Jesus Garcia Garcia
HINF 5650 Final Report
Due 05/12/2021
Professor Dr. Jinhua Wang

To joint scRNA-seq and scATAC-seq PBMC data, I first performed peak calling using macs2. I then performed transformation and PCA analysis on the scRNA-seq data. For the scATAC-seq data, I grabbed the peaks, found the top features, computed the term-frequency inverse-document-frequency, and then ran singular value decomposition. Once the data was processed, I annotated the cells using the PBMC reference dataset from Hao et al. (2020). After the annotation, the datasets are joined and generate a UMAP with annotated clusters. Once the datasets are merged, a signal, motif, footprint, and pseudotime analysis were performed.

*All code used in this project is located in **garcia_FINAL_code** folder.

Study Design

Pre-analysis and software preparation

Step 1) Install **cell ranger arc 1.0.1** from 10x genomics

```
#!/bin/bash -l
#SBATCH --time=01:00:00
#SBATCH --ntasks=4
#SBATCH --mem=16g
#SBATCH --mail-type=ALL
#SBATCH --mail-user=garci624@umn.edu

cd /home/aventeic/garci624/local/src/lsoft/10xGen

# download and install cellranger arc 1.0.1 from 10x genomics
curl -o cellranger-arc-1.0.1.tar.gz .... [KEY]

tar -xzvf cellranger-arc-1.0.1.tar.gz
```

Step 2) Download pre-built **cell ranger arc GRCh38** reference from 10x genomics

```
#!/bin/bash -l
#SBATCH --time=01:00:00
#SBATCH --ntasks=4
#SBATCH --mem=16g
#SBATCH --mail-type=ALL
#SBATCH --mail-user=garci624@umn.edu

cd /home/aventeic/garci624/local/src/ref

# download and install cellranger arc GRCh38-2020 reference from 10x genomics
curl -O https://cf.10xgenomics.com/supp/cell-arc/refdata-cellranger-arc-GRCh38-2020-A.tar.gz

tar -xzvf refdata-cellranger-arc-GRCh38-2020-A.tar.gz
```

Step 3) Download **sequencing data and csv** files from 10x genomics

```
#!/bin/bash -l
#SBATCH --time=04:00:00
#SBATCH --ntasks=4
#SBATCH --mem=16g
#SBATCH --mail-type=ALL
#SBATCH --mail-user=garci624@umn.edu

cd /scratch.global/garci624/Single_Cell_Multiome/data

# sequencing data (FASTQ - 95.6 GB)
curl -O https://s3-us-west-2.amazonaws.com/10x.files/samples/cell-arc/1.0.0/pbmc_granulocyte_sorted_10k/pbmc_granulocyte_sorted_10k_fastqs.tar

# untar pbmc_granulocyte_sorted_10k_fastqs
tar -xvf pbmc_granulocyte_sorted_10k_fastqs.tar

# library (CSV - 220 B  )
curl -O https://cf.10xgenomics.com/samples/cell-arc/1.0.0/pbmc_granulocyte_sorted_10k/pbmc_granulocyte_sorted_10k_library.csv
```

*__Step_1.sh, Step_2.sh and Step_3.sh__ are located in **garcia_FINAL_code** folder

