Supplementary : data analysis methods

April 1, 2024

Contents

1 Quality control

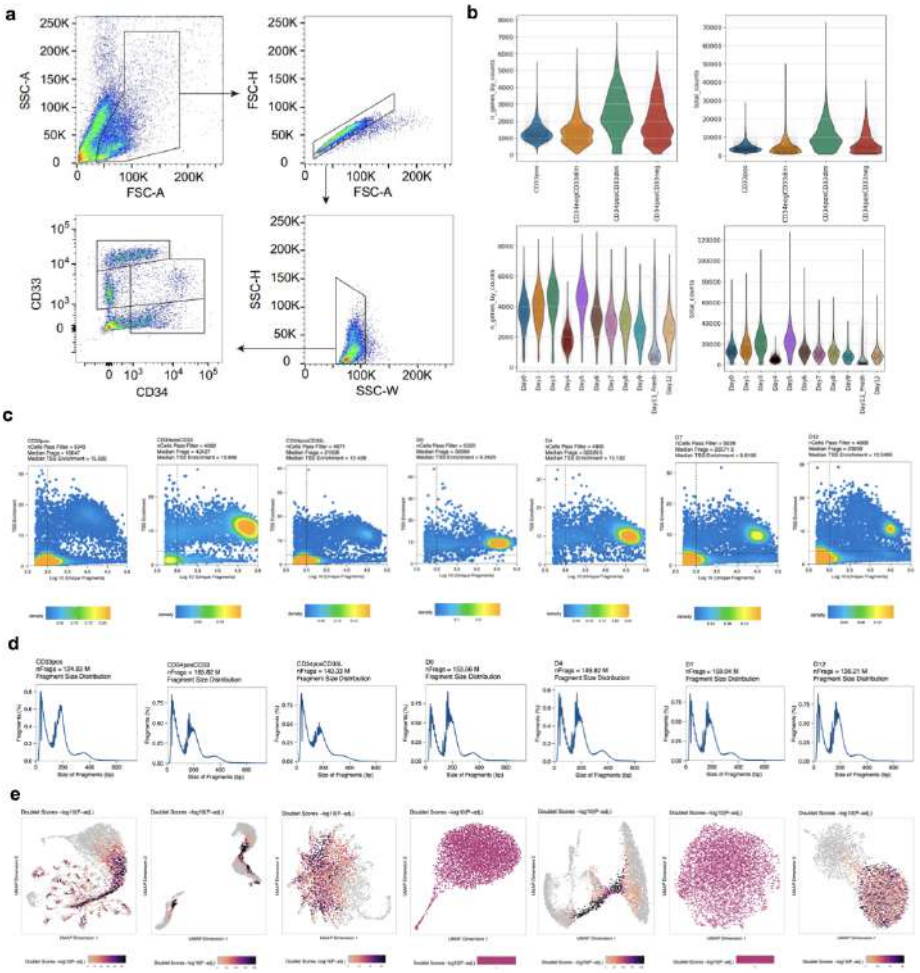


Figure 1: Supplementary: a. Gating strategy used for FACS sorting on human bone marrow cells. b. Quality control for RNA samples. Violin plots show the number of non-zero genes and total counts per cell for indifivudla samples (top for sorted samples, bottom for timepoints samples). c,d,e. Quality control for ATAC samples: TSS by unique fragments, fragment size distribution, doublet scores

