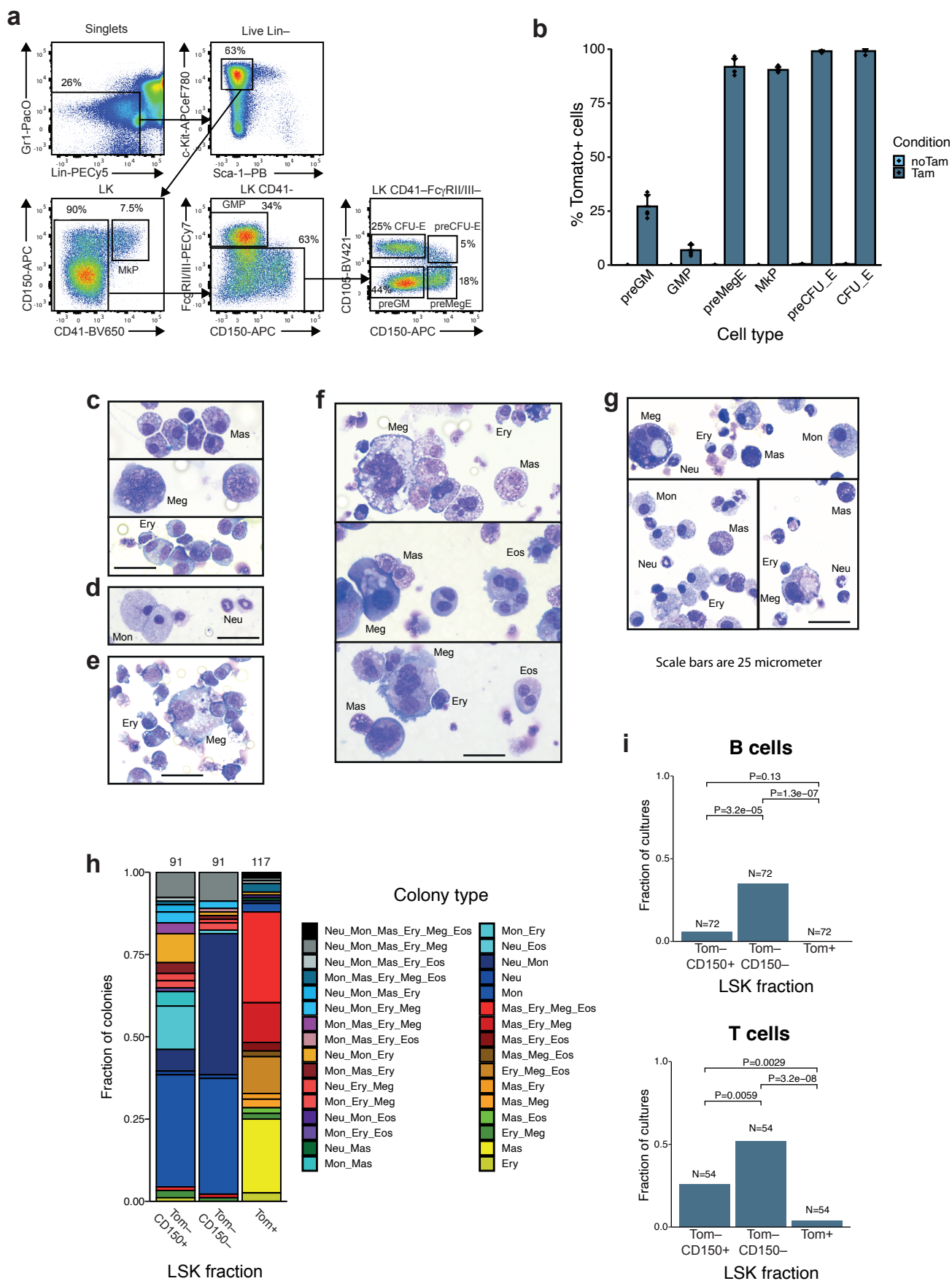
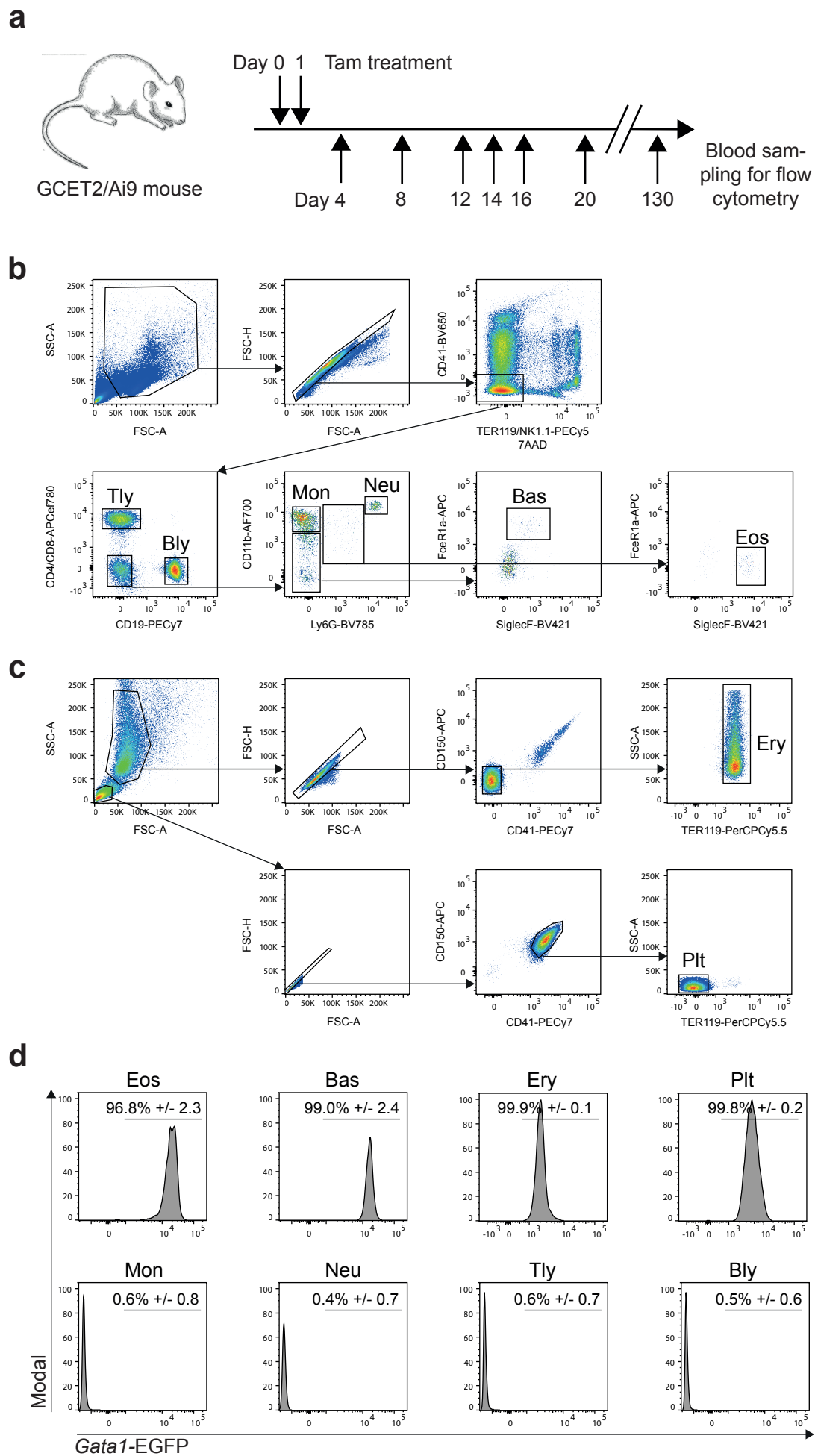




