genes that are used in velocity analysis: red dots highlight the gene filter (‘spliced’) for PCA analysis. c) PC1-2 display with arrows indicating the transition probability of cells (imputation and transitioning probability estimation with k=90, quiver scale=0.7, scale type = “relative”.). The vector calculation was performed with (top) and without (bottom) including cell cycle genes (as defined by gene ontology annotation (using Goatools), used in (La Manno et al., 2018)). d) Scatterplots (left panels) display the un-spliced vs spliced isoform distributions after imputation, and gamma fit of the rates of RNA processing for individual genes. Red-blue heatmaps (center plots)

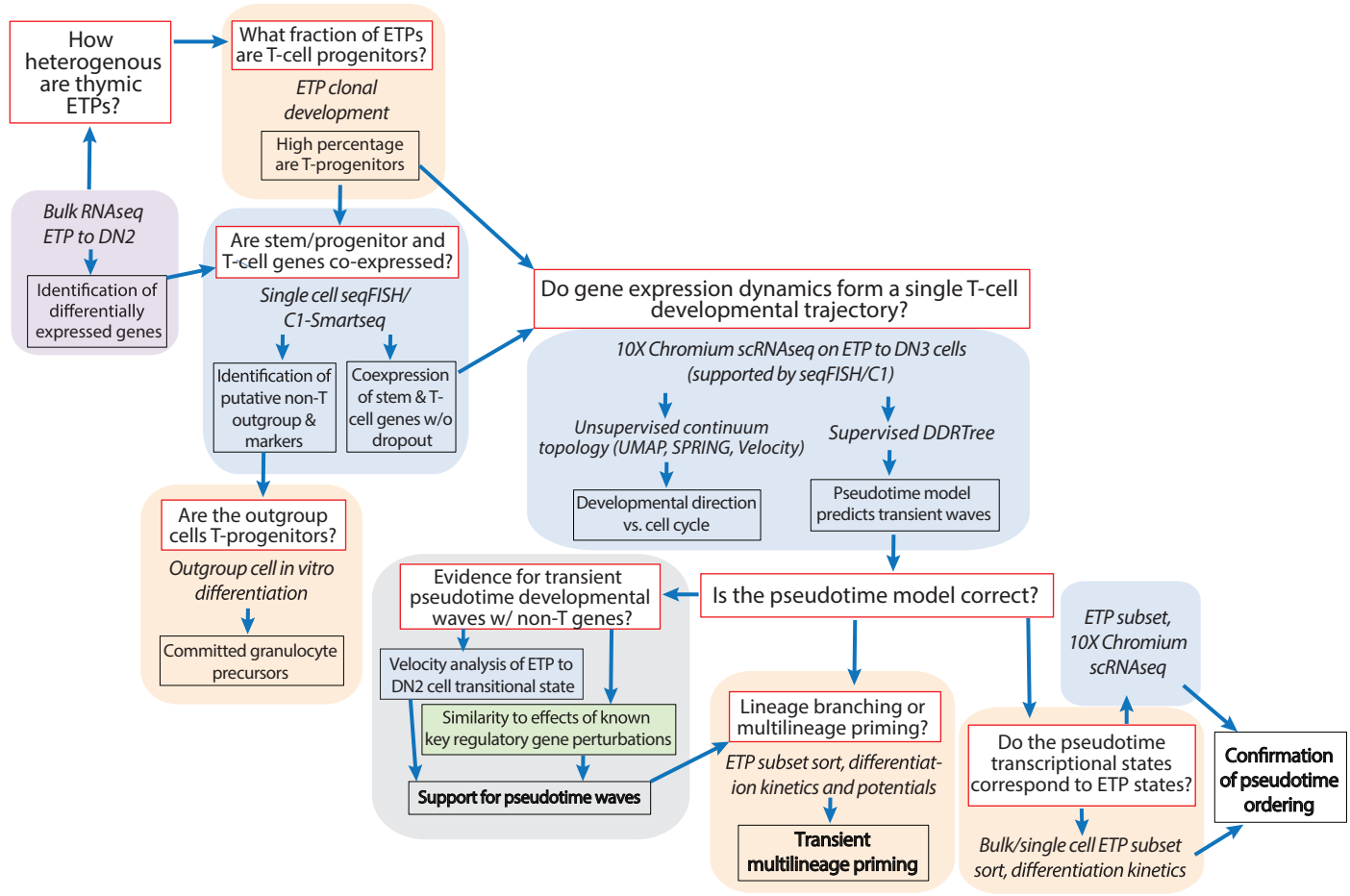
highlight the unspliced fraction of the individual genes, indicating active synthesis of transcripts (red), and apparently decreasing synthesis (blue), on PC1-2 displays as shown in (c). Green (right-hand plots) highlight spliced transcripts on the same axes. e) PC2-3 display with arrows indicating the transition probability of cells as described in the top panel of (c)(and Fig. 5b) plus corresponding gene plots as described in (d).