2 Cell type annotation

2.1 Scatter plots

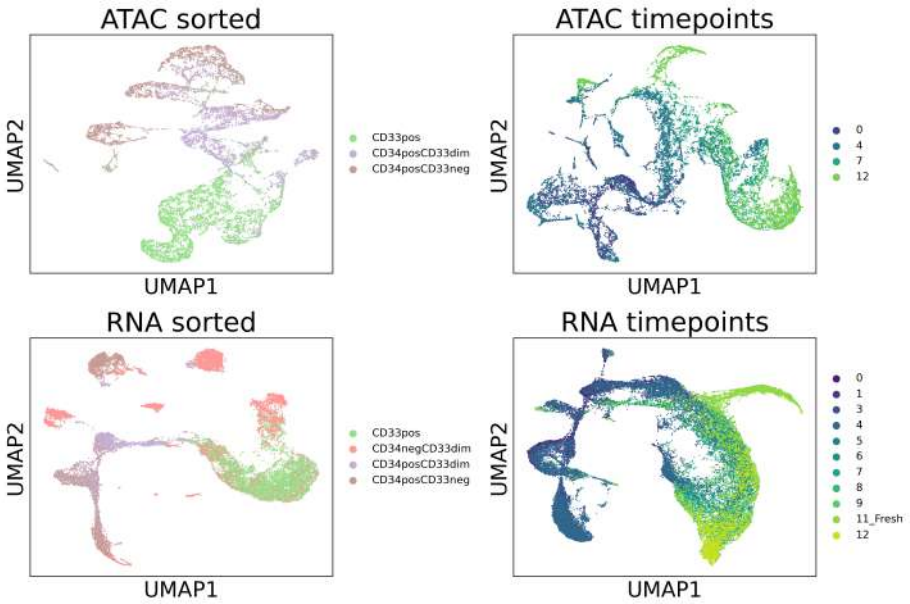


Figure 2: Supplementary: Time points (days) and sorted assays (CD) annotation

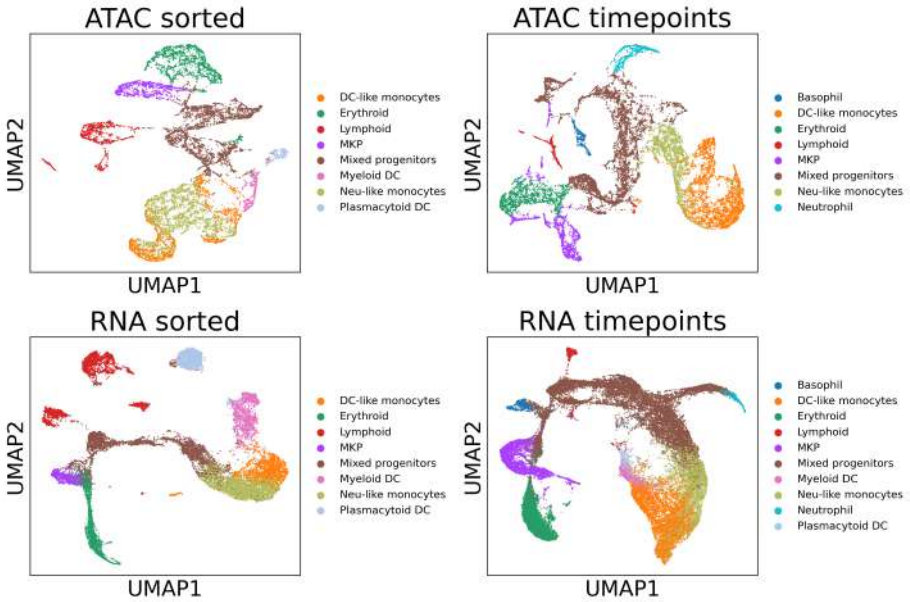


Figure 3: Supplementary: Leiden identity simple annotation

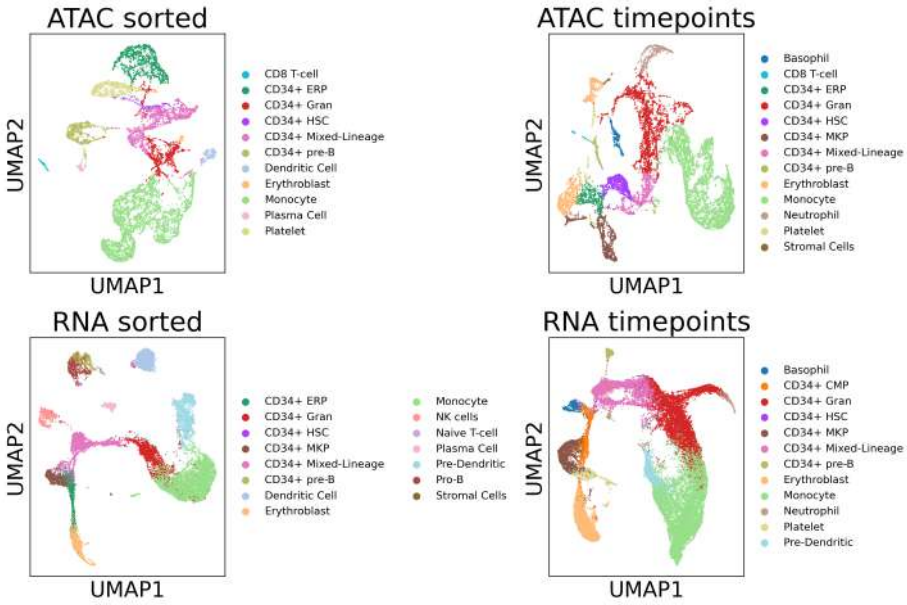


Figure 4: Supplementary: Leiden identity annotation

2.2 Bar plots

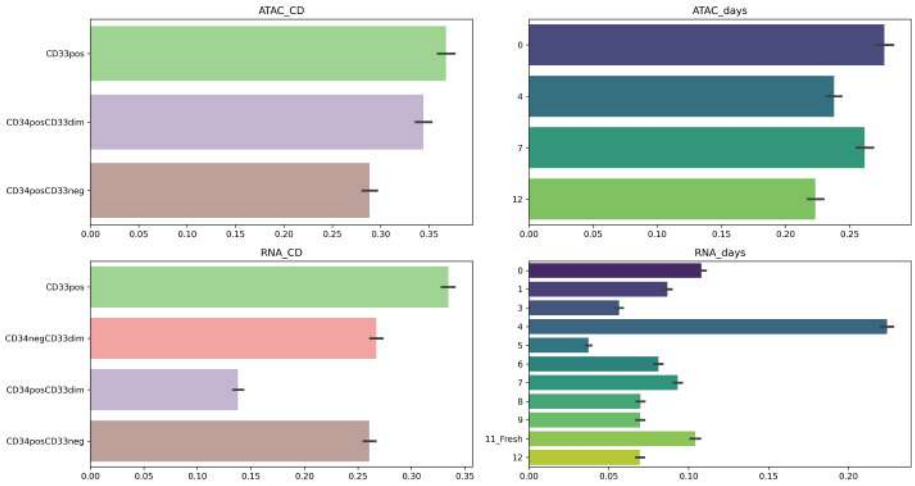


Figure 5: Supplementary: Leiden identity simple annotation

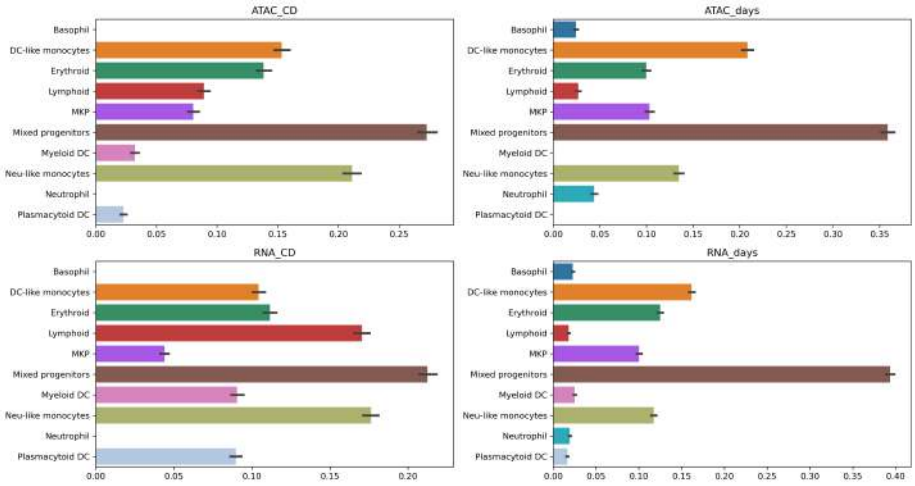


Figure 6: Supplementary: Leiden identity simple annotation

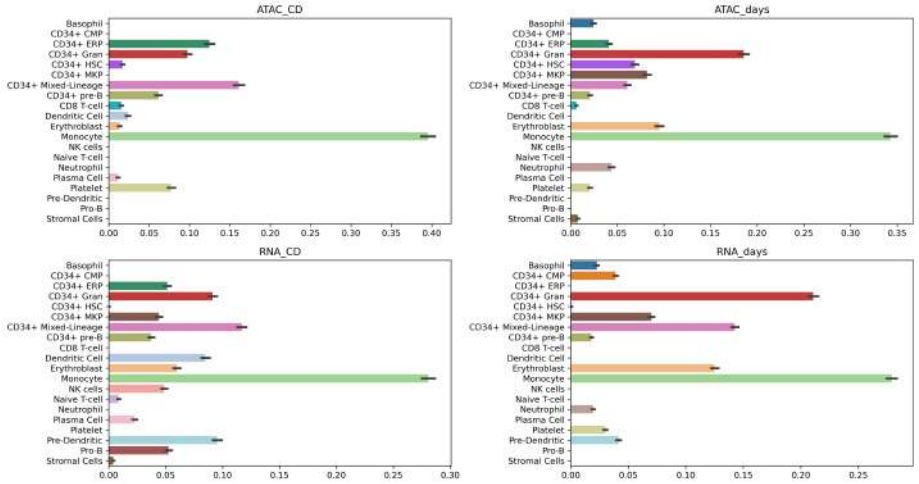
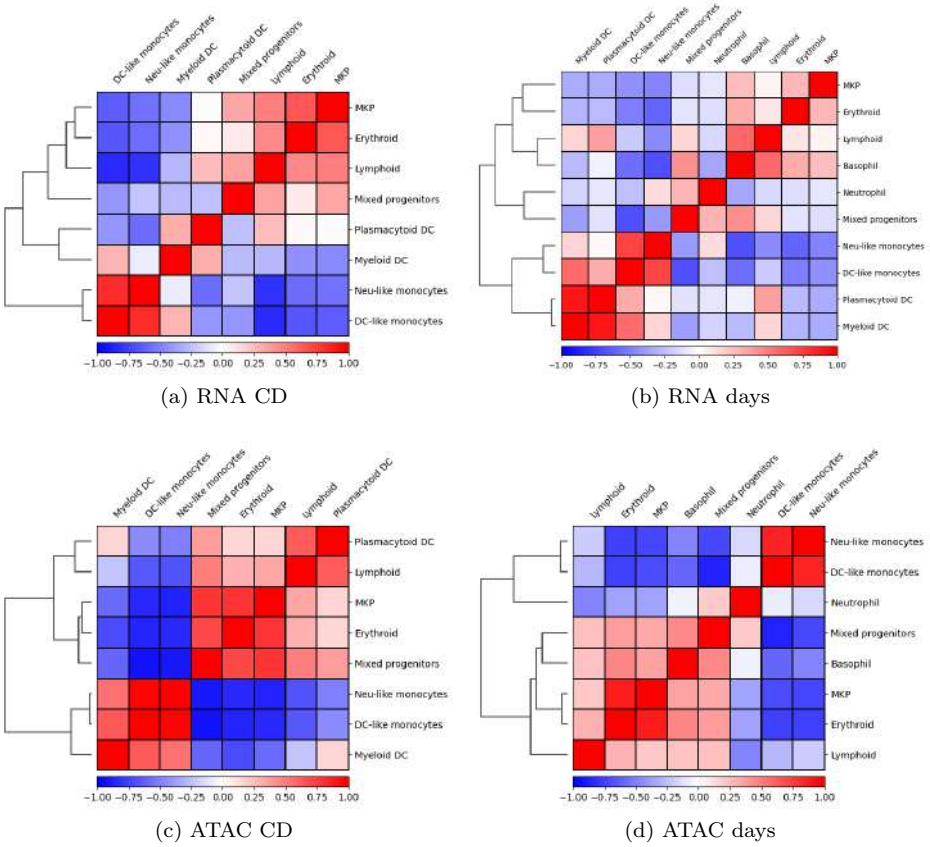


Figure 7: Supplementary: Leiden identity annotation

2.3 Correlation plots



(a) RNA CD

(b) RNA days

(c) ATAC CD

(d) ATAC days

Figure 8: Correlation between cell type averages (obtained using top PCs)

3 Data integration

3.1 Linear optimal transport

3.1.1 Scatter plots

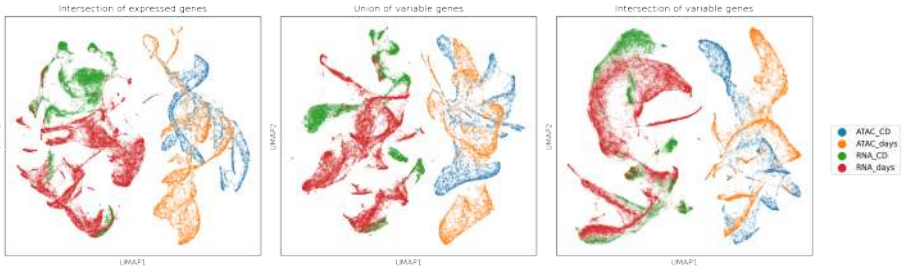


Figure 9: Assays in RNA days PCA subspace

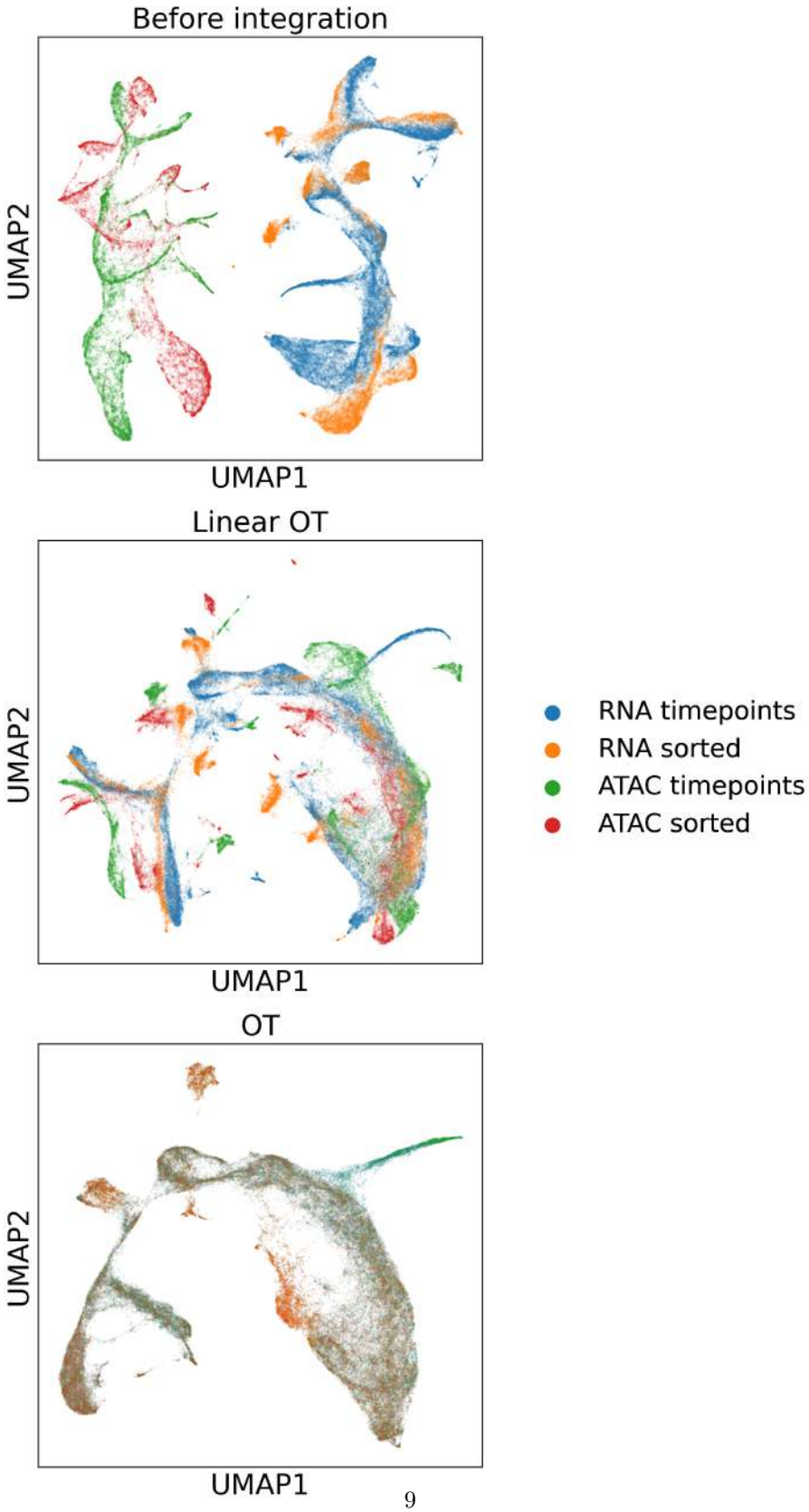


Figure 10: Assays in RNA days PCA subspace, after alignment with linear optimal transport