Step 4) Running **cell ranger arc 1.0.1 (count)** on 10x multiomics data

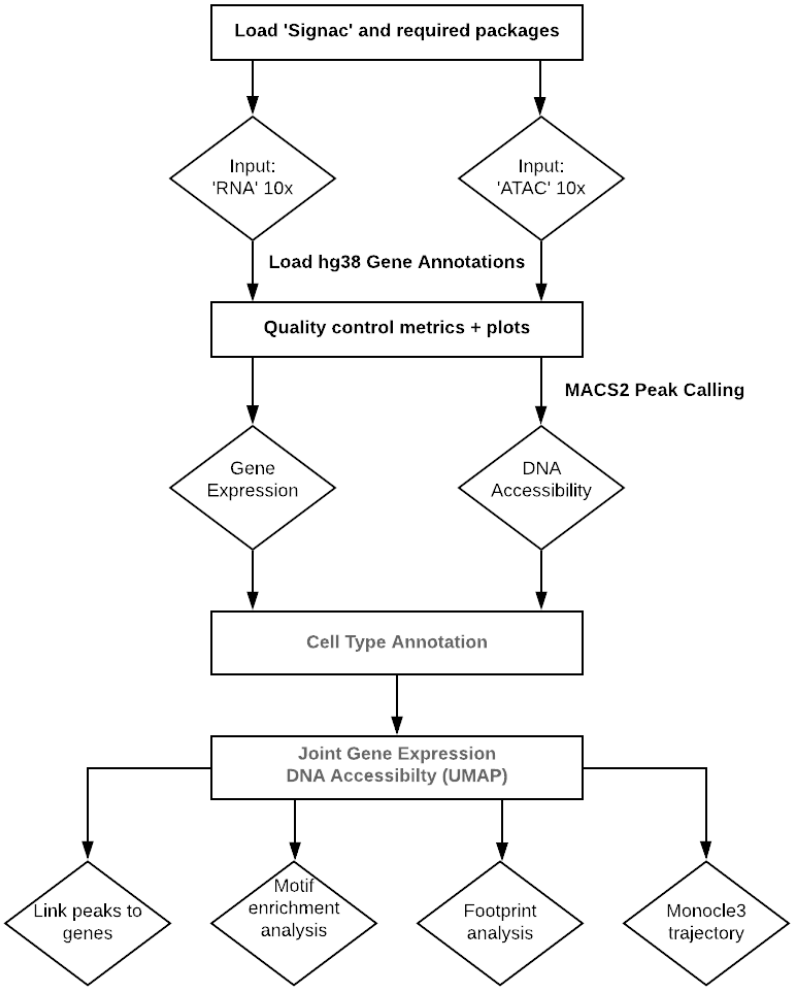
```
#!/bin/bash -l
#SBATCH --time=16:00:00
#SBATCH --ntasks=16
#SBATCH --mem=64g
#SBATCH --mail-type=ALL
#SBATCH --mail-user=garci624@umn.edu

cd /scratch.global/garci624/Single_Cell_Multiome

export PATH=/home/aventeic/garci624/local/src/lsoft/10xGen/cellranger-arc-1.0.1:$PATH

cellranger-arc count --id=10xMultiomic \
--reference=/home/aventeic/garci624/local/src/ref/refdata-cellranger-arc-GRCh38-2020-A-1.0.0 \
--libraries=/scratch.global/garci624/Single_Cell_Multiome/data/pbmc_granulocyte_sorted_10k_library.csv \
--localcores=16 \
--localmem=64
```

Running downstream analyses on 10x PBMC Multiomic data using 'Signac'