Supplementary Figure S8, related to Fig. 6. Characterization of ETP subsets based on model of developmental pseudotime: transcriptome pseudotime position relative to directly tested differentiation kinetics

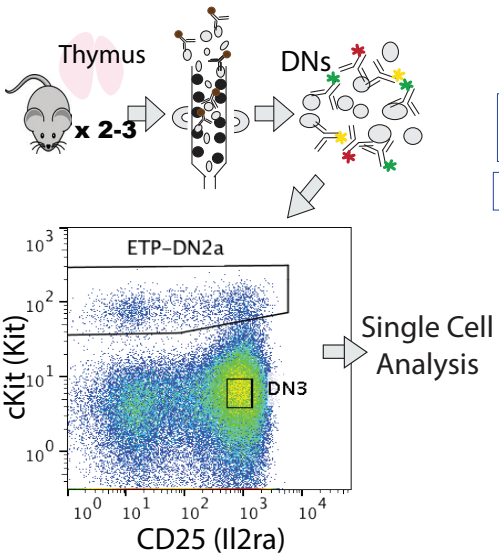
Supervised analysis of 10X Chromium data: low dimensional representation based on the curated, instructive gene list in Table S2. a) PC loading of first 2 PCs with supervised analysis. b) PC (PC1 vs. PC2) and tSNE (tSNE2 vs. tSNE3) displays of 10X data (replicate 1, 4627 cells) with clusters color projected. tSNE and the SLM clustering algorithm were performed based on PC1-6 (same as seqFISH analysis). c) Clustered expression patterns of members of the curated instructive gene list (Table S2) on pseudo-time (same pseudo-time scale and calculations as in Fig. 5; displayed are genes that were detected in $\geq$11 cells). Colored by log transformed and row normalized relative expression level. d) FACS gating strategy for ETP sub-population sorts in Fig. 6, both on population level and single cell level. e) Precursors in individual gates shown in (d) were profiled for transcriptome expression and pseudotime prediction using Cell Hashing. Top panel labels the mean and interquartile ranges of individual-cell pseudotime predictions from each subpopulation. Bottom heatmap displays the key genes' expression pattern on this recalculated, 'ETP-enriched' pseudotime scale, aligned to the top panel. Note enrichment of *Tyrobp, Mpo, Cd7, Spi1,* and *Hhex* expression corresponding to the sorted ETP subsets 4 and 6.

Fig.S1

a


How heterogenous are thymic ETPs?
What fraction of ETPs are T-cell progenitors?
ETP clonal development
High percentage are T-progenitors
Bulk RNAseq ETP to DN2
Identification of differentially expressed genes
Are stem/progenitor and T-cell genes co-expressed?
Single cell seqFISH/ C1-Smartseq
Identification of putative non-T outgroup & markers
Coexpression of stem & T-cell genes w/o dropout
Do gene expression dynamics form a single T-cell developmental trajectory?
10X Chromium scRNAseq on ETP to DN3 cells (supported by seqFISH/C1)
Unsupervised continuum topology (UMAP, SPRING, Velocity)
Supervised DDRTree
Developmental direction vs. cell cycle
Pseudotime model predicts transient waves
Are the outgroup cells T-progenitors?
Outgroup cell in vitro differentiation
Committed granulocyte precursors
Evidence for transient pseudotime developmental waves w/ non-T genes?
Is the pseudotime model correct?
ETP subset, 10X Chromium scRNAseq
Velocity analysis of ETP to DN2 cell transitional state
Similarity to effects of known key regulatory gene perturbations
Support for pseudotime waves
Lineage branching or multilineage priming?
ETP subset sort, differentiation kinetics and potentials
Transient multilineage priming
Do the pseudotime transcriptional states correspond to ETP states?
Bulk/single cell ETP subset sort, differentiation kinetics
Confirmation of pseudotime ordering

c


Thymus
x 2-3
DNs
ETP–DN2a
DN3
cKit (Kit)
CD25 (Il2ra)
Single Cell Analysis

b

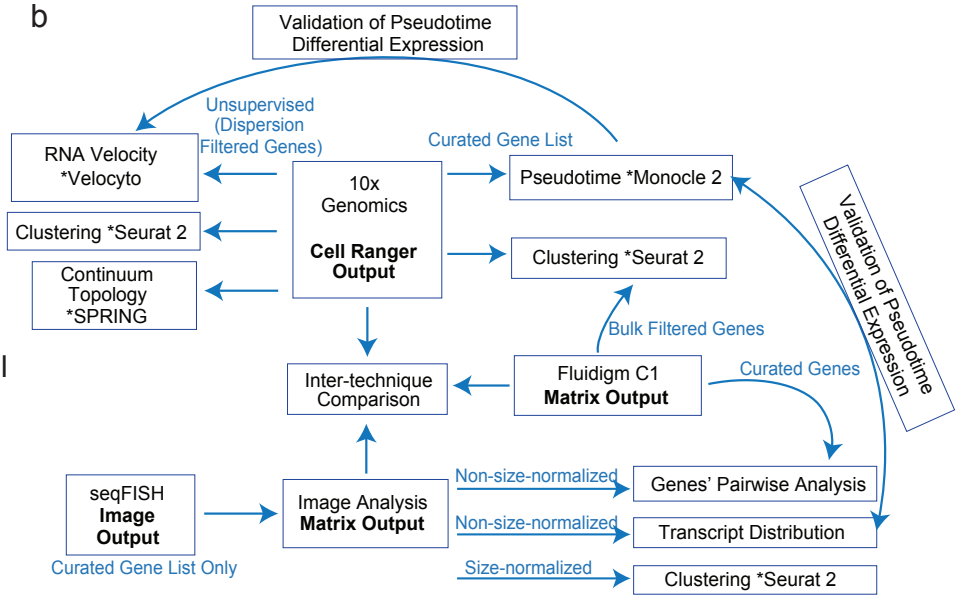

Validation of Pseudotime Differential Expression
Unsupervised (Dispersion Filtered Genes)
Curated Gene List
RNA Velocity *Velocyto
Clustering *Seurat 2
Continuum Topology *SPRING
10x Genomics
Cell Ranger Output
Pseudotime *Monocle 2
Clustering *Seurat 2
Bulk Filtered Genes
Inter-technique Comparison
Fluidigm C1 Matrix Output
Curated Genes
Validation of Pseudotime Differential Expression
seqFISH Image Output
Curated Gene List Only
Image Analysis Matrix Output
Non-size-normalized
Genes' Pairwise Analysis
Non-size-normalized
Transcript Distribution
Size-normalized
Clustering *Seurat 2