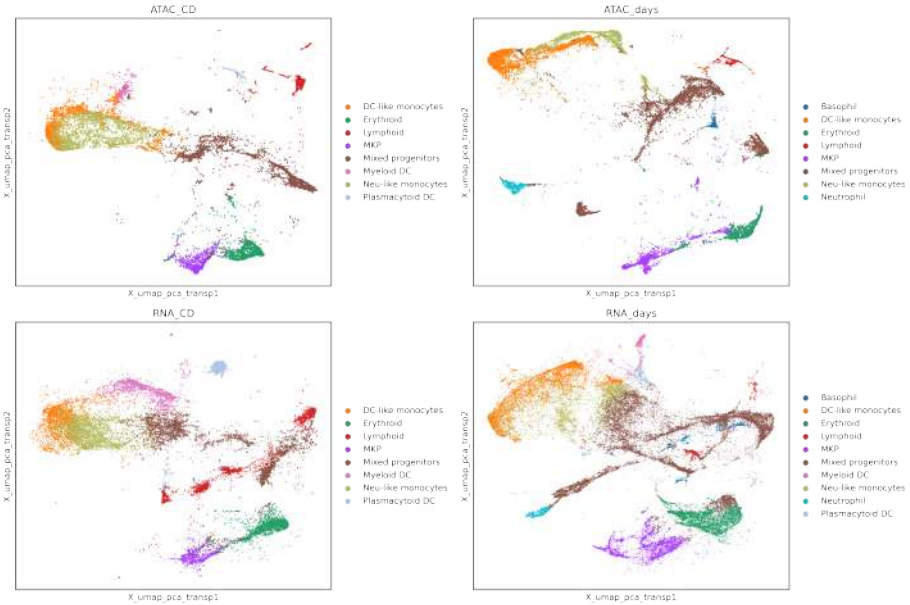


Figure 11: Assays in RNA days PCA subspace obtained using the intersection of expressed genes, after alignment with linear optimal transport.

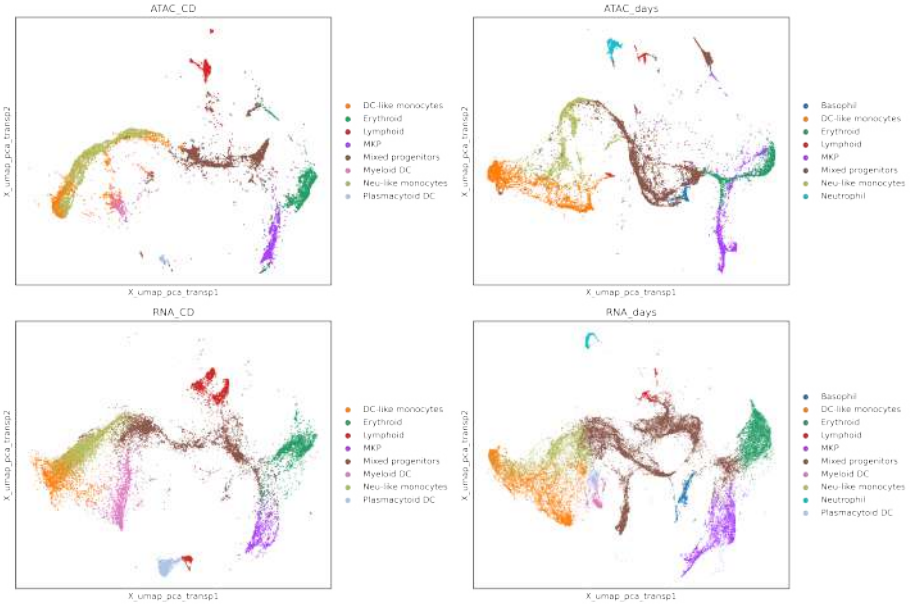


Figure 12: Assays in RNA days PCA subspace obtained using the union of variable genes, after alignment with linear optimal transport.

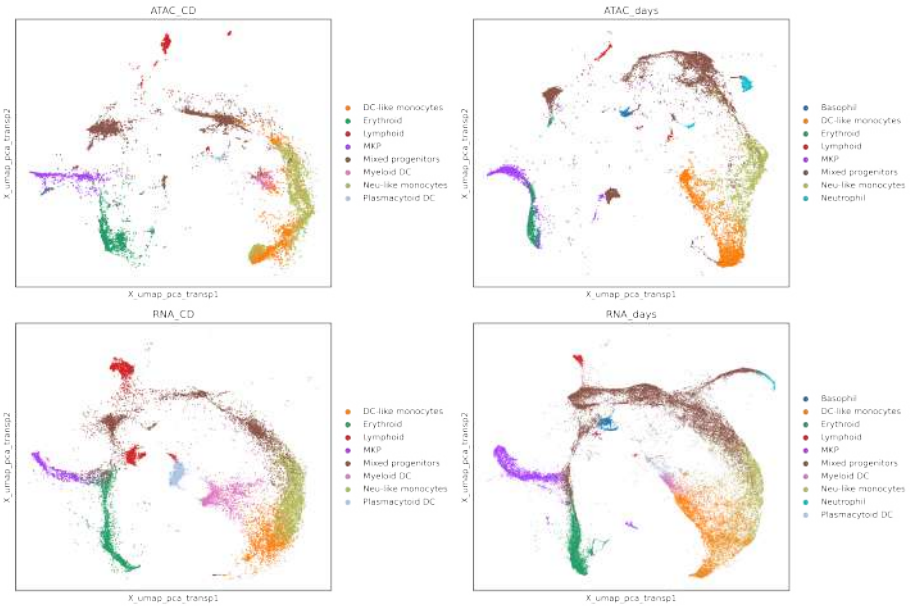


Figure 13: Assays in RNA days PCA subspace obtained using the intersection of variable genes, after alignment with linear optimal transport.

3.1.2 Correlation plots

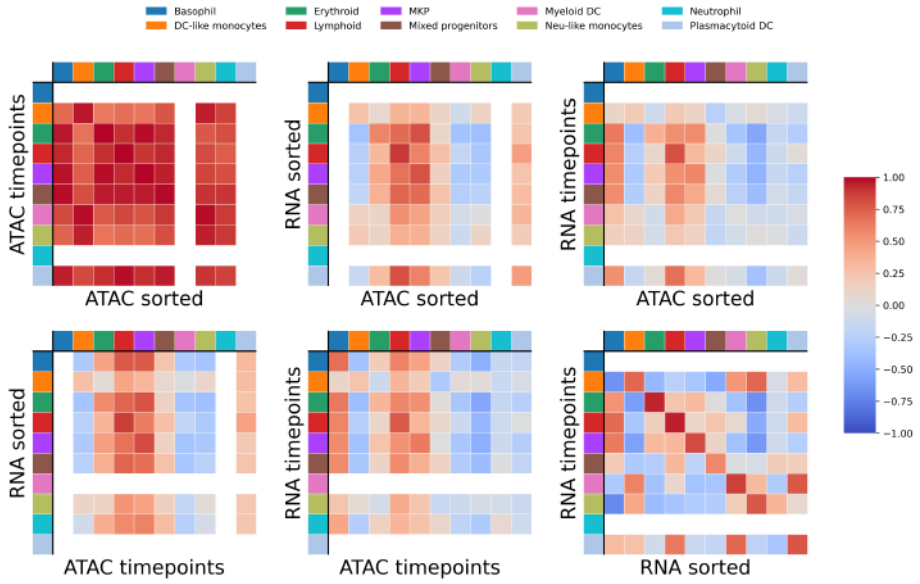


Figure 14: Correlation between assays' cell type averages (obtained using top PCs of assays, each in their own PCA subspace)

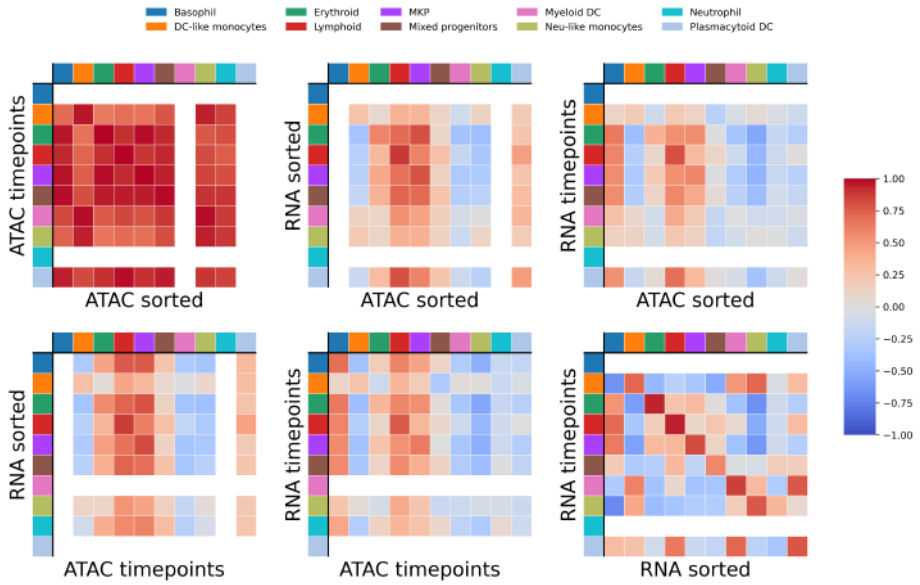


Figure 15: Correlation between assays' cell type averages (obtained using top PCs of assays in RNA days PCA subspace)

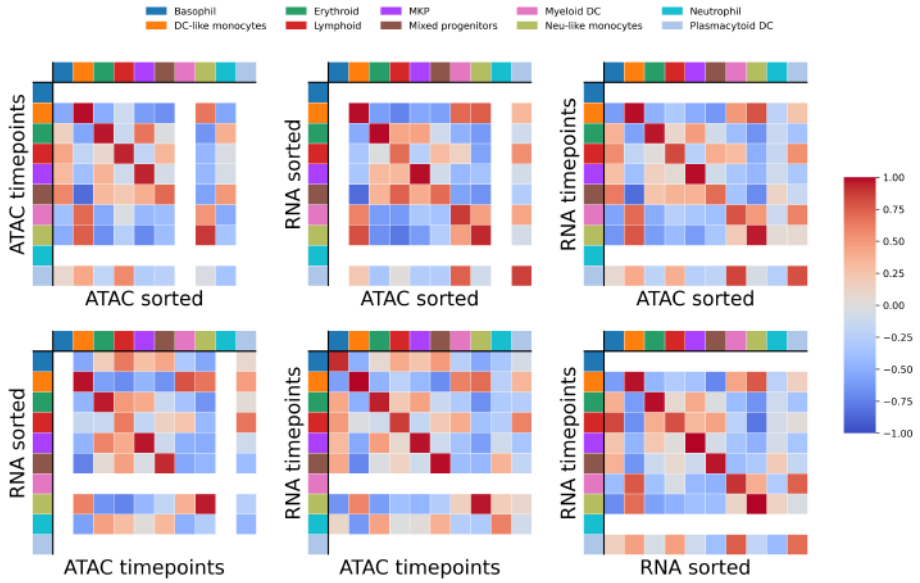


Figure 16: Correlation between assays' cell type averages (obtained using top PCs of assays in RNA days PCA subspace, after alignment with linear optimal transport)