Extended Data Figure 1

Extended Data Fig. 1 | *Gata1* lineage tracing of progenitor populations.

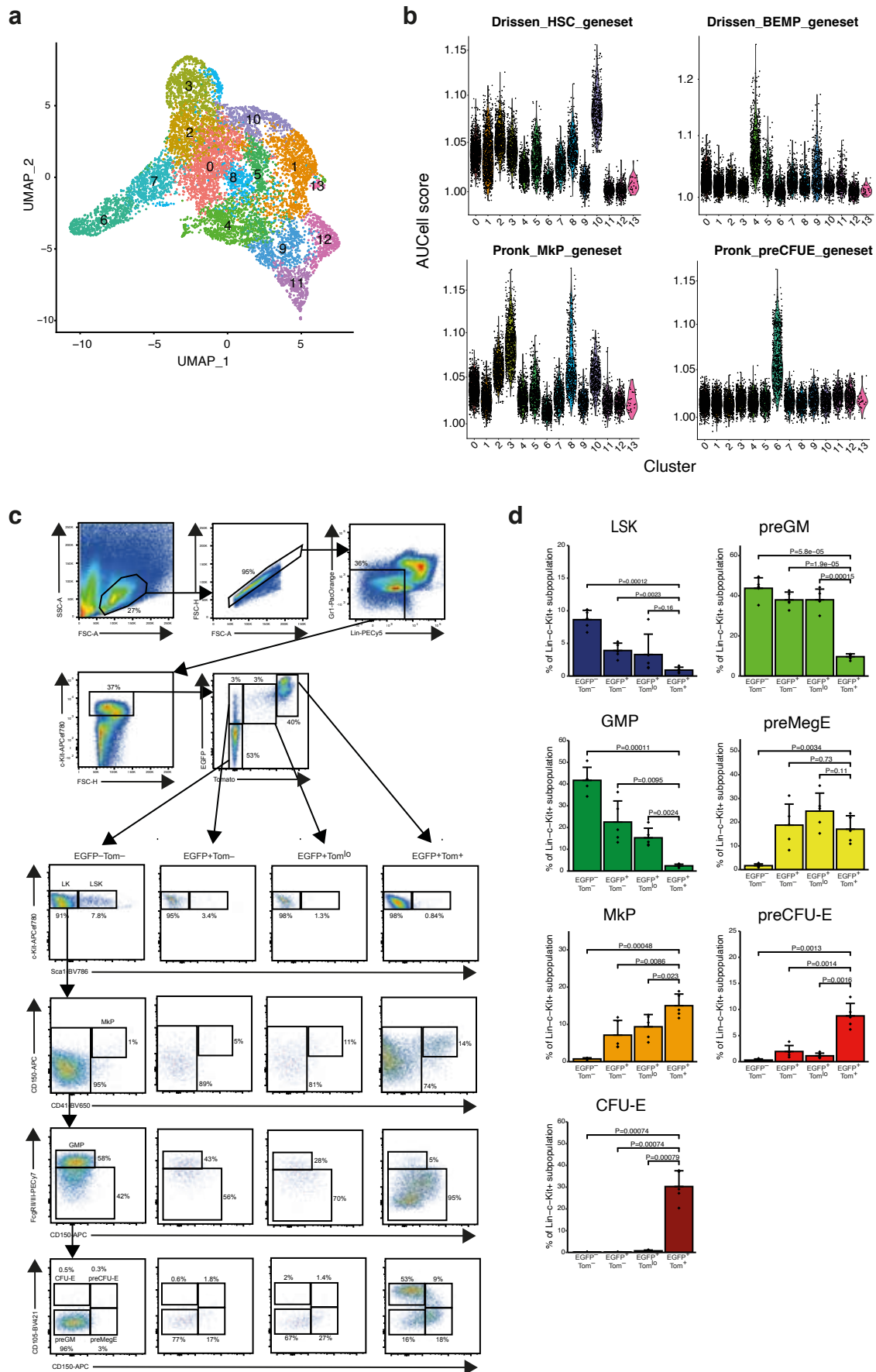
- a)** Gating strategy for identification of myelo-erythroid progenitor populations.
- b)** Bone marrow from untreated GCET2/Ai9/GE mice and GCET2/Ai9/GE mice treated with tamoxifen as shown in Fig.1b was analysed as in (a). Graph shows the mean percentage of Tom+ cells for the indicated progenitor populations. Error bars: SEM. N=5 biological replicates in 2 independent experiments.
- c)** Representative examples of uni-potent Mas, Meg and Ery colonies derived from a single Lin-c-Kit+ cell.
- d)** Representative examples of a bi-potent Neu-Mon colony derived from a single Lin-c-Kit+ cell.
- e)** Representative example of a bi-potent Meg-Ery colony derived from a single Lin-c-Kit+ cell.
- f)** Representative example of an oligo-potent Mas-Eos-Meg-Ery colonies derived from a single Lin-c-Kit+ cell.
- g)** Representative examples of multi-potent colonies derived from a single Lin-c-Kit+ cell.
- h)** Colony types generated by LSK sub-populations from Fig.1e. Data from 5 independent experiments. The total number of colonies analysed for each LSK fraction is shown above the columns.
- i)** Single cells from the indicated LSK subpopulations were cultured under T or B supportive conditions. Graph shows fraction of cultures that produced T or B cells. Total number of cultures are shown. P-values show the significance of the difference in frequency of positive cultures (Chi-square test). Data are from 2 independent experiments.



Extended Data Figure 2

Extended Data Fig. 2 | Peripheral blood analysis of *Gata1* lineage traced mice.