Fig. S2

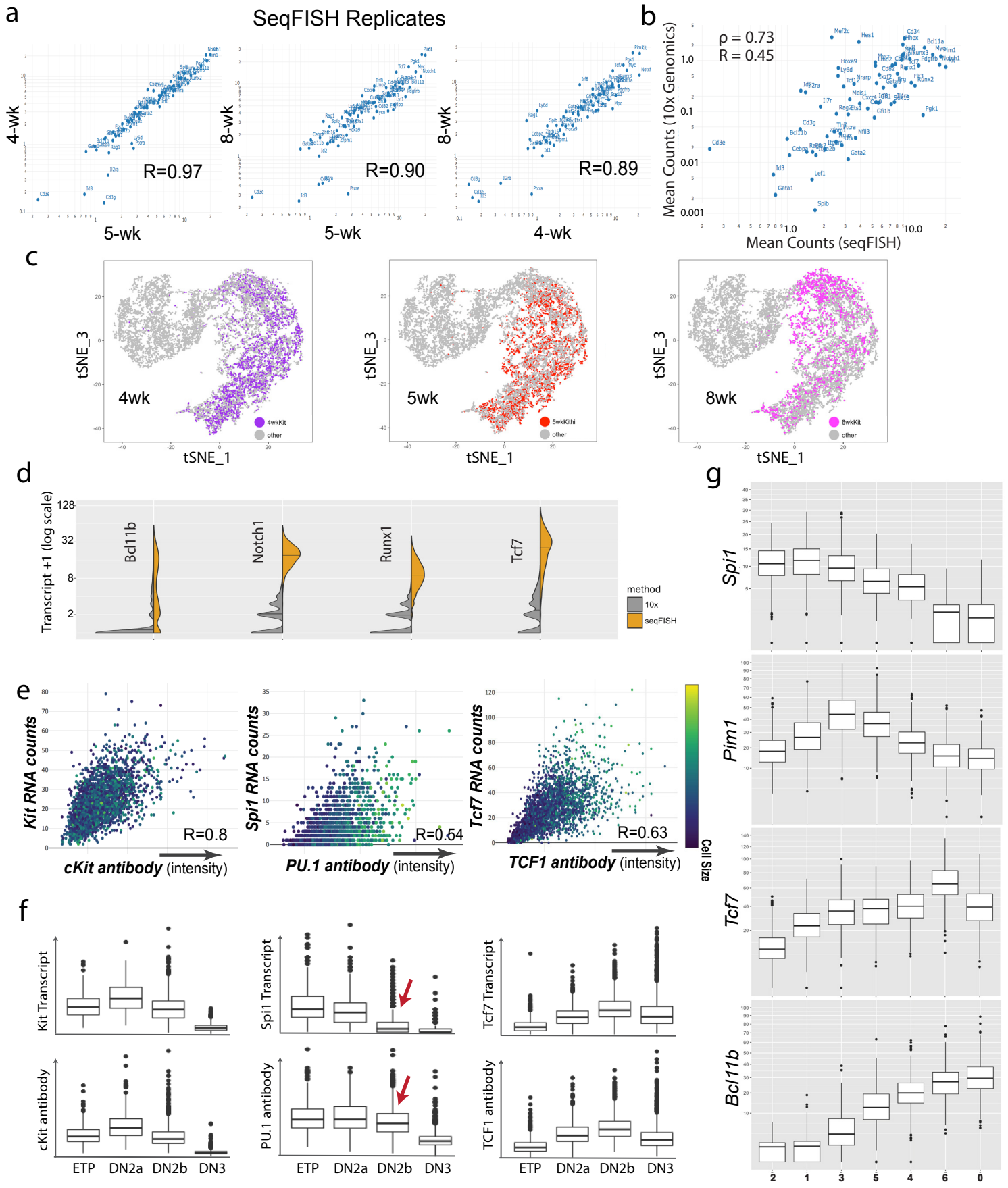

a
SeqFISH Replicates
4-wk
5-wk
R=0.97
8-wk
5-wk
R=0.90
8-wk
4-wk
R=0.89
b
ρ = 0.73
R = 0.45
Mean Counts (10x Genomics)
Mean Counts (seqFISH)
c
tSNE_3
tSNE_1
4wk
4wkKit
other
5wk
5wkKithi
other
8wk
8wkKit
other
d
Transcript +1 (log scale)
Bcl11b
Notch1
Runx1
Tcf7
method
10x
seqFISH
e
Kit RNA counts
R=0.8
cKit antibody (intensity)
Spi1 RNA counts
R=0.54
PU.1 antibody (intensity)
Tcf7 RNA counts
R=0.63
TCF1 antibody (intensity)
Cell Size
f
Kit Transcript
cKit antibody
Spi1 Transcript
PU.1 antibody
Tcf7 Transcript
TCF1 antibody
ETP
DN2a
DN2b
DN3
g
Spi1
Pim1
Tcf7
Bcl11b
2
1
3
5
4
6
0