3.2 Supervised optimal transport

Grouping of categories to go from full to simple annotation

- 'Early-ERP','Erythroblast','CD34+ ERP' = 'Erythroid'
- 'Platelet','CD34+ MKP' = 'MKP'
- 'Pre-Dendritic','Dendritic Cell' = 'Dendritic'
- 'CD34+ CLP','CD34+ pre-B','Pro-B','Plasma Cell', 'NK cells','Naive T-cell','CD8 T-cell' = 'Lymphoid'
- 'CD34+ Mixed-Lineage', 'CD34+ HSC','CD34+ CMP', 'CD34+ Gran', 'Eosinophil','Stromal Cells' = 'Mixed-Lineage'
- change unlikely 'Erythroid' annotation (outlier, mixed cluster) for late ATAC days (day 7 and 12) to 'Mixed-Lineage'

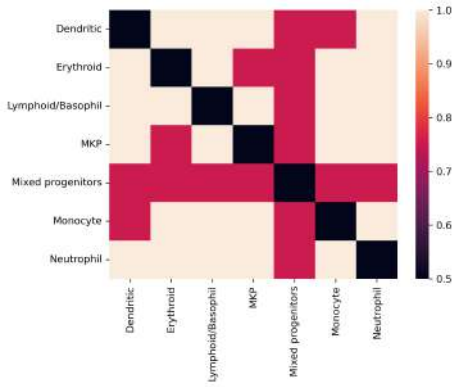


Figure 17: Supplementary: supervision of OT cost matrix

4 HSC score

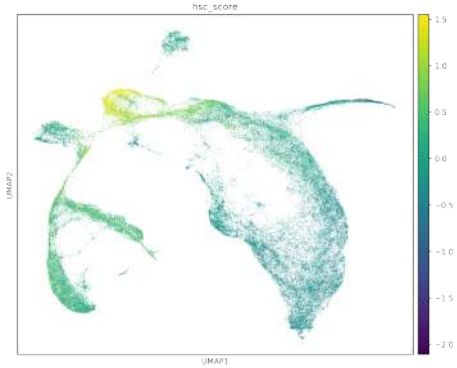


Figure 18: Supplementary: hematopoeitic stem cell score

5 Trajectory inference

5.1 Trajectory in PCA space

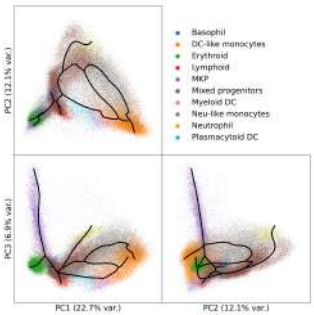


Figure 19: Supplementary: Trajectory visualized in PCA space

6 Trajectory analysis

6.1 Erythroid branch

6.1.1 Pseudotime

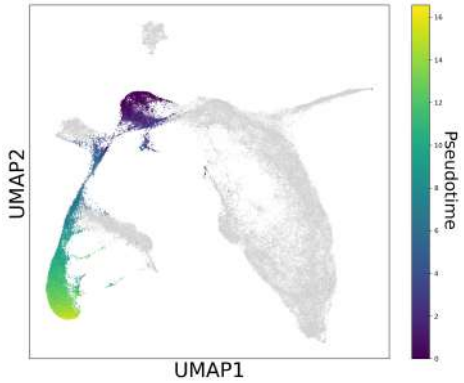


Figure 20: Supplementary: Erythroid branch pseudotime

6.1.2 Volcanos

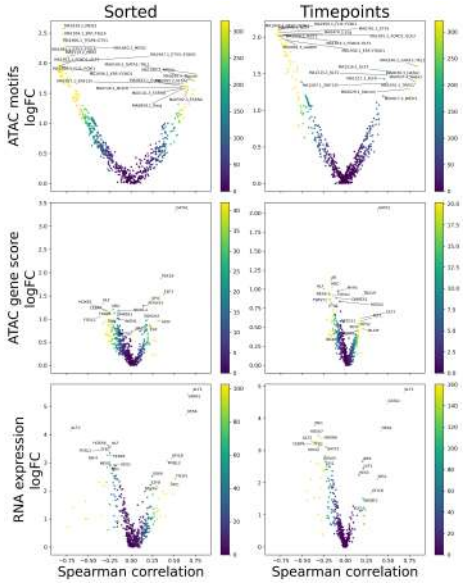


Figure 21: Supplementary: Erythroid branch marker detection with Spearman correlation. The volcano plot is colored by -log10 adjusted p-values

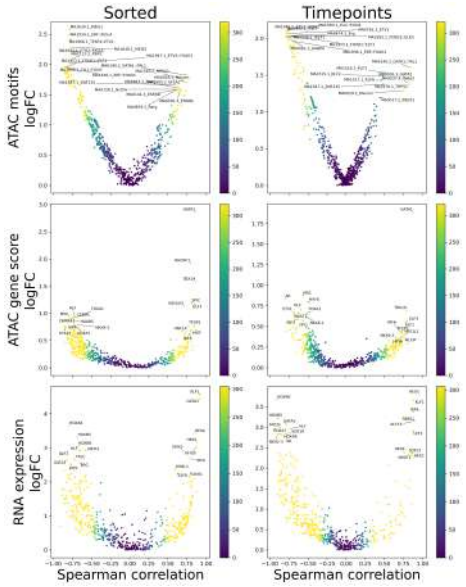


Figure 22: Supplementary: Erythroid branch marker detection with Spearman correlation and MAGIC imputed matrices for gene score and RNA expression. The volcano plot is colored by -log10 adjusted p-values

6.1.3 Heatmap

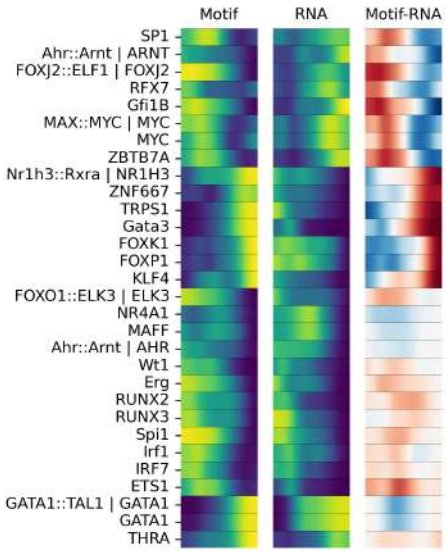


Figure 23: Supplementary: Heatmap of top negative and positive correlations against pseudotime, grouping all RNA and chromVAR assays

6.1.4 X plots

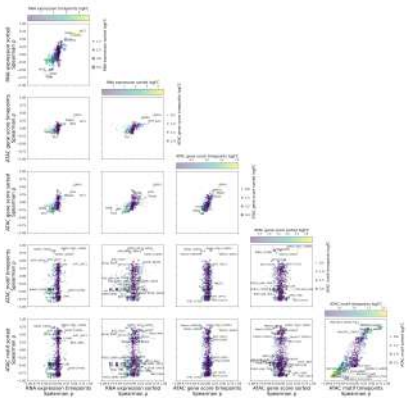


Figure 24: Supplementary: X-plot showing Spearman correlations against pseudotime, for all pairs of assays

6.2 MKP branch

6.2.1 Pseudotime

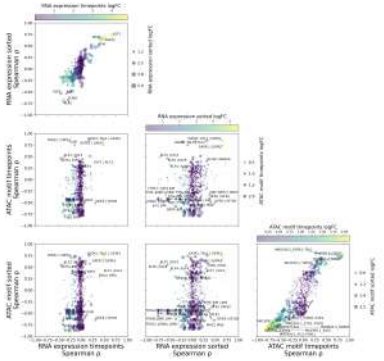


Figure 25: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping sorted and timepoints assays for RNA and gene score

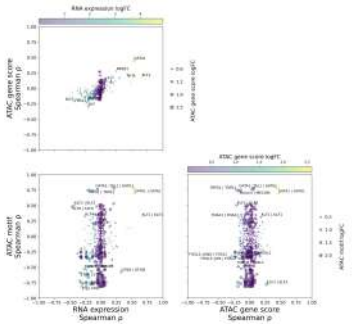


Figure 26: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping all RNA and chromVAR assays

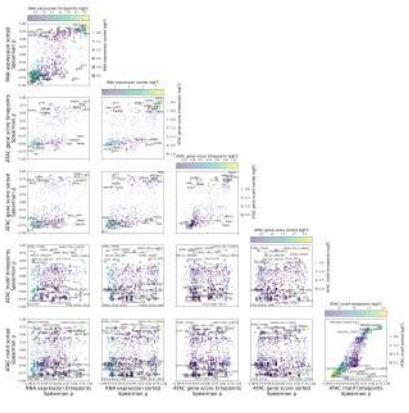


Figure 27: Supplementary: X-plot showing Spearman correlations against pseudotime, for all pairs of assays, with MAGIC smoothing for gene score and RNA expression

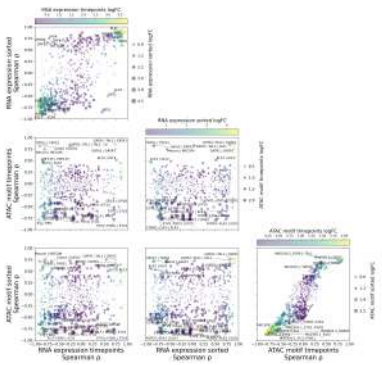


Figure 28: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping sorted and timepoints assays for RNA and gene score, with MAGIC smoothing for gene score and RNA expression

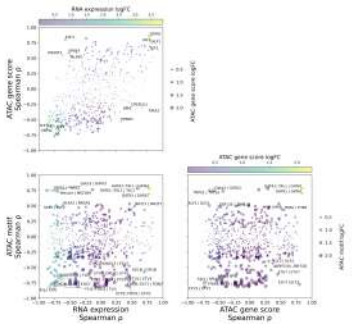


Figure 29: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping all RNA and chromVAR assays with MAGIC smoothing for gene score and RNA expression