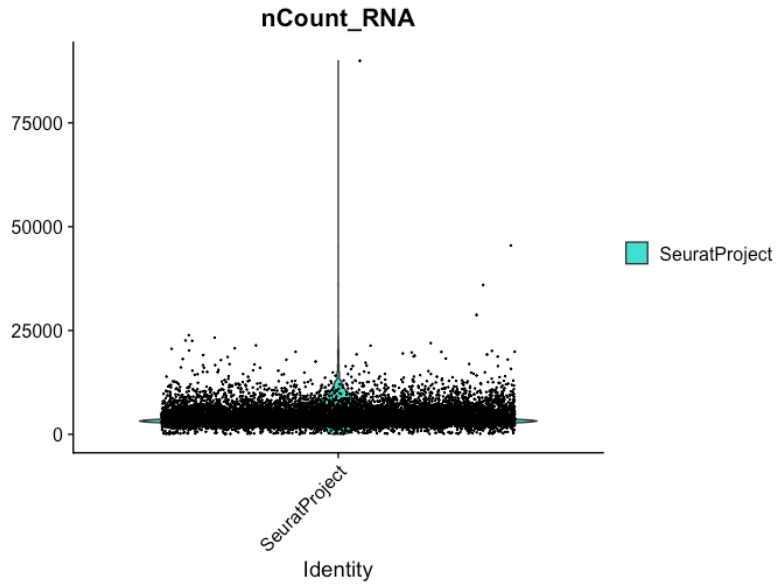
---

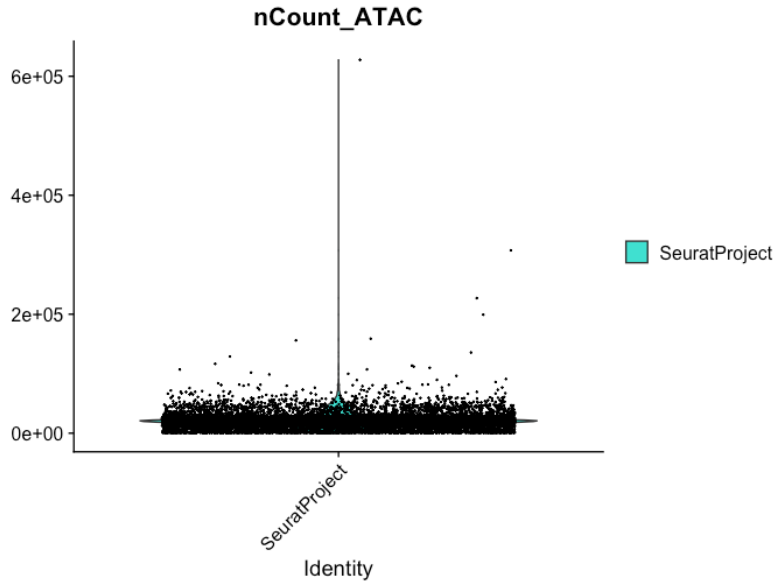
*__10xPBMC_Multiomic.R__ are located in **garcia_FINAL_code** folder