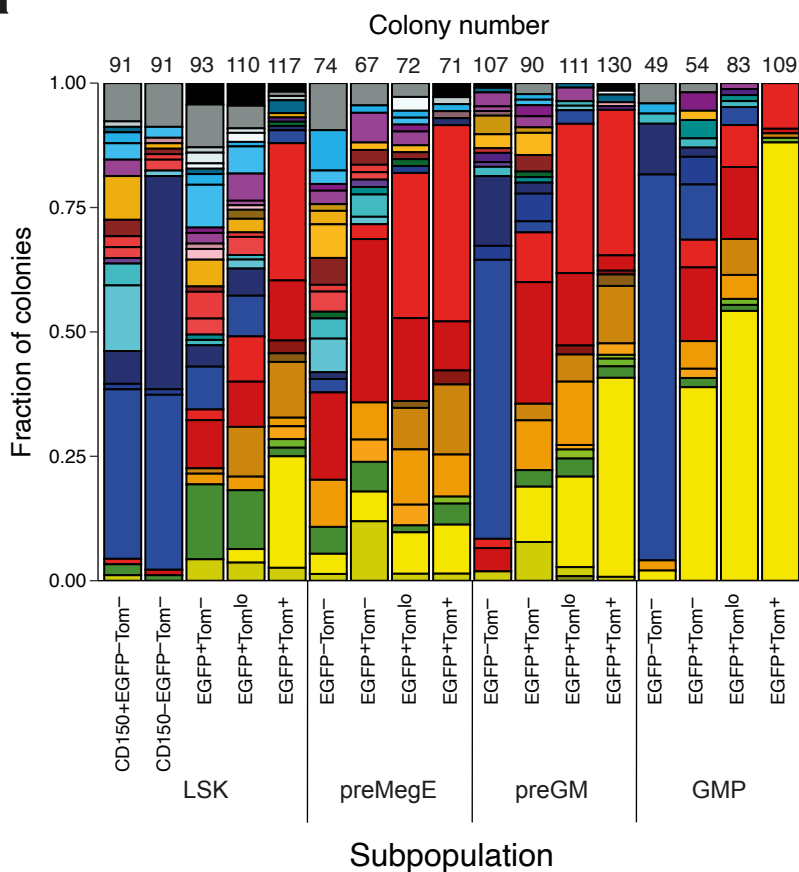
- a)** Workflow of tamoxifen treatment and peripheral blood analysis for *Gata1* lineage tracing of peripheral blood cell types.
- b)** Gating strategy for identification of leukocytes from peripheral blood. Tly: T-lymphocytes; Bly: B-lymphocytes; Mon: monocytes; Neu: neutrophils; Bas: basophils; Eos: eosinophils.
- c)** Gating strategy for identification of erythrocytes (Ery) and platelets (Plt) from peripheral blood.
- d)** Histograms showing the *Gata1*-EGFP signal in peripheral blood cell types from GCET2/Ai9/GE mice. Numbers show percentage EGFP⁺ cells as mean $\pm$ SEM. N=15 from 3 mice analysed at 5 timepoints.



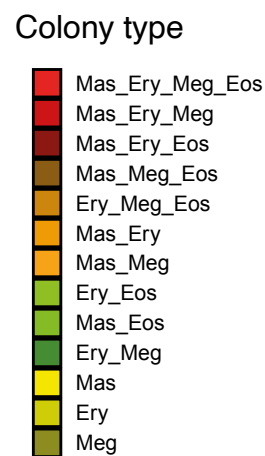
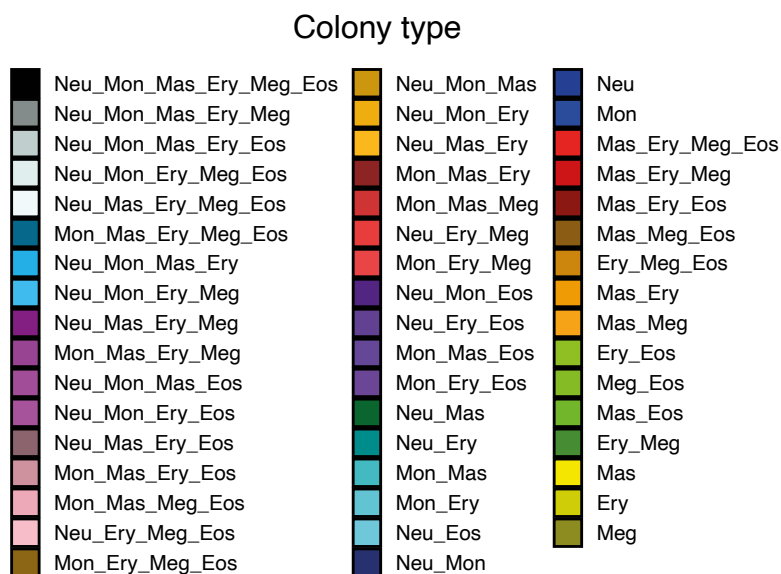
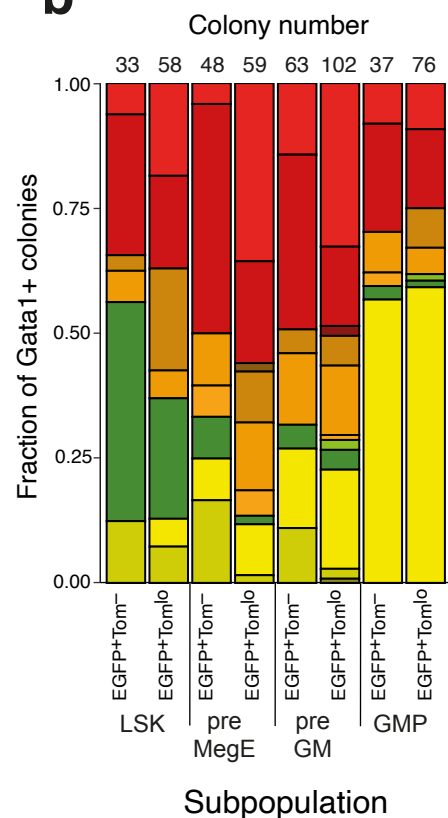
Extended Data Fig. 3 | Differentiation trajectories in LSK and LK cell populations.