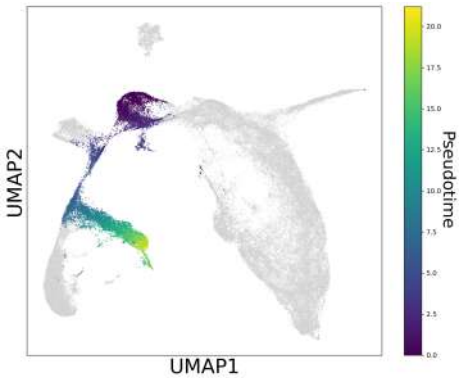


Figure 30: Supplementary: MKP branch pseudotime

6.2.2 Volcanos

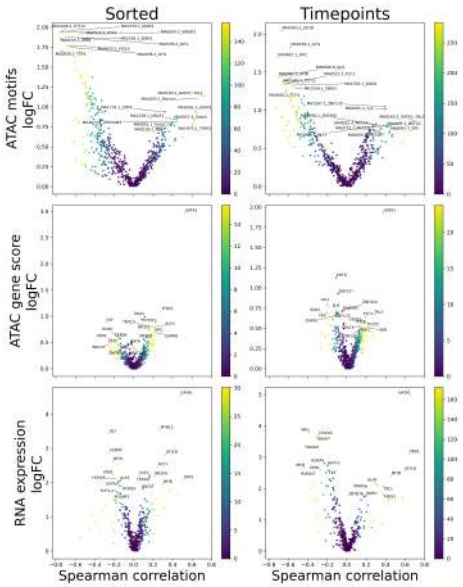


Figure 31: Supplementary: Erythroid branch marker detection with Spearman correlation. The volcano plot is colored by -log10 adjusted p-values

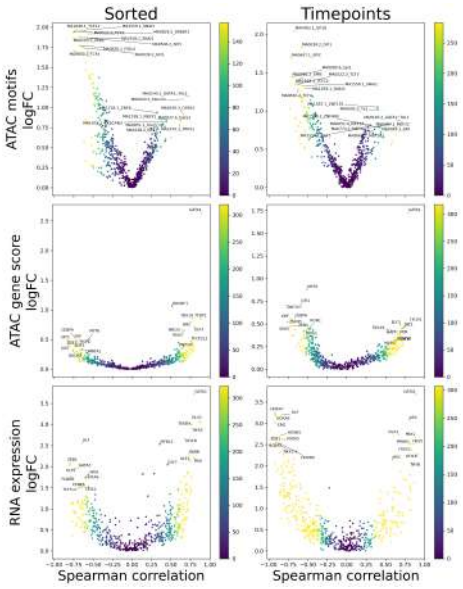


Figure 32: Supplementary: Erythroid branch marker detection with Spearman correlation and MAGIC imputed matrices for gene score and RNA expression. The volcano plot is colored by -log10 adjusted p-values

6.2.3 Heatmap

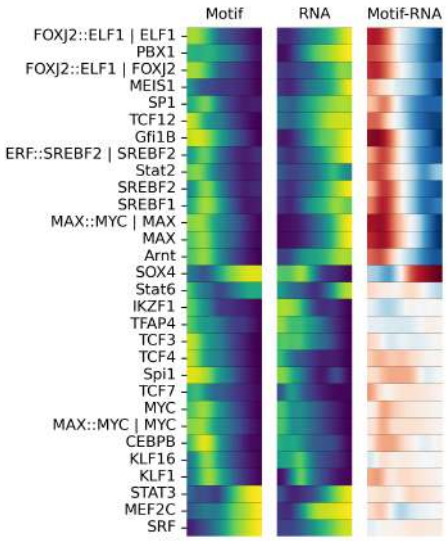


Figure 33: Supplementary: Heatmap of top negative and positive correlations against pseudotime, grouping all RNA and chromVAR assays

6.2.4 X plots

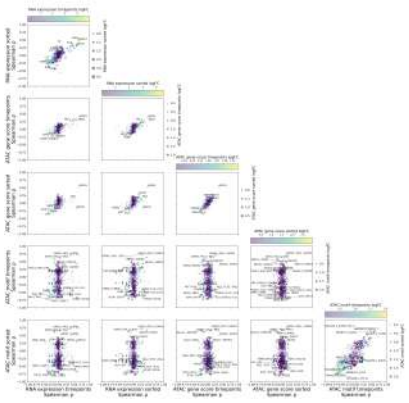


Figure 34: Supplementary: X-plot showing Spearman correlations against pseudotime, for all pairs of assays

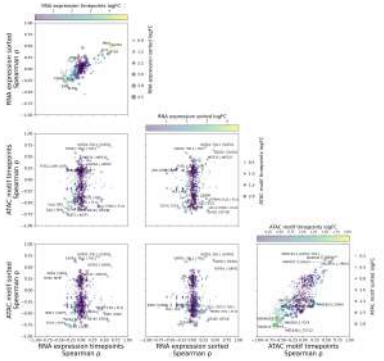


Figure 35: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping sorted and timepoints assays for RNA and gene score

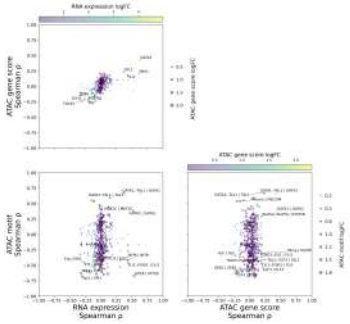


Figure 36: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping all RNA and chromVAR assays

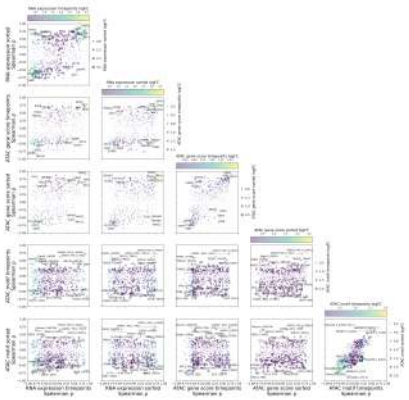


Figure 37: Supplementary: X-plot showing Spearman correlations against pseudotime, for all pairs of assays, with MAGIC smoothing for gene score and RNA expression

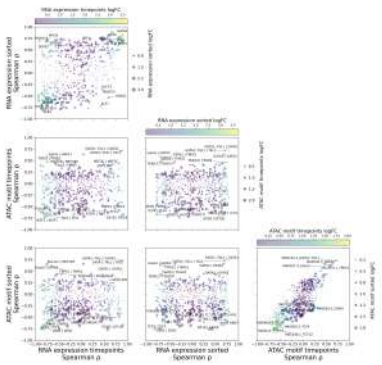


Figure 38: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping sorted and timepoints assays for RNA and gene score, with MAGIC smoothing for gene score and RNA expression

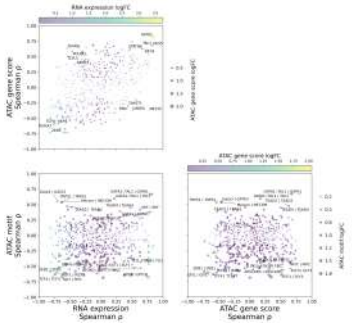


Figure 39: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping all RNA and chromVAR assays with MAGIC smoothing for gene score and RNA expression

6.3 Basophil branch

6.3.1 Pseudotime

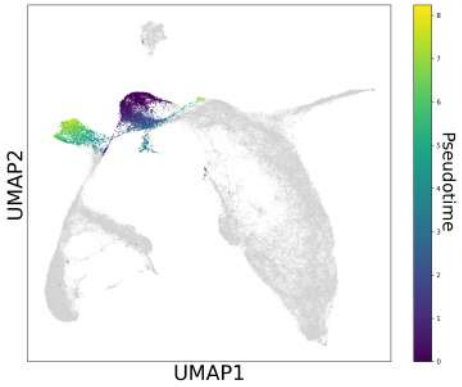


Figure 40: Supplementary: MKP branch pseudotime

6.3.2 Volcanos

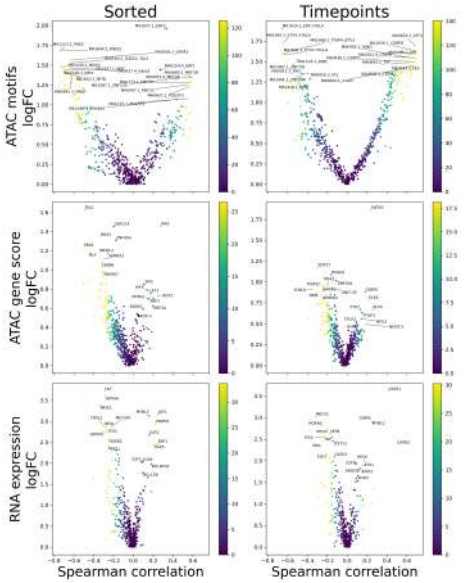


Figure 41: Supplementary: Erythroid branch marker detection with Spearman correlation. The volcano plot is colored by -log10 adjusted p-values

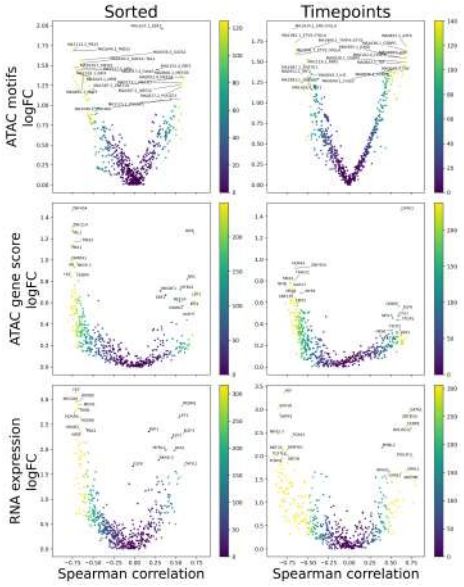


Figure 42: Supplementary: Erythroid branch marker detection with Spearman correlation and MAGIC imputed matrices for gene score and RNA expression. The volcano plot is colored by -log10 adjusted p-values