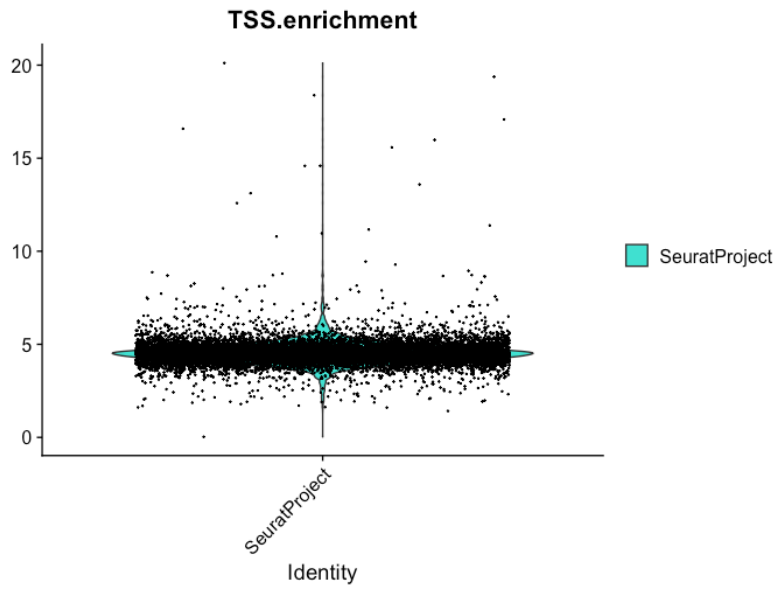
Final Results

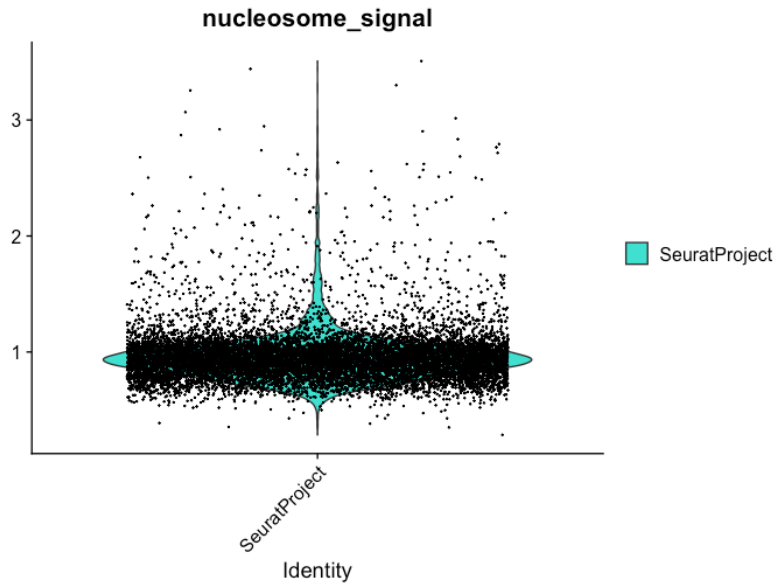
Quality Control plots

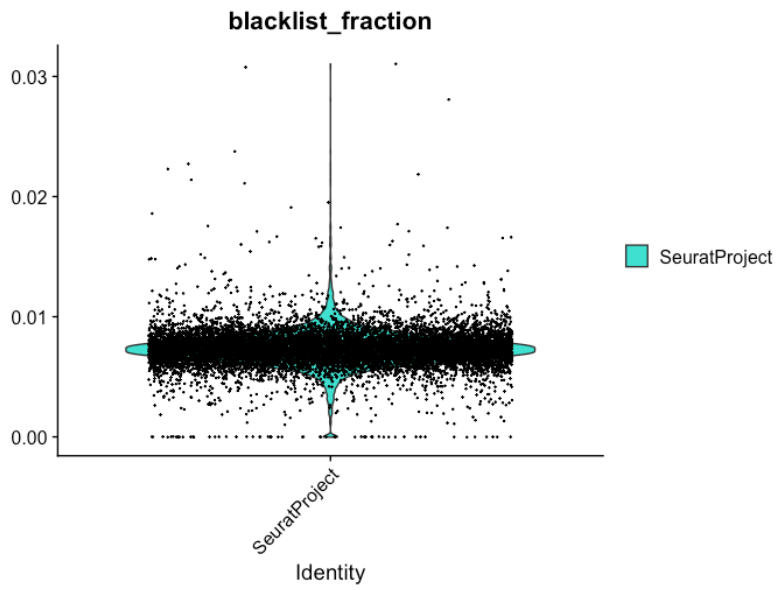
Five quality control violin plots were generated to remove low quality cells. The filter applied was nCount_RNA < 25000 & nCount_RNA > 1000 & nCount_ATAC < 100000 & nCount_ATAC > 1000 & TSS.enrichment > 1 & nucleosome_signal < 2 & blacklist_fraction < 0.015.