Fig. S3

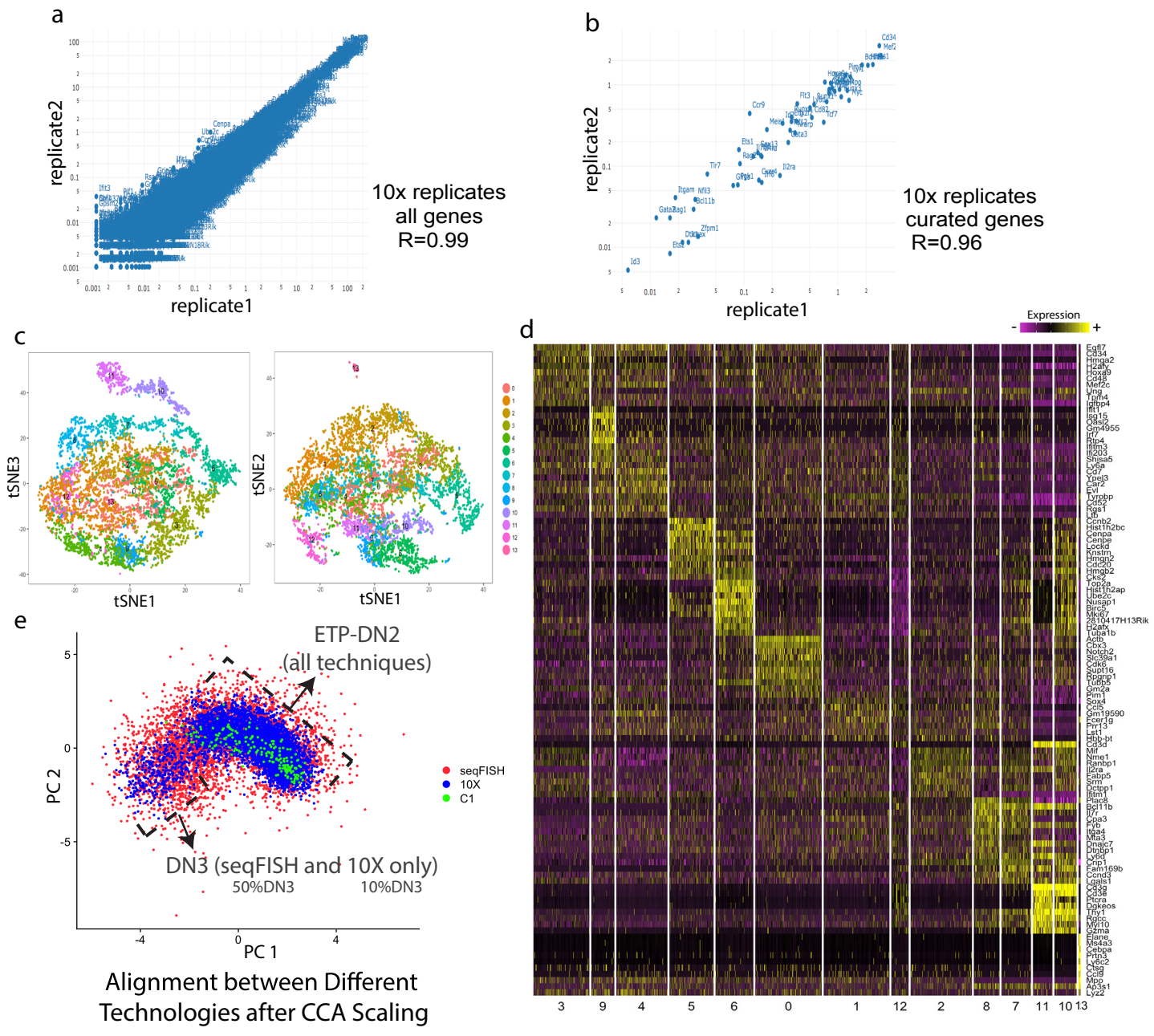
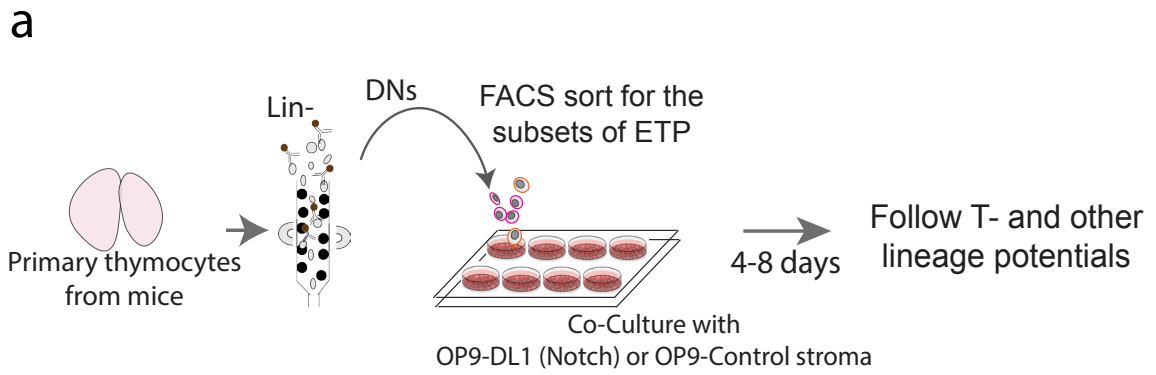
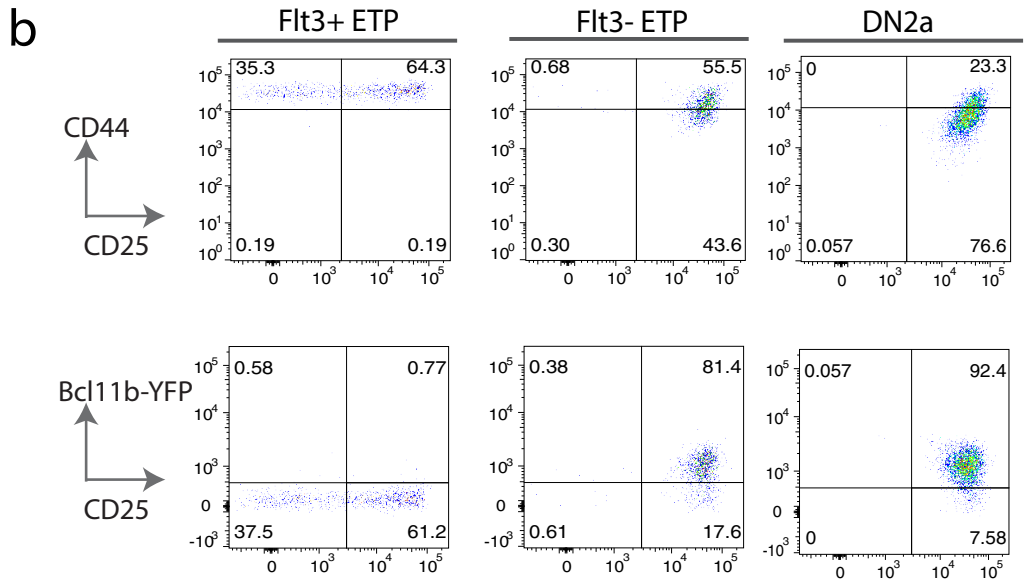
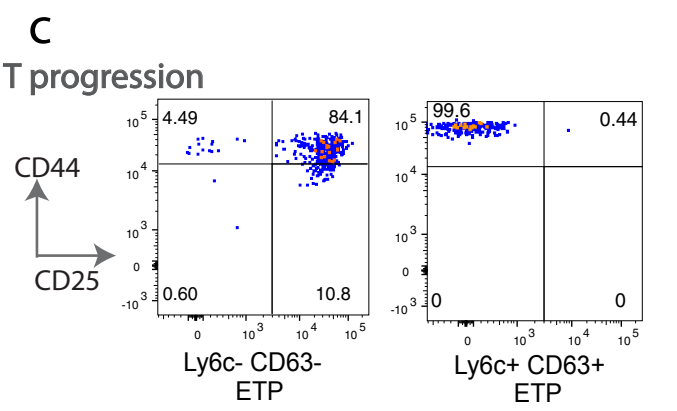