- a)** UMAP projection of clustered single cell transcriptomes of populations from Fig.2b.
- b)** Violinplot showing AUCell scores of the indicated signatures in single cell clusters from (a)
- c)** Gating strategy for the analysis of immunophenotypic progenitor populations in Lin–c-Kit⁺ subsets defined by EGFP and Tom expression. Numbers show the gated population as a percentage of the parental gate, and the analysis is representative of 5 samples from 4 independent experiments.
- d)** Quantification of the progenitor populations defined in (c) within each Lin–c-Kit⁺ subset. The abundance of progenitors is shown as a percentage of the respective Lin–c-Kit⁺ subsets. N=5, 4 independent experiments. Error bars: SEM. P-values show the significance of the difference in abundance between the EGFP⁺Tom⁺ and the other Lin–c-Kit⁺ subsets (Student's t-test).

a



b

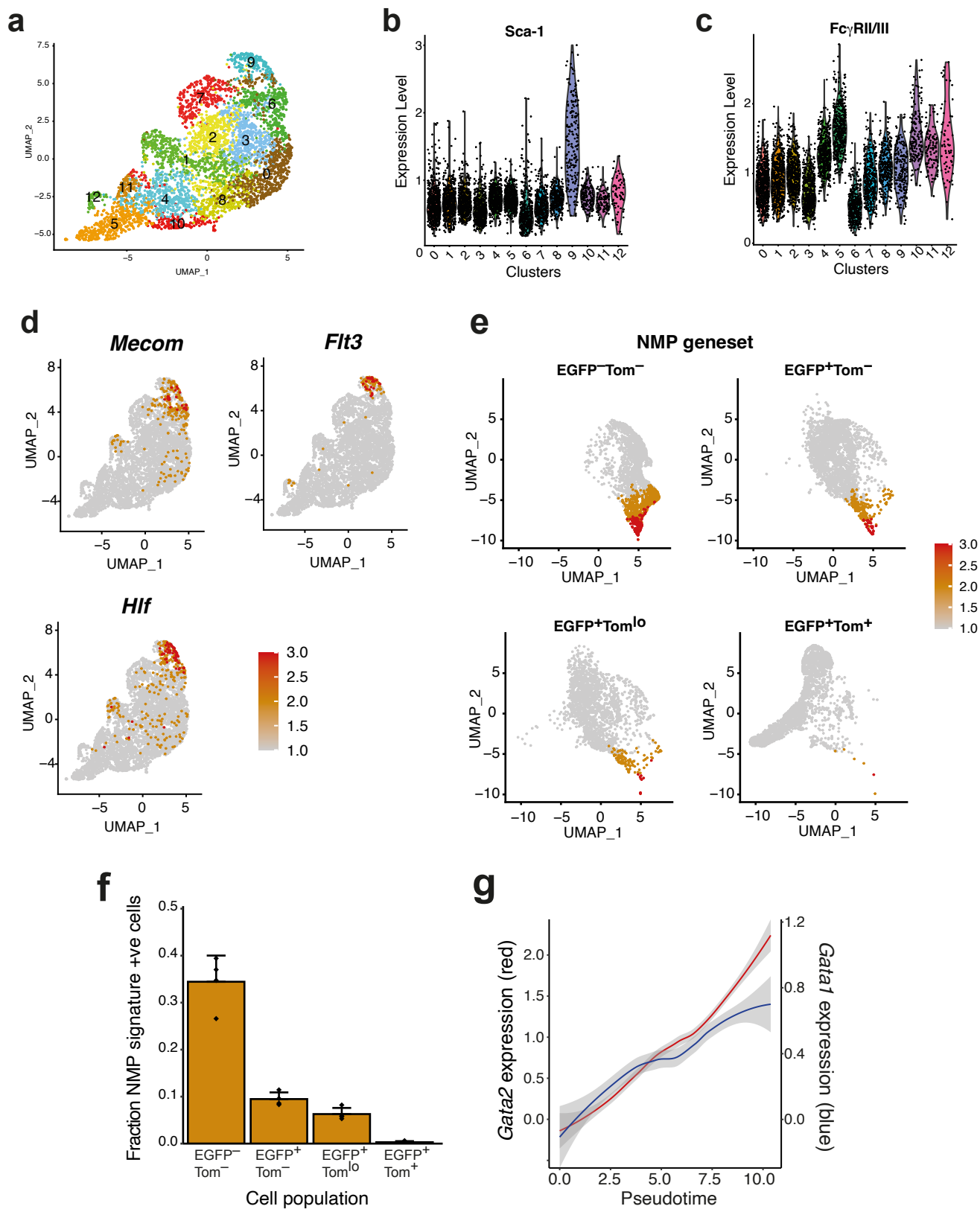


Extended Data Figure 4

Extended Data Fig. 4 | Single cell lineage potential.

a) Cell types detected in cytopins of single cell colonies from the indicated progenitor populations. Number of analysed colonies for each population is shown on top. Data from each population is from 2 or more independent experiments.

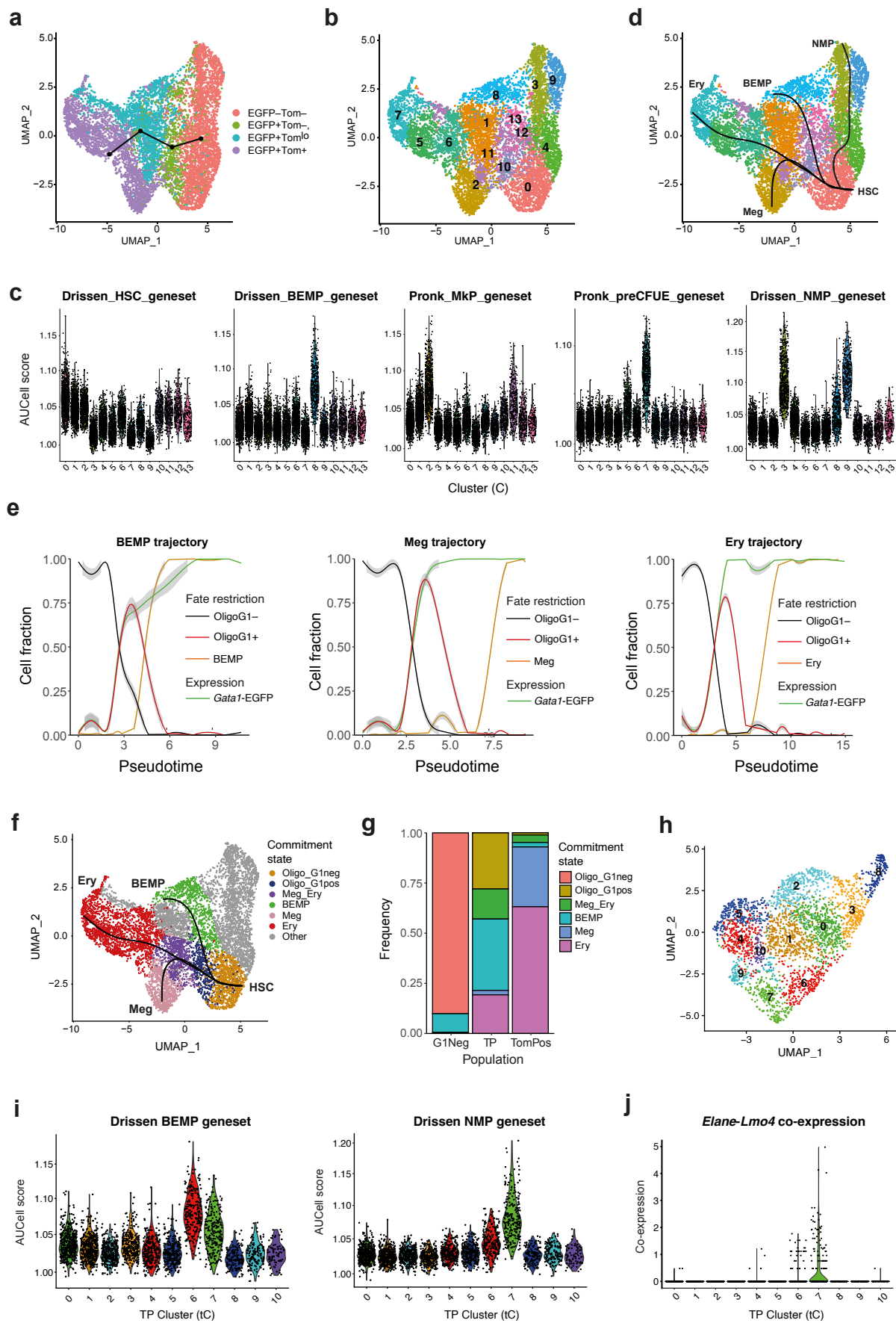
b) Data from (a), showing only *Gata1*⁺ lineage readouts from transitional progenitor populations.