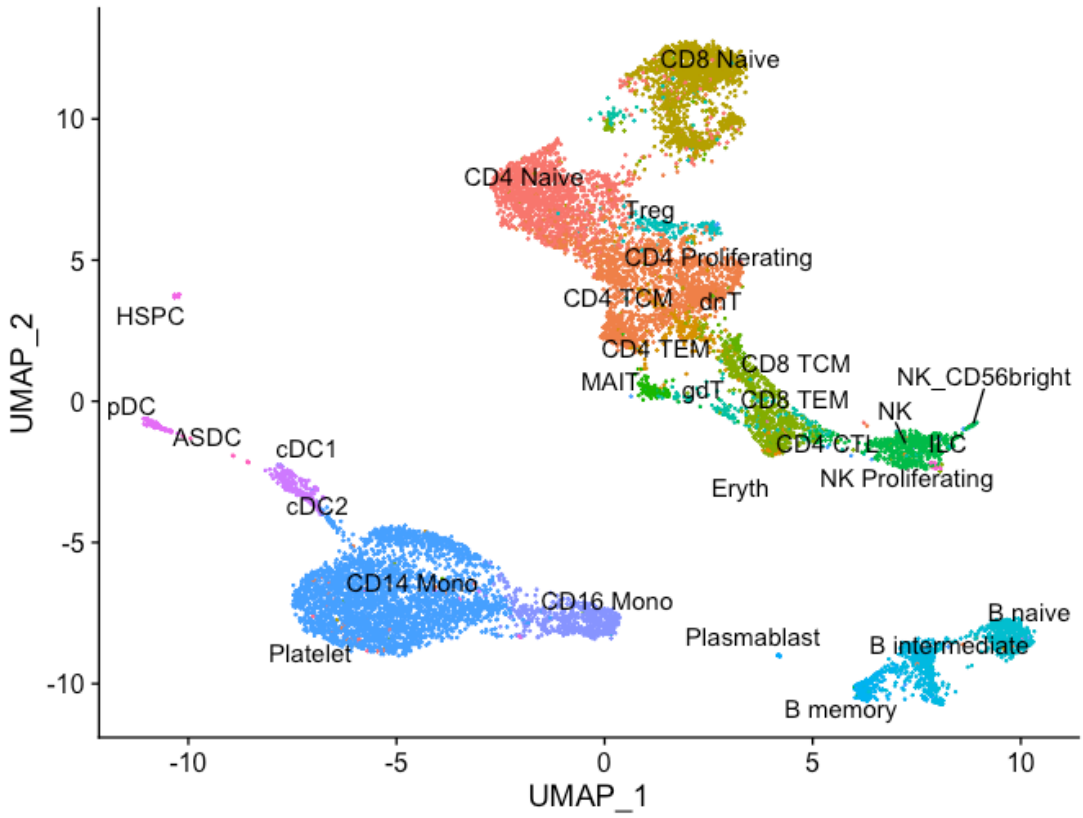



TSS.enrichment
20
15
10
5
0
SeuratProject
Identity
SeuratProject

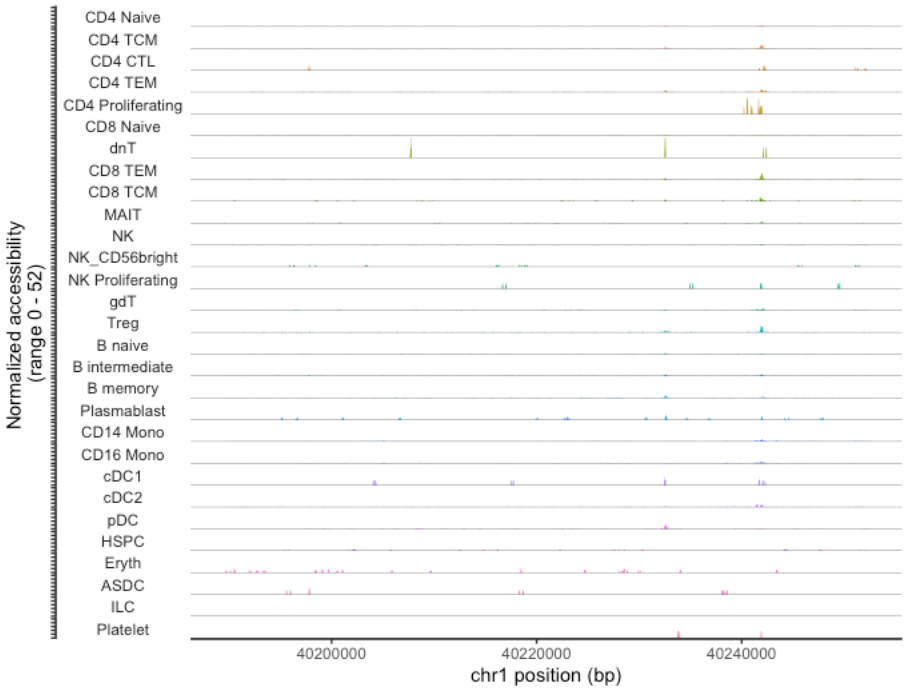

nucleosome_signal
3
2
1
SeuratProject
Identity
SeuratProject



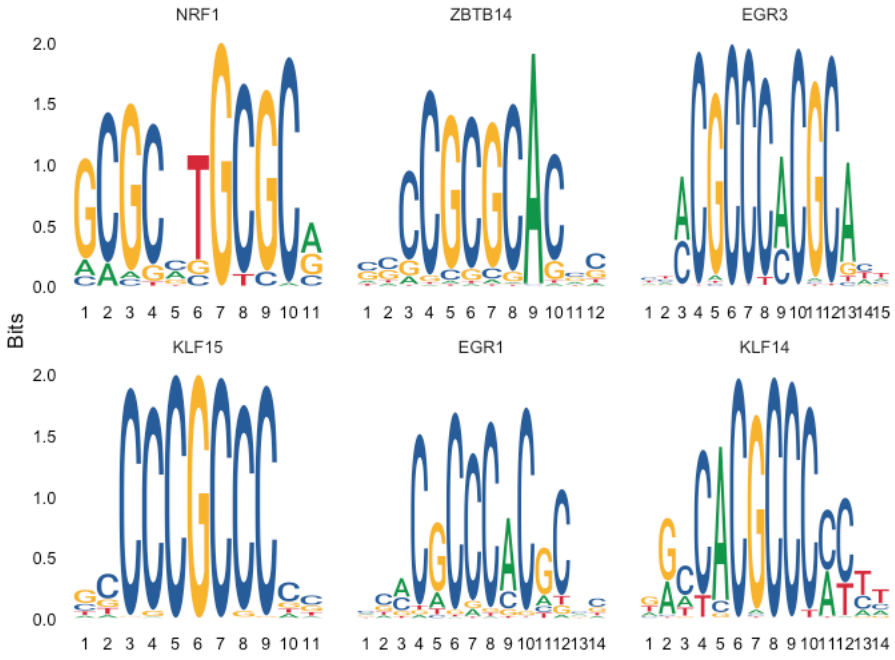
Joint UMAP visualization (scRNA-seq + scATAC-seq Multiomics analsysis)