a
replicate2
replicate1
10x replicates
all genes
R=0.99
b
replicate2
replicate1
10x replicates
curated genes
R=0.96
c
tSNE3
tSNE1
tSNE2
tSNE1
d
Expression
e
ETP-DN2
(all techniques)
seqFISH
10X
C1
DN3 (seqFISH and 10X only)
50%DN3
10%DN3
PC 2
PC 1
Alignment between Different
Technologies after CCA Scaling

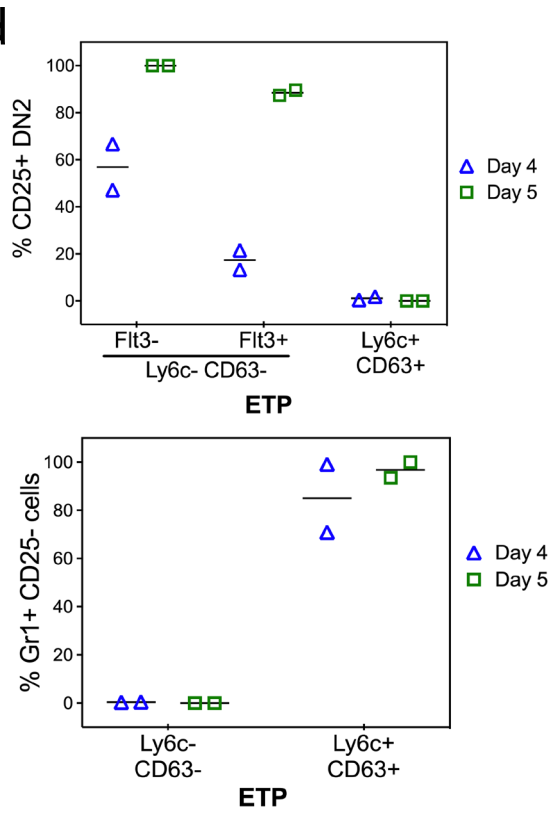
Fig S4



With Notch Signaling







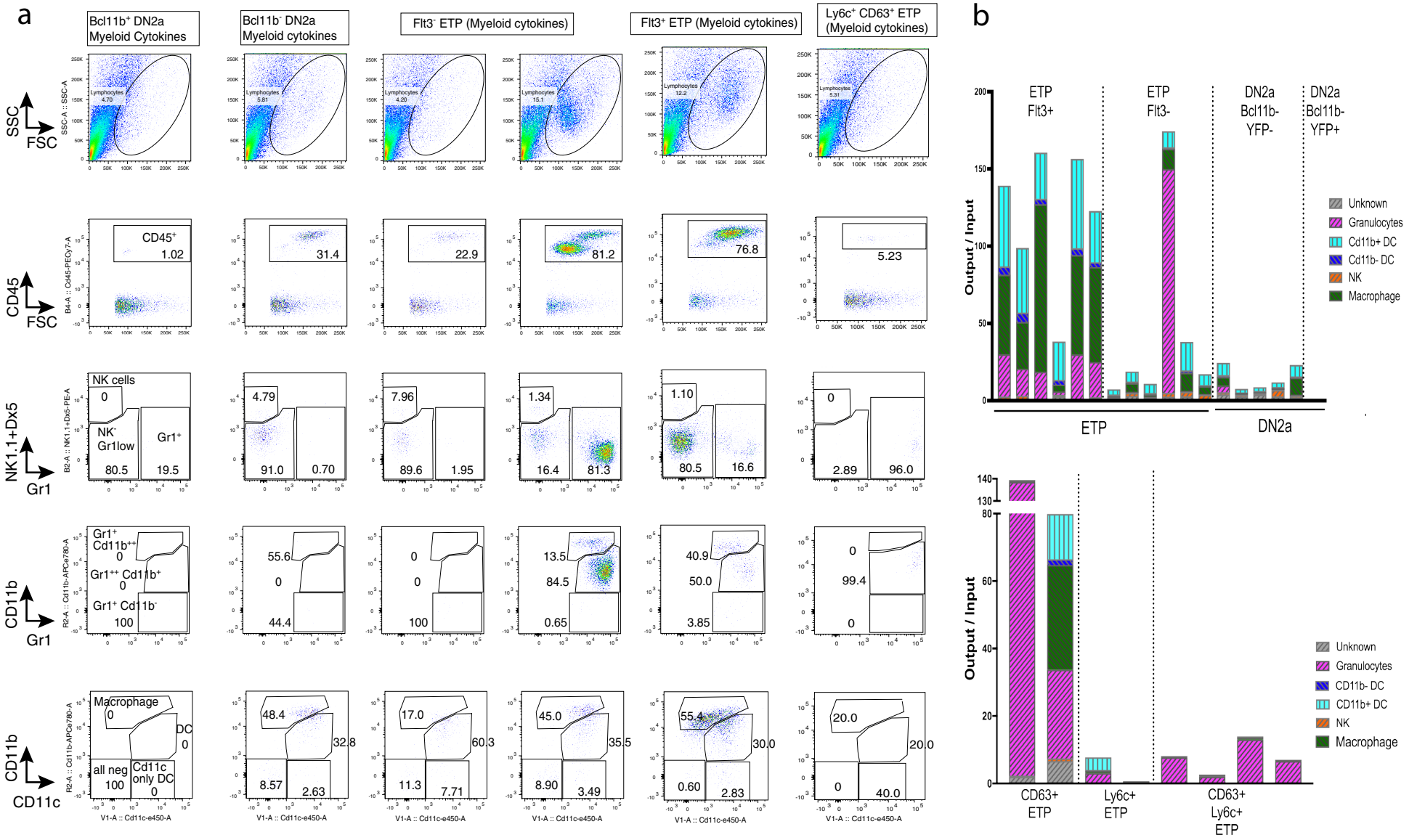

a
Bcl11b+ DN2a Myeloid Cytokines
Bcl11b- DN2a Myeloid cytokines
Flt3- ETP (Myeloid cytokines)
Flt3+ ETP (Myeloid cytokines)
Ly6c+ CD63+ ETP (Myeloid cytokines)
SSC
FSC
CD45
NK1.1+Dx5
Gr1
CD11b
CD11c
Lymphocytes
CD45+ 1.02
31.4
22.9
81.2
76.8
5.23
NK cells
NK- Gr1low
Gr1+
80.5
19.5
4.79
91.0
0.70
7.96
89.6
1.95
1.34
16.4
81.3
1.10
80.5
16.6
0
2.89
96.0
Gr1+ Cd11b++
Gr1++ Cd11b+
Gr1+ Cd11b-
55.6
44.4
13.5
84.5
0.65
40.9
50.0
3.85
99.4
Macrophage
DC
all neg
Cd11c only DC
48.4
32.8
8.57
2.63
17.0
60.3
11.3
7.71
45.0
35.5
8.90
3.49
55.4
30.0
0.60
2.83
20.0
40.0
V1-A :: Cd11c-e450-A
b
ETP Flt3+
ETP Flt3-
DN2a Bcl11b- YFP-
DN2a Bcl11b- YFP+
Output / Input
Unknown
Granulocytes
Cd11b+ DC
Cd11b- DC
NK
Macrophage
ETP
DN2a
CD63+ ETP
Ly6c+ ETP
CD63+ Ly6c+ ETP