6.3.3 Heatmap

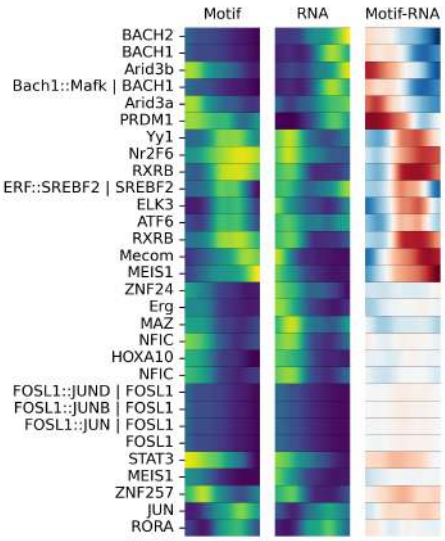


Figure 43: Supplementary: Heatmap of top negative and positive correlations against pseudotime, grouping all RNA and chromVAR assays

6.3.4 X plots

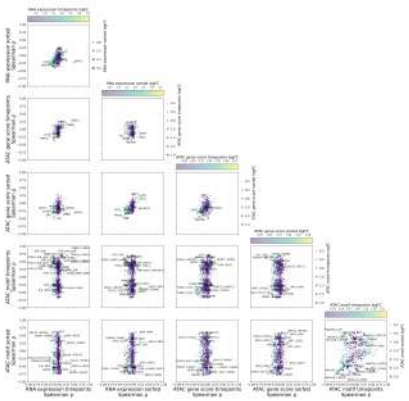


Figure 44: Supplementary: X-plot showing Spearman correlations against pseudotime, for all pairs of assays

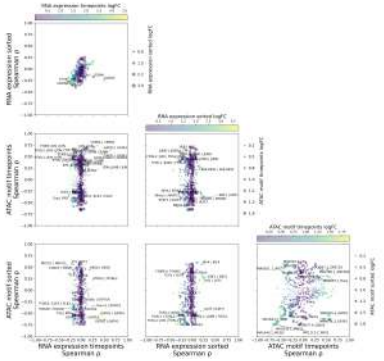


Figure 45: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping sorted and timepoints assays for RNA and gene score

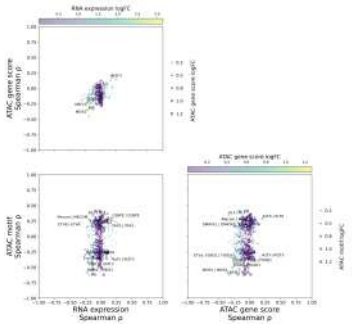


Figure 46: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping all RNA and chromVAR assays

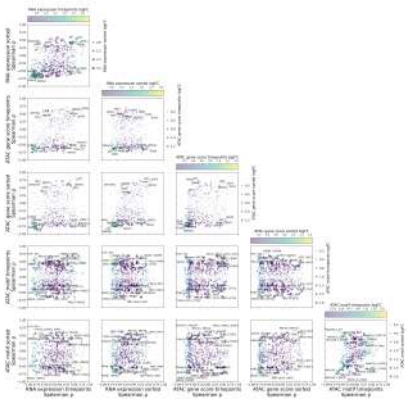


Figure 47: Supplementary: X-plot showing Spearman correlations against pseudotime, for all pairs of assays, with MAGIC smoothing for gene score and RNA expression

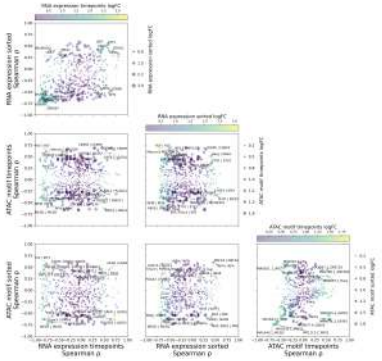


Figure 48: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping sorted and timepoints assays for RNA and gene score, with MAGIC smoothing for gene score and RNA expression

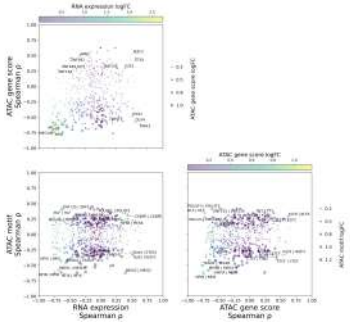


Figure 49: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping all RNA and chromVAR assays with MAGIC smoothing for gene score and RNA expression

6.4 Neutrophil branch

6.4.1 Pseudotime

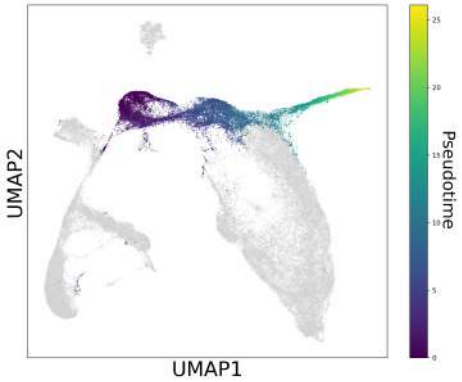


Figure 50: Supplementary: MKP branch pseudotime

6.4.2 Volcanos

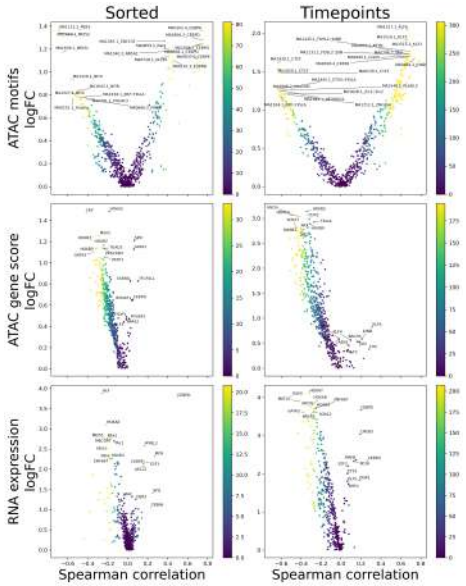


Figure 51: Supplementary: Erythroid branch marker detection with Spearman correlation. The volcano plot is colored by -log10 adjusted p-values

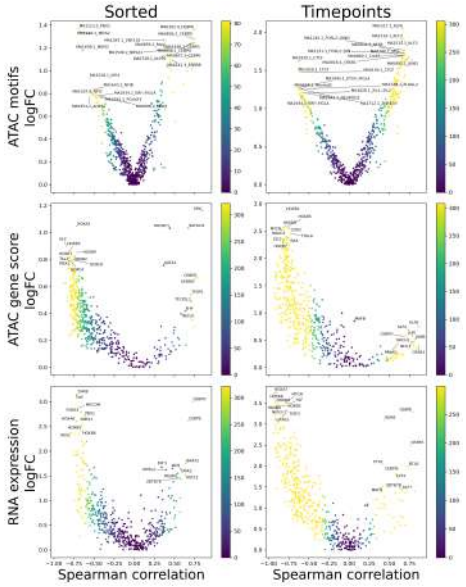


Figure 52: Supplementary: Erythroid branch marker detection with Spearman correlation and MAGIC imputed matrices for gene score and RNA expression. The volcano plot is colored by -log10 adjusted p-values

6.4.3 Heatmap

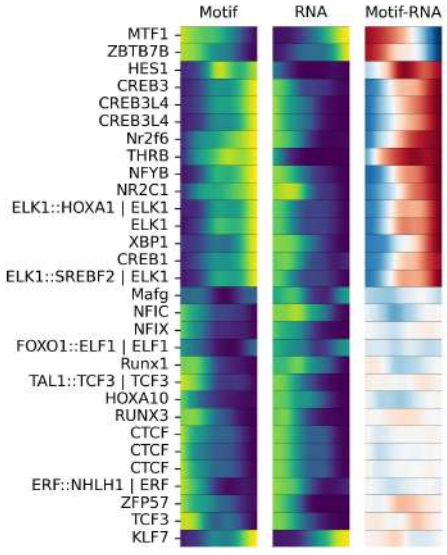


Figure 53: Supplementary: Heatmap of top negative and positive correlations against pseudotime, grouping all RNA and chromVAR assays

6.4.4 X plots

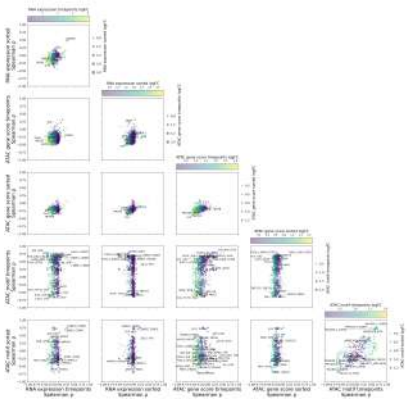


Figure 54: Supplementary: X-plot showing Spearman correlations against pseudotime, for all pairs of assays

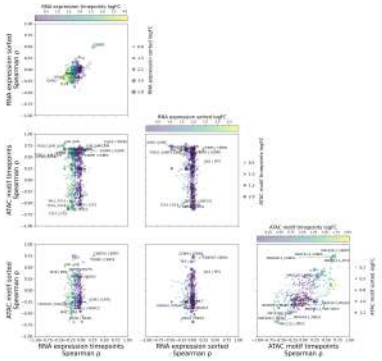


Figure 55: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping sorted and timepoints assays for RNA and gene score

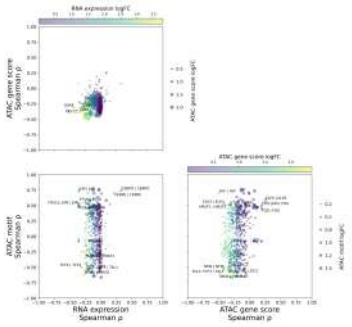


Figure 56: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping all RNA and chromVAR assays

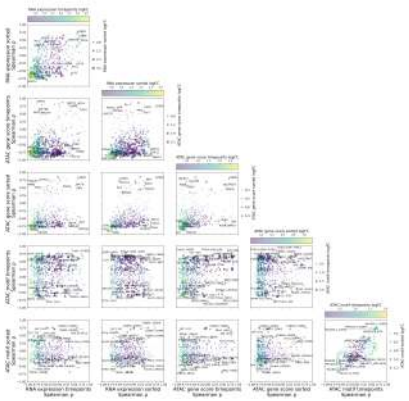


Figure 57: Supplementary: X-plot showing Spearman correlations against pseudotime, for all pairs of assays, with MAGIC smoothing for gene score and RNA expression

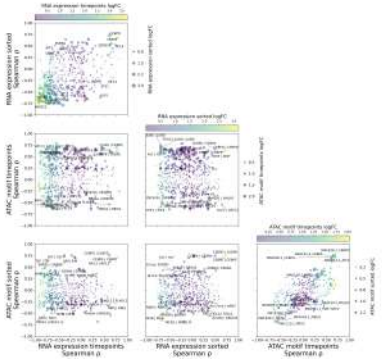


Figure 58: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping sorted and timepoints assays for RNA and gene score, with MAGIC smoothing for gene score and RNA expression