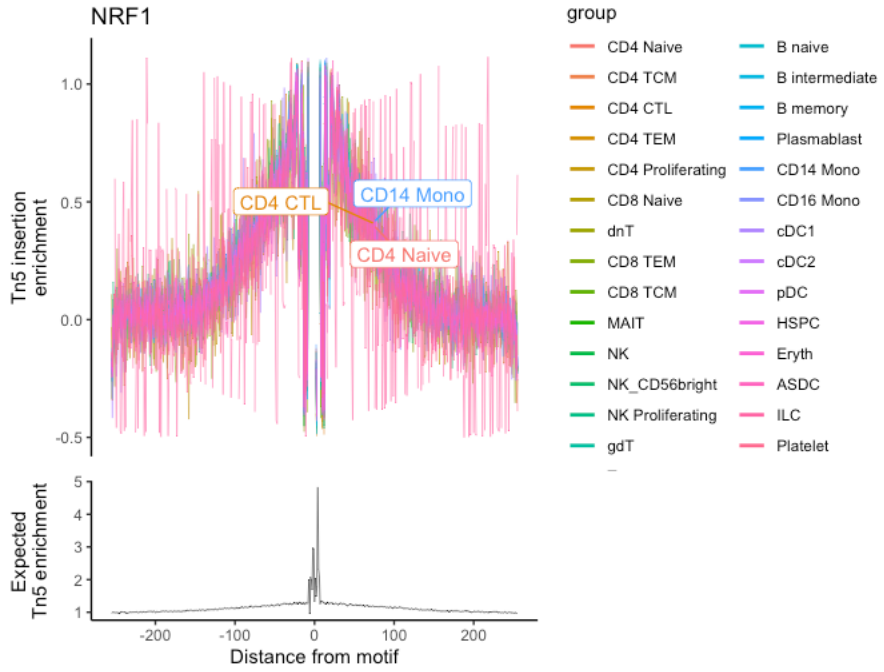
Peak Signal Analysis

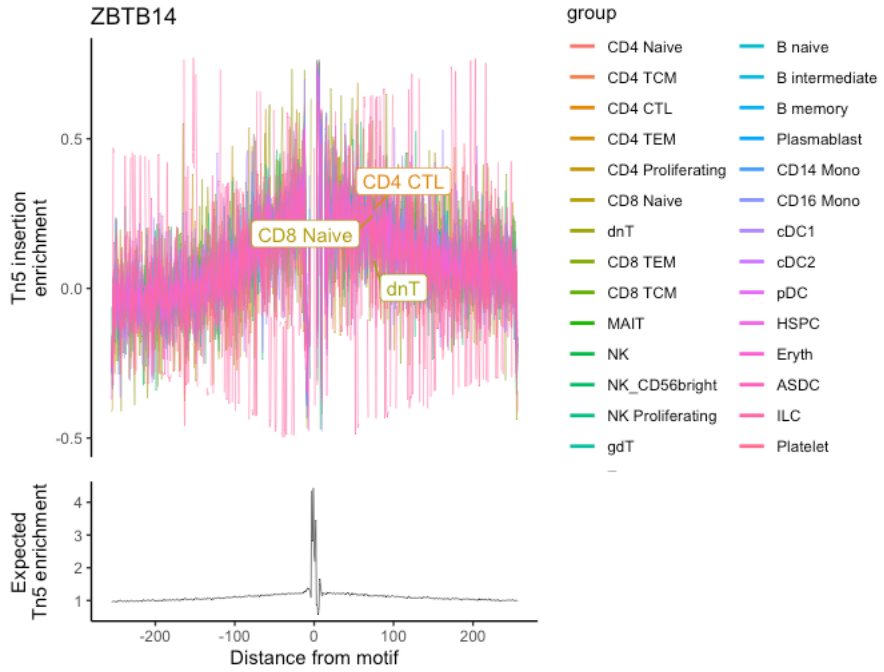