Extended Data Figure 5

Extended Data Fig. 5 | Clustering and analysis of TP CITE-seq data.

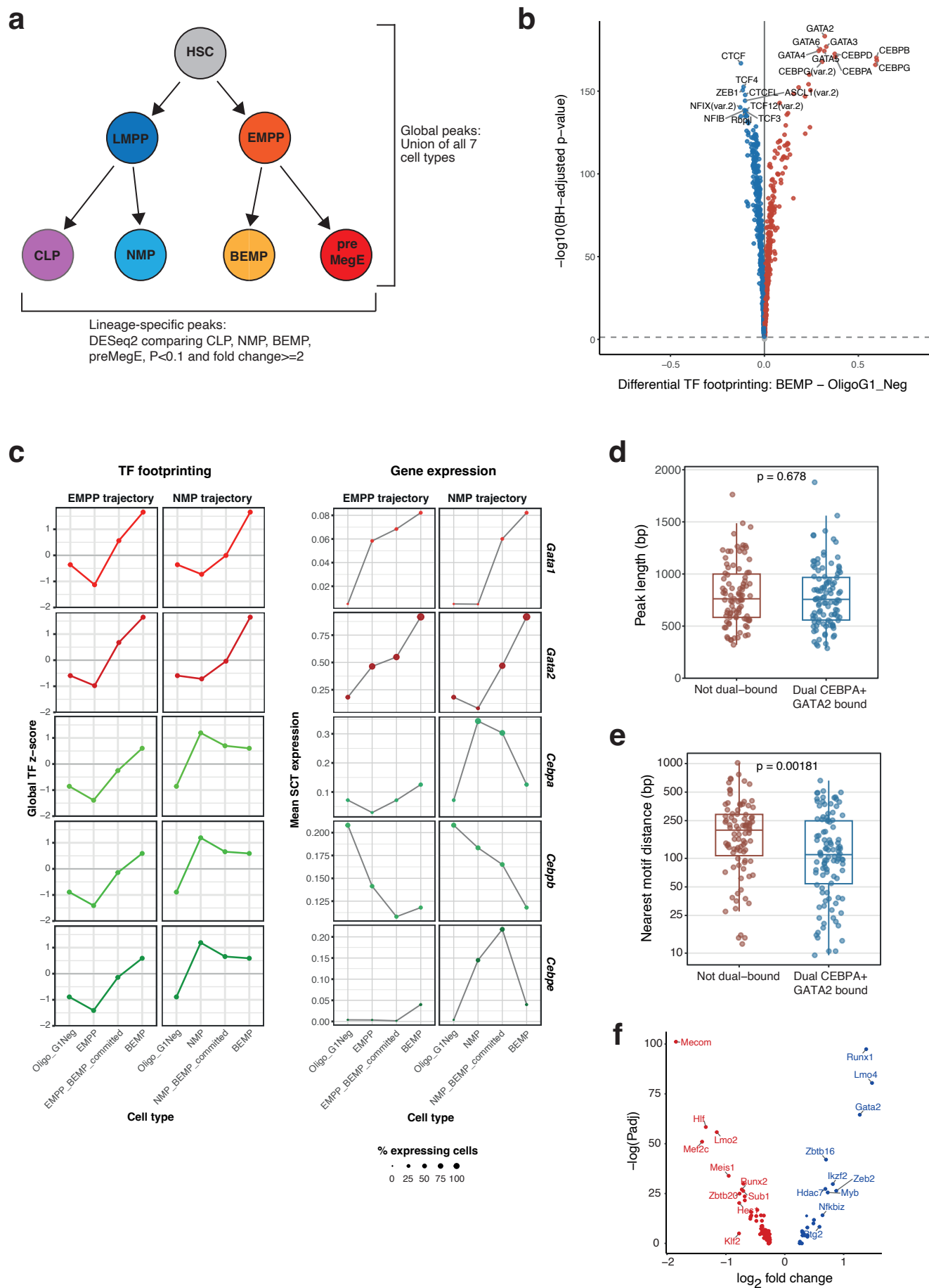
- a)** UMAP projection showing clustering of CITE-seq TP single cell transcriptomes. Cluster numbers are shown.
- b)** Violinplot showing the expression of the Sca-1 surface marker in clusters from using CITEseq (a)
- c)** Violinplot showing the expression of the FcγRII/III surface marker in clusters from using CITEseq (a)
- d)** Feature plots showing mRNA expression of *Mecom*, *Hlf* and *Flt3* on UMAP projection of TP single cell transcriptomes. Scale bar shows the normalised expression value.
- e)** Feature plots showing expression of the NMP geneset AUCell score on UMAP projection of the indicated Lin[−]c-Kit⁺ subpopulations. Scale bar shows the normalised expression value.
- f)** Fraction of NMP signature positive cells from (e) in the indicated Lin[−]c-Kit⁺ subpopulations. Cells were quantified in each of the 4 biological replicates, and the mean value is shown. Error bars: SEM.
- g)** Expression of *Gata1* (blue line) and *Gata2* (red line) across pseudotime of the C5-C10 trajectory from Fig.4e. Grey shaded areas show 95% confidence intervals.