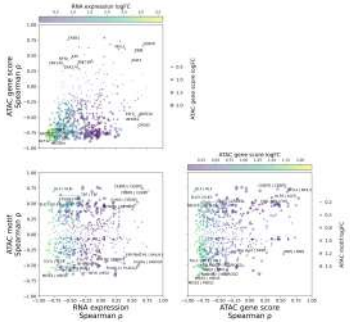


Figure 59: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping all RNA and chromVAR assays with MAGIC smoothing for gene score and RNA expression

6.5 Monocyte branch

6.5.1 Pseudotime

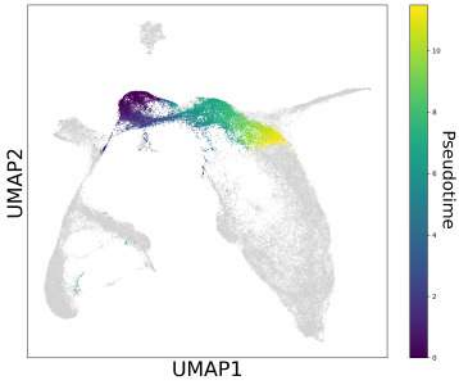


Figure 60: Supplementary: MKP branch pseudotime

6.5.2 Volcanos

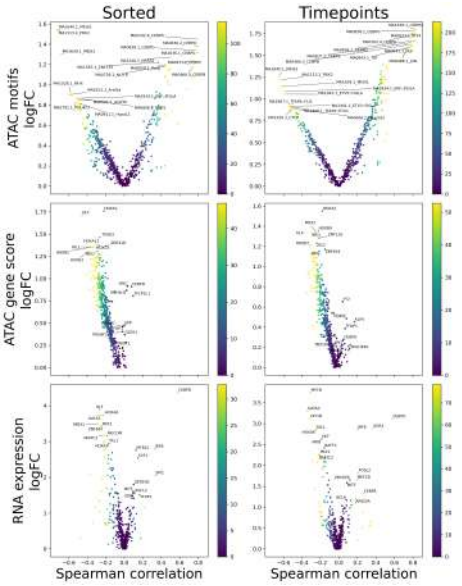


Figure 61: Supplementary: Erythroid branch marker detection with Spearman correlation. The volcano plot is colored by -log10 adjusted p-values

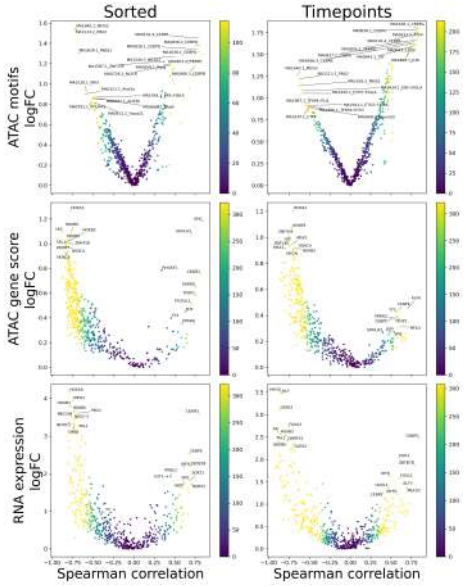


Figure 62: Supplementary: Erythroid branch marker detection with Spearman correlation and MAGIC imputed matrices for gene score and RNA expression. The volcano plot is colored by -log10 adjusted p-values

6.5.3 Heatmap

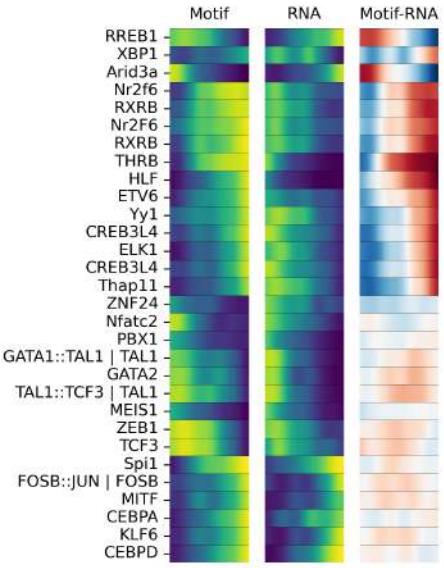


Figure 63: Supplementary: Heatmap of top negative and positive correlations against pseudotime, grouping all RNA and chromVAR assays

6.5.4 X plots

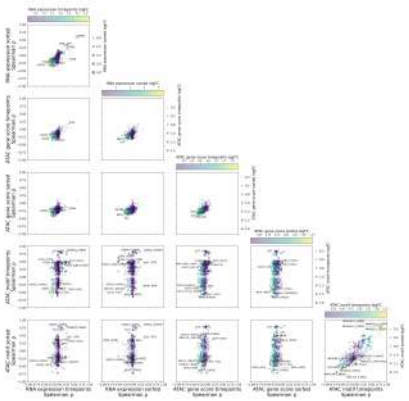


Figure 64: Supplementary: X-plot showing Spearman correlations against pseudotime, for all pairs of assays

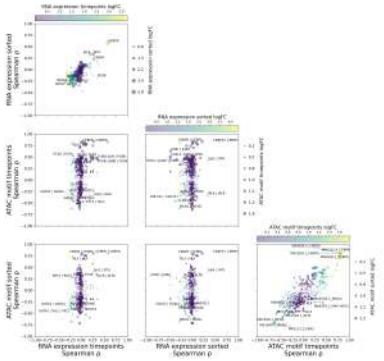


Figure 65: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping sorted and timepoints assays for RNA and gene score

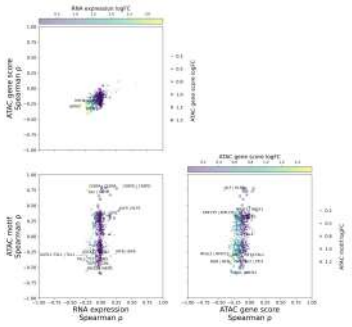


Figure 66: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping all RNA and chromVAR assays

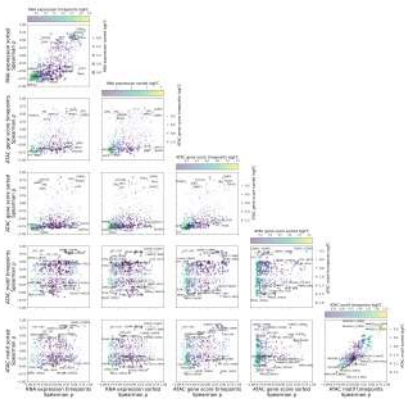


Figure 67: Supplementary: X-plot showing Spearman correlations against pseudotime, for all pairs of assays, with MAGIC smoothing for gene score and RNA expression

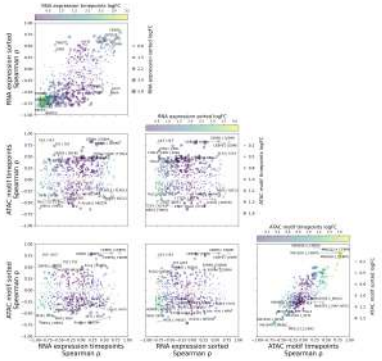


Figure 68: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping sorted and timepoints assays for RNA and gene score, with MAGIC smoothing for gene score and RNA expression

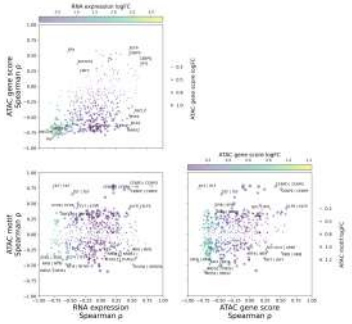


Figure 69: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping all RNA and chromVAR assays with MAGIC smoothing for gene score and RNA expression

6.6 Myeloid DC branch

6.6.1 Pseudotime

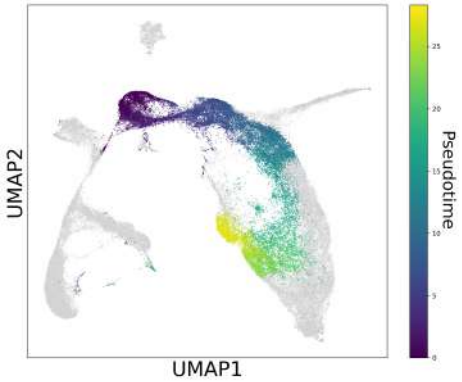


Figure 70: Supplementary: MKP branch pseudotime

6.6.2 Volcanos

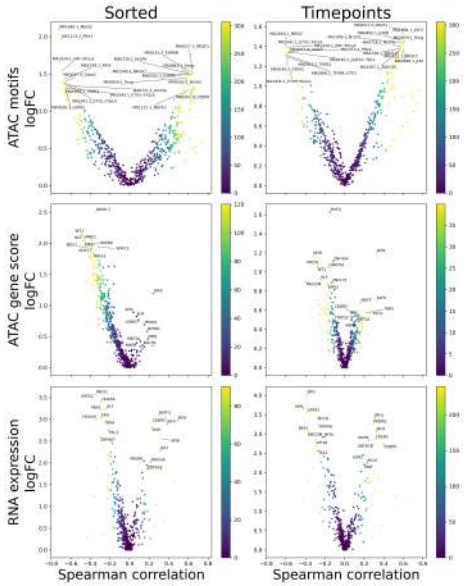


Figure 71: Supplementary: Erythroid branch marker detection with Spearman correlation. The volcano plot is colored by -log10 adjusted p-values

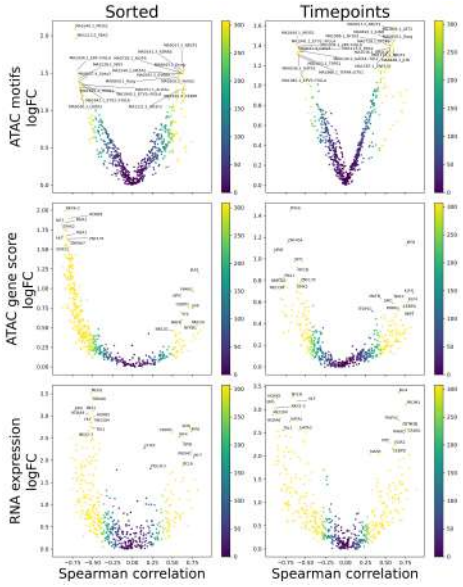


Figure 72: Supplementary: Erythroid branch marker detection with Spearman correlation and MAGIC imputed matrices for gene score and RNA expression. The volcano plot is colored by -log10 adjusted p-values