Fig. S5

Fig. S6

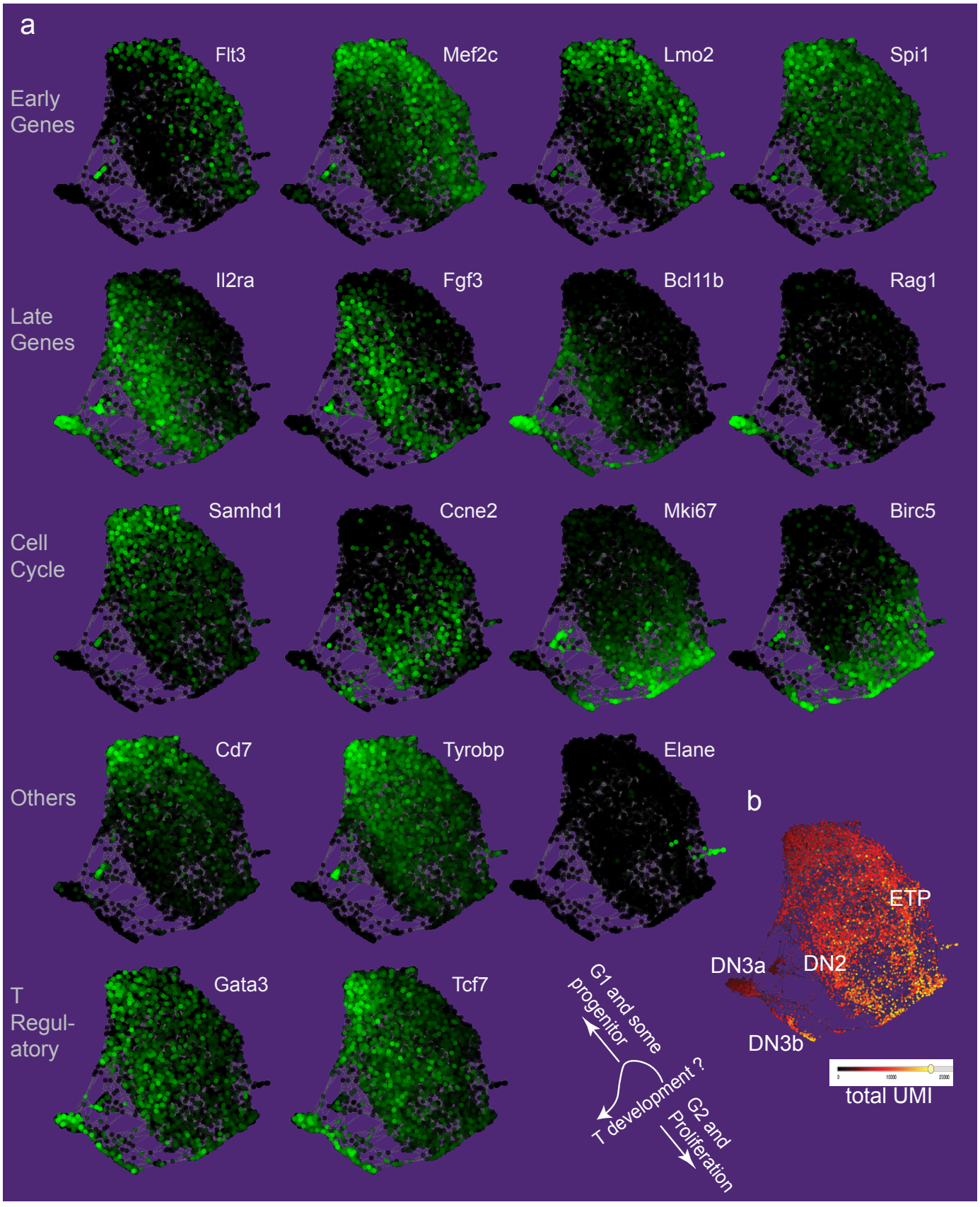

a
Early Genes
Flt3
Mef2c
Lmo2
Spi1
Late Genes
Il2ra
Fgf3
Bcl11b
Rag1
Cell Cycle
Samhd1
Ccne2
Mki67
Birc5
Others
Cd7
Tyrobp
Elane
T Regul-atory
Gata3
Tcf7
G1 and some progenitor
T development ?
G2 and Proliferation
b
ETP
DN3a
DN2
DN3b
total UMI