Extended Data Figure 6

Extended Data Fig. 6 | Clustering of single cell multiomes and trajectory analysis.

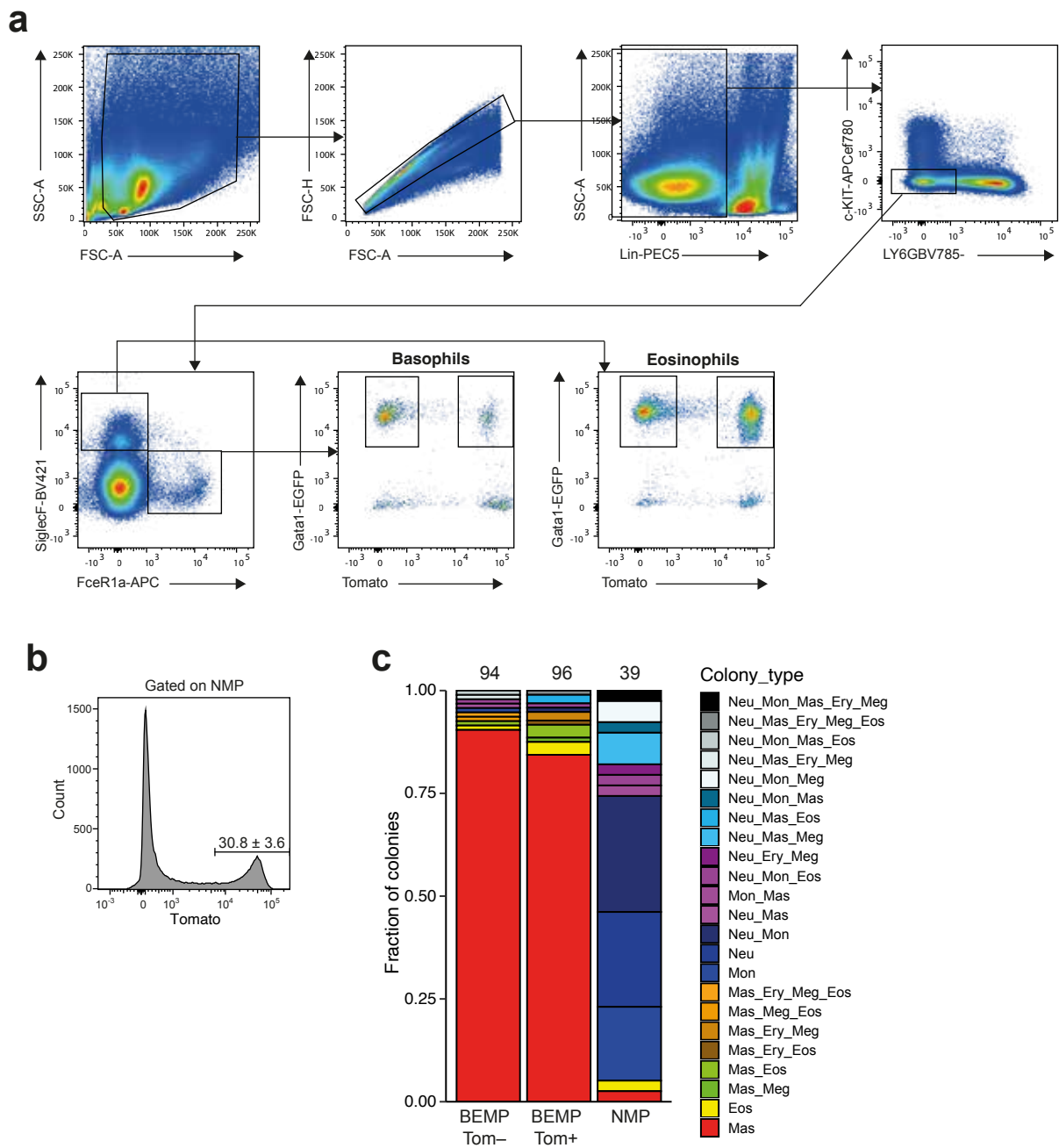
- a)** UMAP projection of single cell multiomes from the indicated Lin⁺c-Kit⁺ subpopulations, and the slingshot trajectory connecting the centroids of the 4 populations, generated with the EGFP⁺Tom⁺ population as the starting point.
- b)** UMAP projection of single cell multiomes from (a) clustered using weighted nearest neighbours. Cluster numbers are shown.
- c)** Violinplot showing AUCell scores of the indicated signatures in single cell clusters from (b)
- d)** UMAP projection of single cell transcriptomes from (b) showing clustering and Slingshot trajectories connecting the HSC cluster (C0) to the BEMP (C8), NMP (C3), Ery (C7) and Meg (C2) clusters.
- e)** Inferred potencies from (d) were plotted in pseudotime along the indicated trajectories, along with the proportion of cells expressing *Gata1*-EGFP.
- f)** UMAP plot showing the potency of single cells inferred from association with the Ery, Meg and BEMP trajectories in (d). Potencies were defined as in Fig.2g.
- g)** The fraction of cells with the indicated inferred potencies from (f) within the EGFP⁺Tom⁺ (G1neg), TP and EGFP⁺Tom⁺ (TomPos) LK subpopulations.
- h)** Clustering of TPs (LK EGFP⁺Tom⁺ and LK EGFP⁺Tom^{lo} cells) from the multiome analysis in a).
- i)** Single cell AUCell scores of the clusters from (h) using the indicated signatures.
- j)** Co-expression of *Elane* and *Lmo4* in single cells in clusters from (h). Co-expression was calculated as the product of the SCT expression values for the two genes.



Extended Data Figure 7

Extended Data Fig. 7| TF footprinting and gene expression across myeloid trajectories.