6.6.3 Heatmap

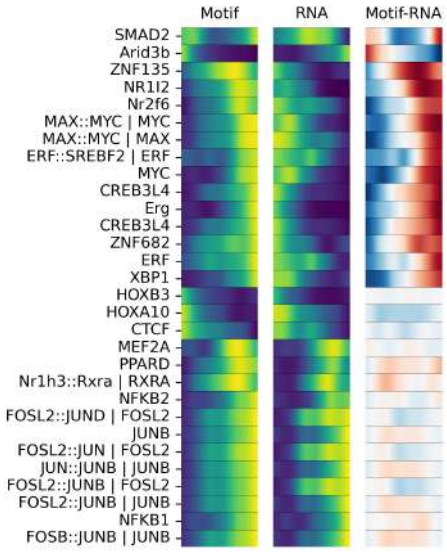


Figure 73: Supplementary: Heatmap of top negative and positive correlations against pseudotime, grouping all RNA and chromVAR assays

6.6.4 X plots

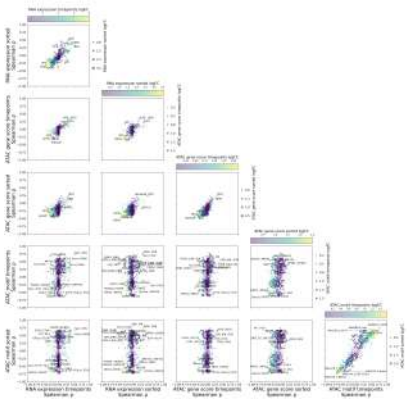


Figure 74: Supplementary: X-plot showing Spearman correlations against pseudotime, for all pairs of assays

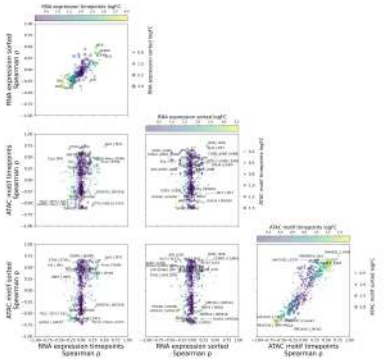


Figure 75: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping sorted and timepoints assays for RNA and gene score

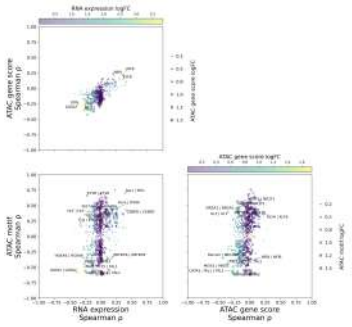


Figure 76: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping all RNA and chromVAR assays

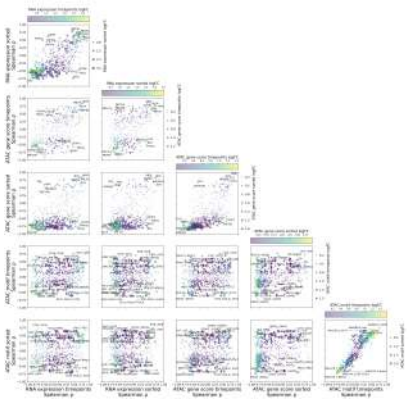


Figure 77: Supplementary: X-plot showing Spearman correlations against pseudotime, for all pairs of assays, with MAGIC smoothing for gene score and RNA expression

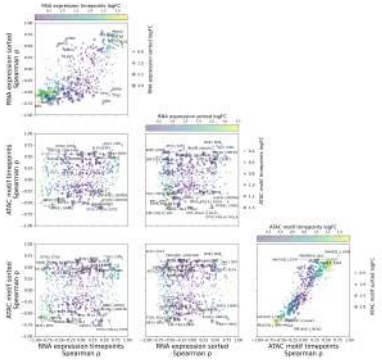


Figure 78: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping sorted and timepoints assays for RNA and gene score, with MAGIC smoothing for gene score and RNA expression

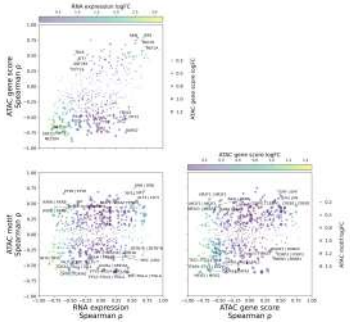


Figure 79: Supplementary: X-plot showing Spearman correlations against pseudotime, grouping all RNA and chromVAR assays with MAGIC smoothing for gene score and RNA expression

7 Early populations

7.1 Top markers from lineage tracing

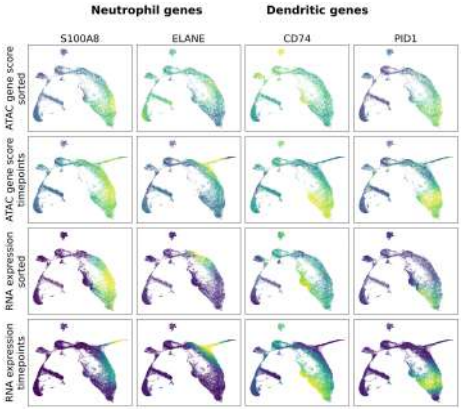


Figure 80: Supplementary: Marker genes previously identified by lineage tracing distinguish early monocyte populations: dendritic-like and neutrophil-like

7.2 Early populations classification by backpropagating labels

7.2.1 UMAP

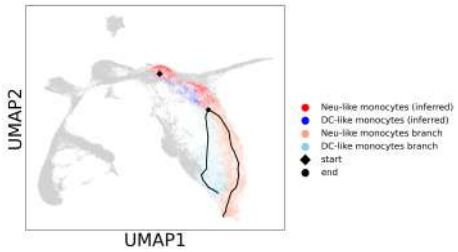


Figure 81: Supplementary: UMAP showing inferred dendritic-like and neutrophil-like early monocyte populations

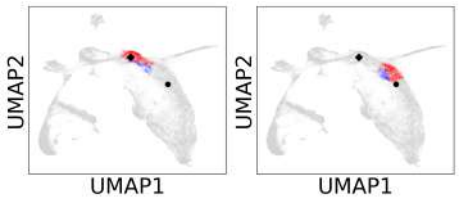


Figure 82: Supplementary: UMAP showing inferred dendritic-like and neutrophil-like early monocyte populations, split into "early" and "late" halves of pseudotime.

7.2.2 Volcanos

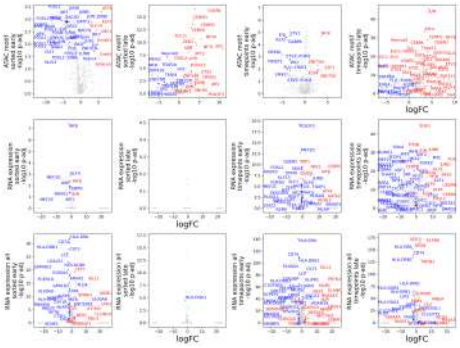


Figure 83: Supplementary: Volcano plots of Wilcoxon test between inferred dendritic-like and neutrophil-like early monocyte populations, for "early" and "late" halves of pseudotime. Motifs and RNA expression show TFs only while RNA expression all shows all genes

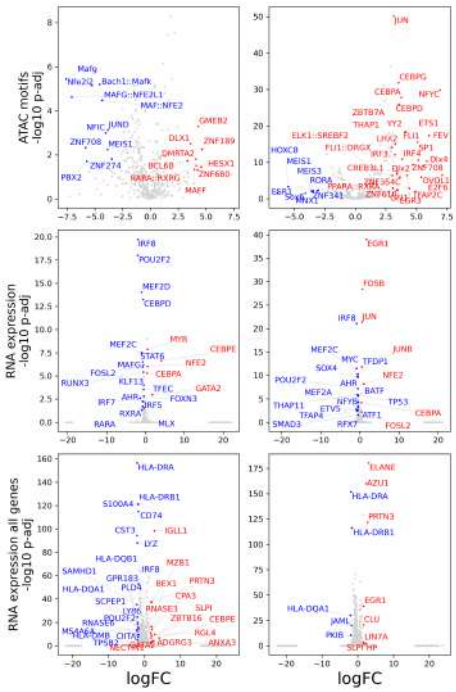


Figure 84: Supplementary: Volcano plots of Wilcoxon test between inferred dendritic-like and neutrophil-like early monocyte populations, for "early" and "late" halves of pseudotime, grouping assays. Motifs and RNA expression show TFs only while RNA expression all shows all genes

7.3 Early populations classification using known lineage tracing markers

7.3.1 UMAP

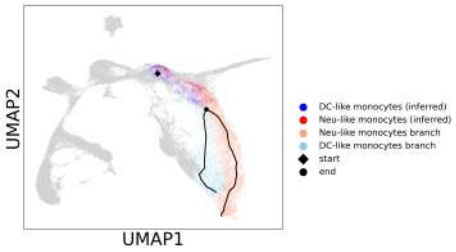


Figure 85: Supplementary: UMAP showing inferred dendritic-like and neutrophil-like early monocyte populations

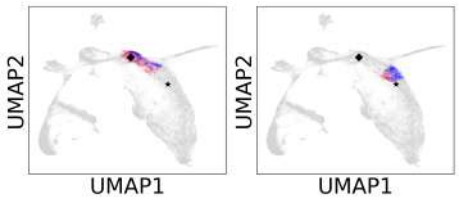


Figure 86: Supplementary: UMAP showing inferred dendritic-like and neutrophil-like early monocyte populations, split into "early" and "late" halves of pseudotime.

7.3.2 Volcanos

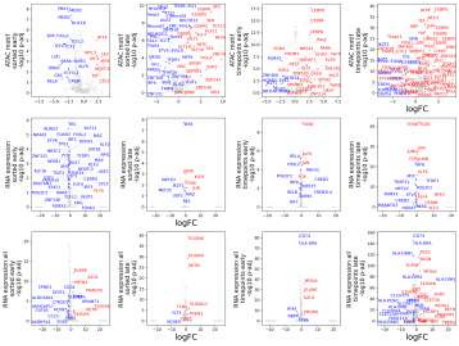


Figure 87: Supplementary: Volcano plots of Wilcoxon test between inferred dendritic-like and neutrophil-like early monocyte populations, for "early" and "late" halves of pseudotime. Motifs and RNA expression show TFs only while RNA expression all shows all genes