Fig. S7

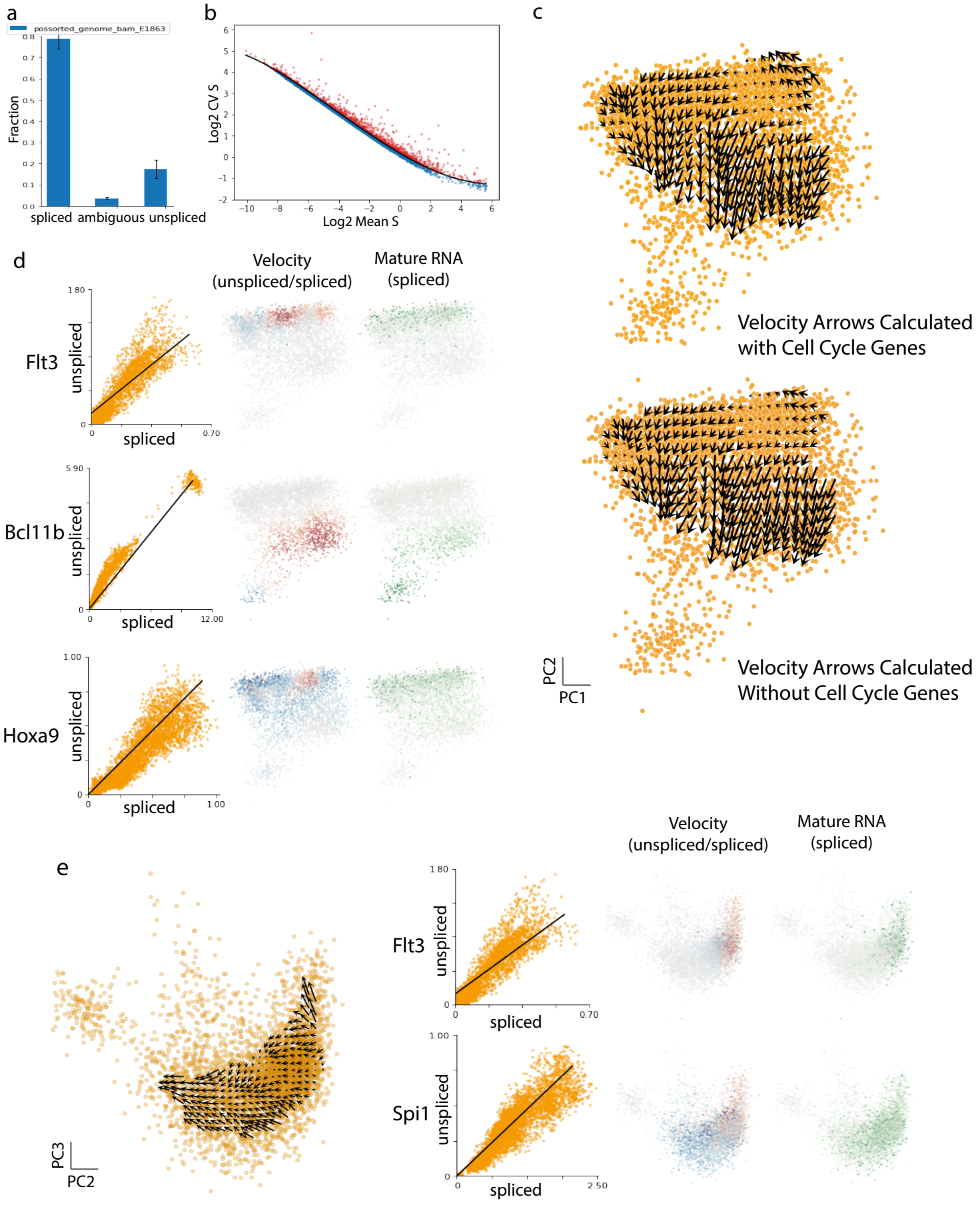
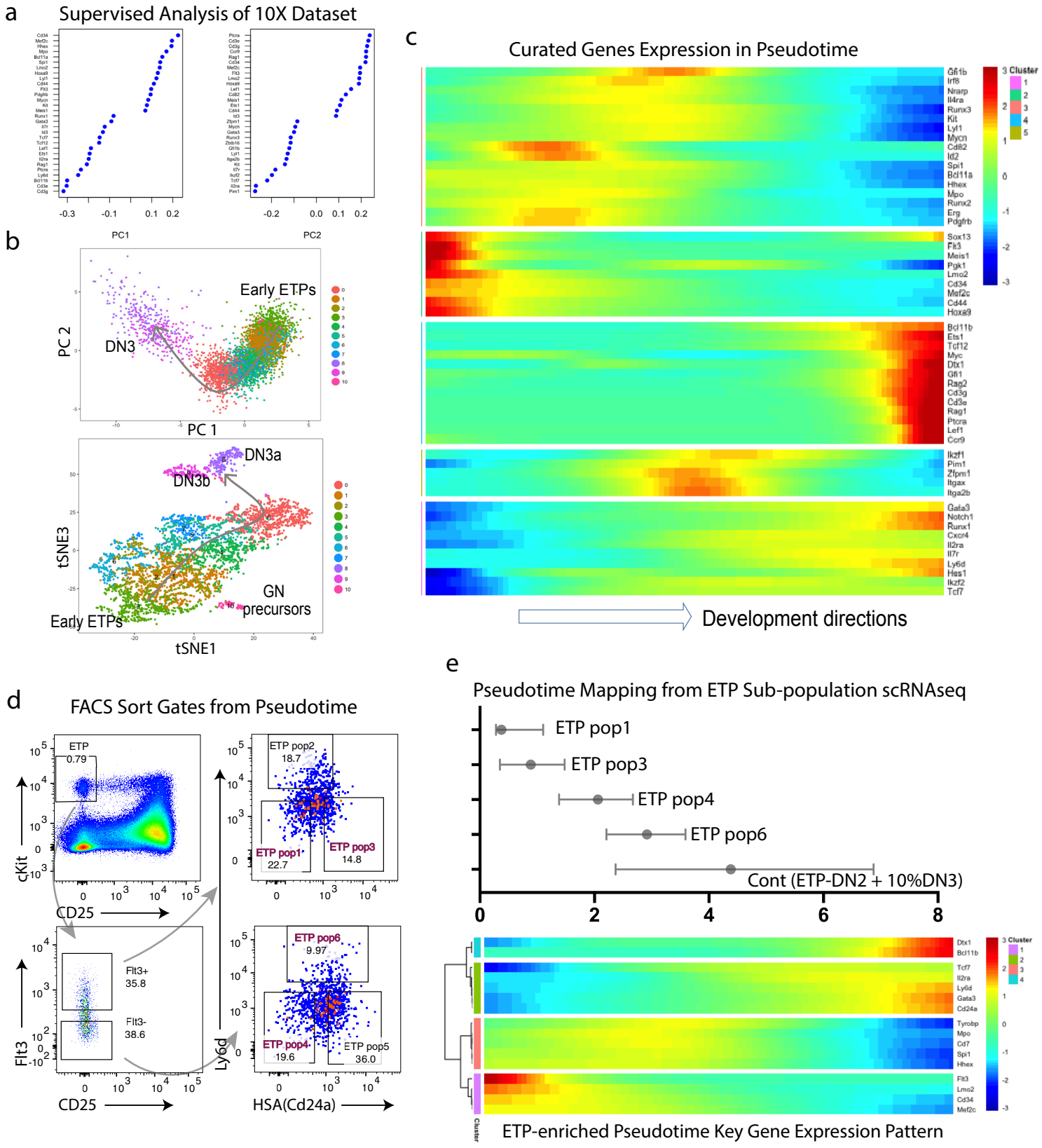

a
possorted_genome_bam_E1863
Fraction
spliced
ambiguous
unspliced
b
Log2 CV S
Log2 Mean S
c
Velocity Arrows Calculated with Cell Cycle Genes
Velocity Arrows Calculated Without Cell Cycle Genes
PC2
PC1
d
Velocity (unspliced/spliced)
Mature RNA (spliced)
Flt3
Bcl11b
Hoxa9
unspliced
spliced
e
PC3
PC2
Velocity (unspliced/spliced)
Mature RNA (spliced)
Flt3
Spi1
unspliced
spliced

Fig. S8


a
Supervised Analysis of 10X Dataset
b
PC1
PC2
Early ETPs
DN3
PC 2
PC 1
DN3a
DN3b
GN precursors
Early ETPs
tSNE3
tSNE1
c
Curated Genes Expression in Pseudotime
3 Cluster
Development directions
d
FACS Sort Gates from Pseudotime
ETP 0.79
ETP pop2 18.7
ETP pop1 22.7
ETP pop3 14.8
cKit
CD25
Flt3+ 35.8
Flt3- 38.6
Flt3
ETP pop6 9.97
ETP pop4 19.6
ETP pop5 36.0
Ly6d
HSA(Cd24a)
e
Pseudotime Mapping from ETP Sub-population scRNAseq
ETP pop1
ETP pop3
ETP pop4
ETP pop6
Cont (ETP-DN2 + 10%DN3)
ETP-enriched Pseudotime Key Gene Expression Pattern