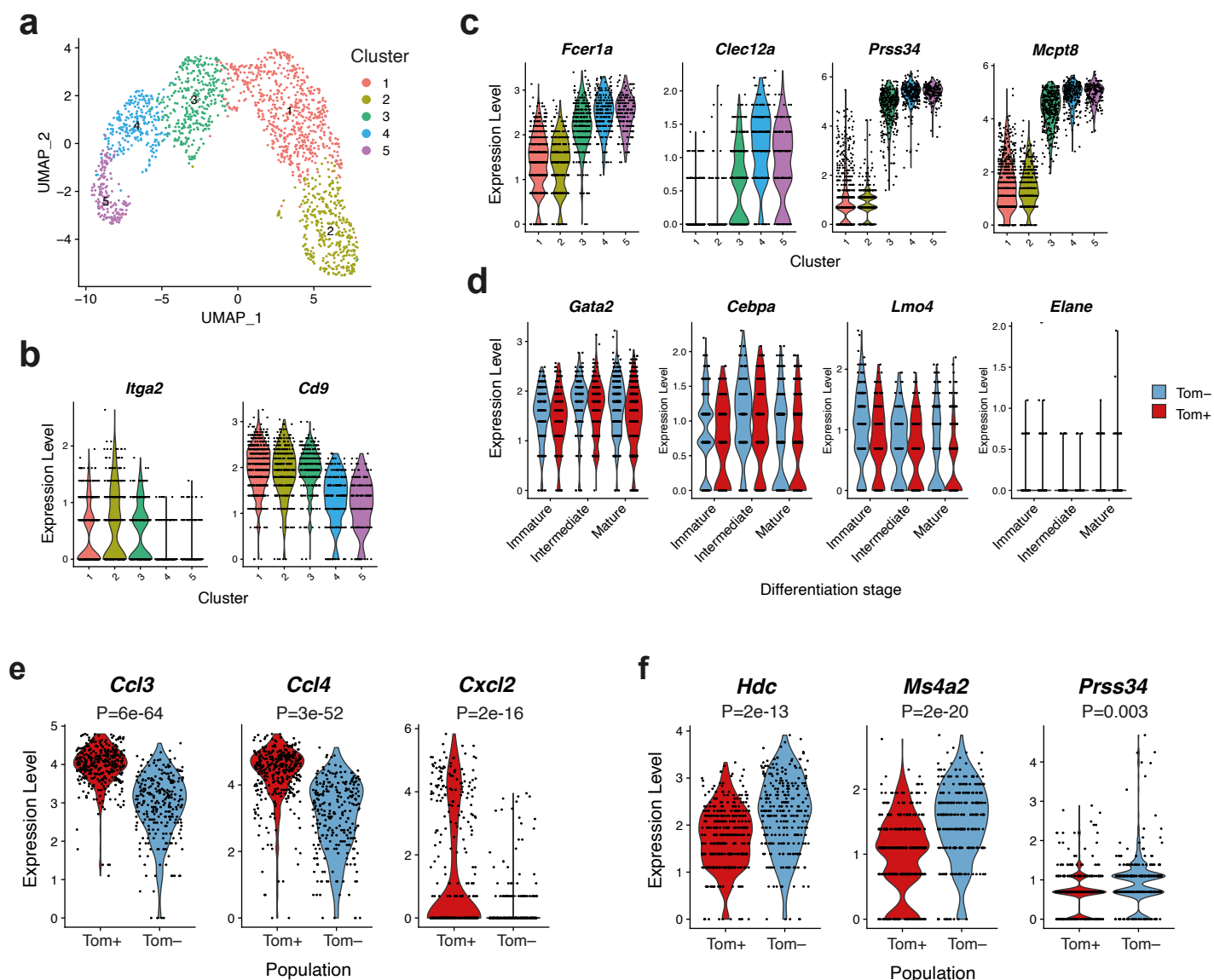
- a)** Hierarchical organisation of cell populations used to identify global and lineage-specific ATAC peaks.
- b)** Volcano plot of differentially footprinted TF motifs comparing Oligo_G1Neg and BEMP populations in Extended Data Figure 6f using TOBIAS and global peaks. The top 10 up- and down-regulated TFs by adjusted p-value are indicated.
- c)** Graphs showing the normalised mean TOBIAS foot-printing scores (left) and mean SCT normalised gene expression (right) across the EMPP and NMP trajectories for the indicated TFs. BEMP peaks were used to generate foot-printing scores.
- d)** Box-and-whisker plot showing the peaks lengths for BEMP peaks that contain both CEBPA and GATA2 motifs. Peaks were grouped into those where foot-printing was detected on both CEBPA and GATA2 motifs, and those where dual binding was not detected. P-value for the group comparison is shown (Wilcoxon test).
- e)** Box-and-whisker plot of the peaks groups from (d) showing the closest distance between CEBPA and GATA2 motifs for dual foot-printed and non-dual foot-printed peaks, as indicated. P-value for the group comparison is shown (Wilcoxon test).
- f)** Volcano plot of genes differentially expressed between the Oligo_G1Neg and BEMP populations from Extended.Data.Fig.6f using MAST. The top up- and down-regulated genes are labelled.



Extended Data Figure 8

Extended Data Fig. 8 | Gating and analysis of GMP subsets.

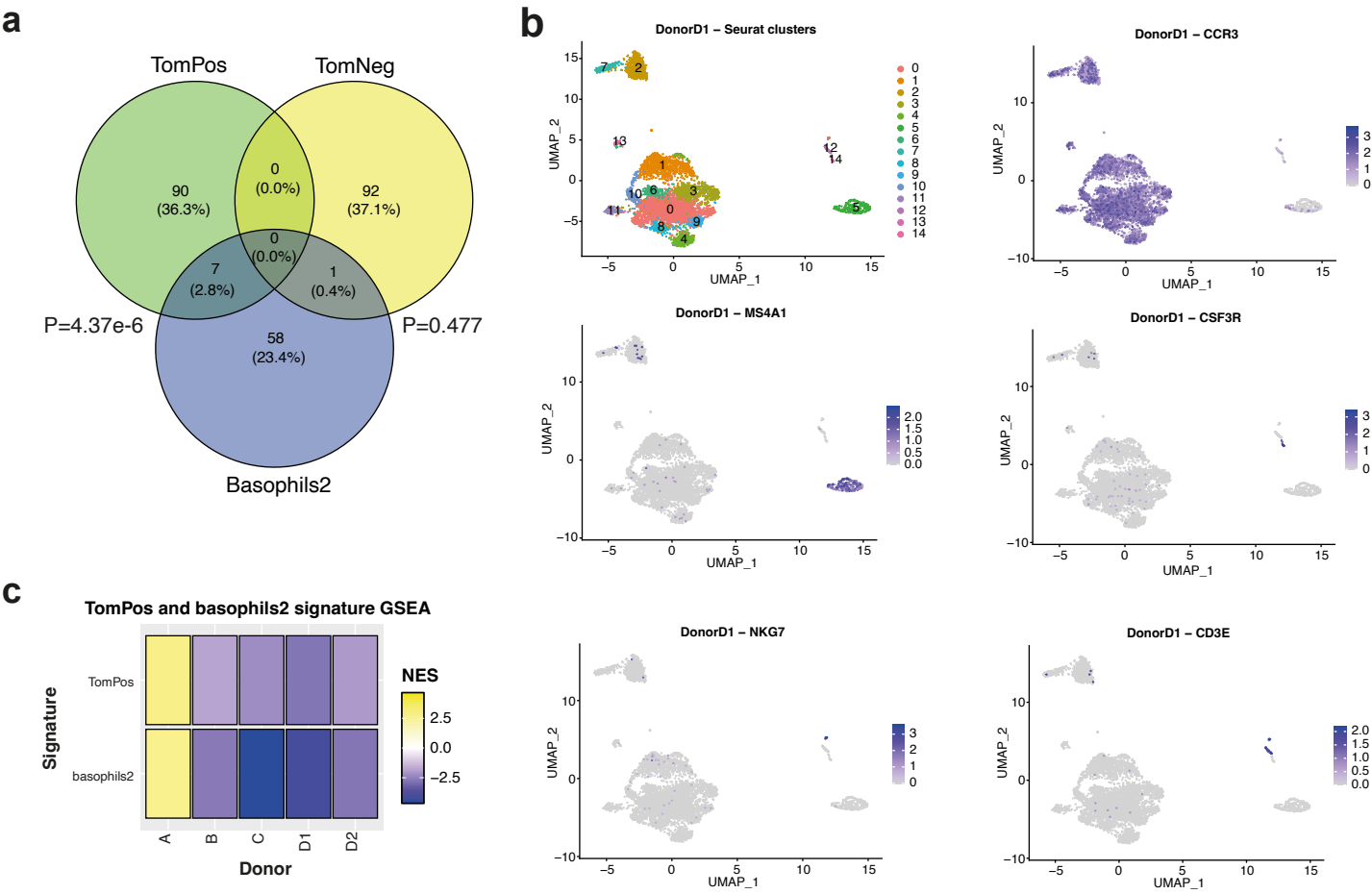
- a)** Gating strategy for identification of Tom⁺ and Tom[−] basophils and eosinophils in EC/Ai9/GE mice.
- b)** Tom expression in NMPs from EC/Ai9/GE mice, defined as in Fig.6e. The mean $\pm$ SEM of Tom⁺ cells is shown. N=6, 2 independent experiments.
- c)** Colony types generated by NMPs and BEMP sub-populations defined in Fig. 6e,f. Data from 2 independent experiments. The total number of colonies analysed for each population is shown above the columns.



Extended Data Figure 9

Extended Data Fig. 9 | Characterization of basophils in EC/Ai9/GE mice.

- a)** UMAP projection of clustering of combined spleen and bone marrow basophil single cell transcriptomes.
- b)** Violin plots of gene expression of the indicated mature basophil markers in clusters from (a) using SCT expression values from the CITEseq analysis.
- c)** Violin plots of gene expression of the indicated immature basophil markers in clusters from (a) using SCT expression values from the CITEseq analysis.
- d)** Violin plots of gene expression of the indicated transcription factors in Tom⁻ and Tom⁺ cells grouped by the differentiation stages defined in Fig. 7a using SCT expression values from the CITEseq analysis. *Elane* expression is shown as a negative control.
- e)** Violin plots of gene expression of selected genes up-regulated in mature Tom⁺ SP basophils from analysis in Fig.7e using SCT expression values from the CITEseq analysis.
- f)** Violin plots of gene expression of selected genes up-regulated in mature Tom⁻ SP basophils from analysis in Fig.7e using SCT expression values from the CITEseq analysis.



Extended Data Figure 10

Extended Data Fig. 10 | Analysis of human basophil single cell transcriptomes.

- a)** Venn diagram showing the overlap between the human baso2 signature and the humanised TomPos and TomNeg signatures from mature mouse SP basophils. The P-values of the overlaps calculated using a hypergeometric test are shown.
- b)** High-resolution clustering of human basophil sample D1, showing the cluster numbers and expression of basophil (CCR3), B cell (MS4A1), NK cell (NKG7), T-cell (CD3E) and neutrophil (CSF3R) markers. Marker expression was used to select basophil clusters for subsequent low-resolution clustering to generate 2 basophil clusters.
- c)** Heatmap of GSEA normalised enrichment scores comparing the two low-resolution clusters of human basophils from 5 donor samples using the basophils2 and TomPos signatures.

Supplementary Table 1 | CITE-seq cluster marker genes.

Marker genes for clusters in Extended.Data.Fig.5a, identified using FindAllMarkers with MAST and the SCT assay. The cluster number, gene ID, p-value, mean fold change (log2 transformed), percentage of expressing cells in each population, and the adjusted p-value for the comparison are shown. The cluster of interest is the test population and the remaining clusters the reference population, and avg_log2FC scores > 0 therefore indicates genes that are more highly expressed in the cluster of interest.

Supplementary Table 2 | Wormhole scores for multiome trajectory-cluster pairs.

Scoring of cross-branch cluster pairs similarity from the Slingshot lineage graph. Each row compares two branch-specific clusters and lists their branch identities, cell counts, trajectory graph separation, SCT expression similarity, chromVAR motif similarity, Fisher-transformed and standardized SCT/chromVAR values, the SCT+chromVAR concordance penalty, and the final concordant wormhole score.

Supplementary Table 3 | Differential foot-printing between Oligo_G1Neg and BEMP clusters.

TFs showing differential foot-printing comparing the Oligo_G1Neg and BEMP populations from Extended.Data.Fig.6f using TOBIAS on BEMP or global chromatin, as indicated. Columns show chromatin domain used, TF name, mean binding scores, binding difference, raw and adjusted p-values, and the negative of the log10-transformed adjusted p-value.

Supplementary Table 4 | Differentially expressed TFs between Oligo_G1Neg and BEMP clusters.

Genes showing differential expression of TF and co-TF encoding genes between the Oligo_G1Neg and BEMP populations from Extended.Data.Fig.6f, identified using FindMarkers with MAST and the SCT assay. Columns show gene ID, p-value, mean fold change (log2 transformed), and the adjusted p-value for the comparison. BEMPs were the test population and Oligo_G1Neg the reference; positive avg_log2FC indicates higher expression in BEMPs.

Supplementary Table 5 | Signatures of mature Tom+ and Tom- SP basophils.

The top 100 genes, ranked by log2 fold change, upregulated in Tom- (TomNeg) and Tom+ mature SP basophils comparing the two populations using FindMarkers with MAST. Columns show gene IDs.

Supplementary Table 6 | Antibodies used for flow cytometry and cell sorting.

List of antibodies used for flow cytometry and cell sorting. Columns show antigen, fluorochrome, clone ID, supplier, concentration used, and the type of analysis each antibody was used for.

Supplementary Table 7 | Gene sets used for AUCell scores and cell cycle regression.

Cell type signatures used to generate AUCell scores and cell cycle regression. Gene IDs are shown for each signature.

Supplementary Table 8 | Genes associated with BEMP peaks.

Genes identified as being proximal to BEMP-specific peaks, classified according to the detection of C/EBP α and GATA-2 motif or foot-printing within the peak region.