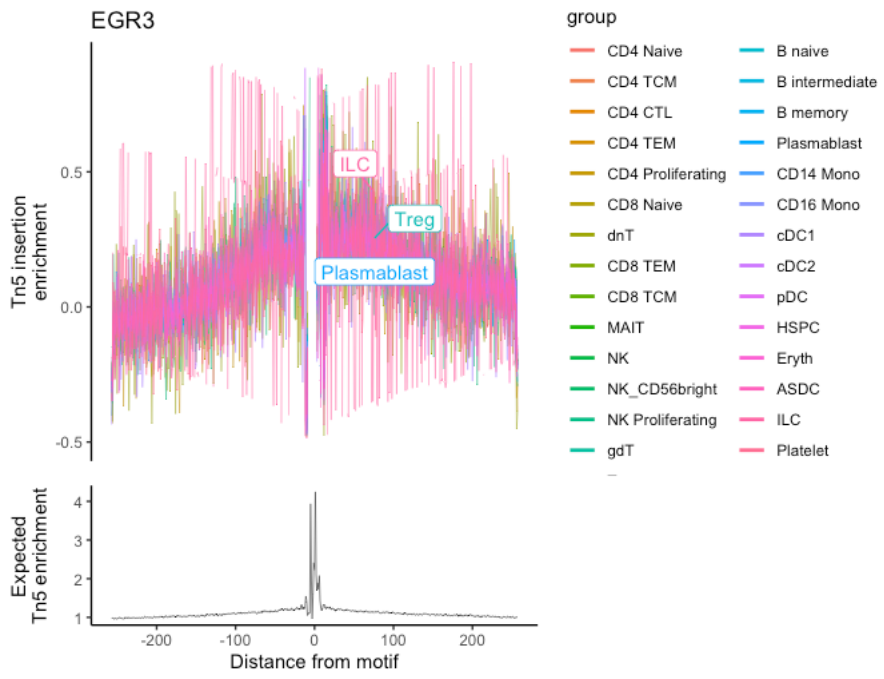
MOTIF Analysis

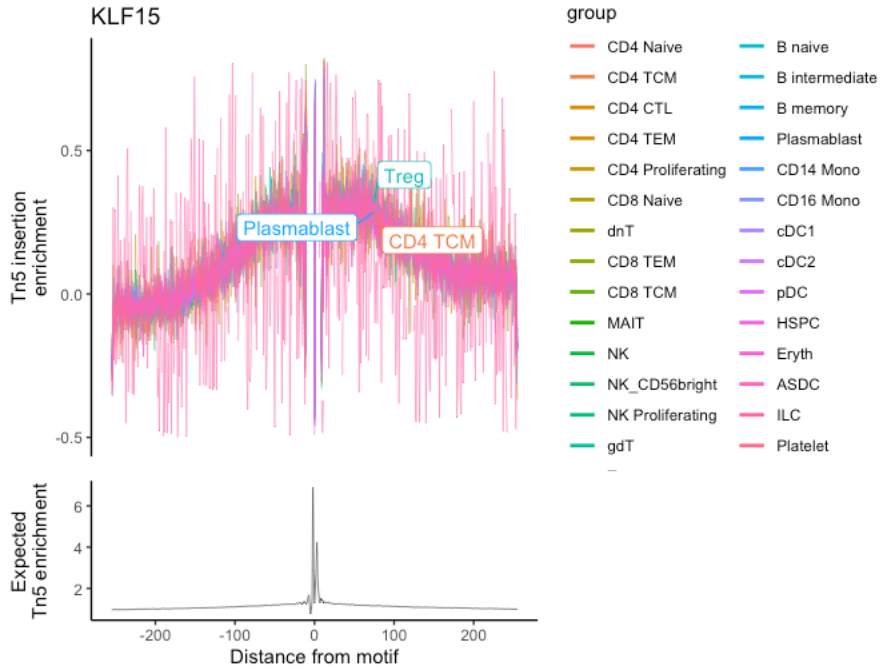


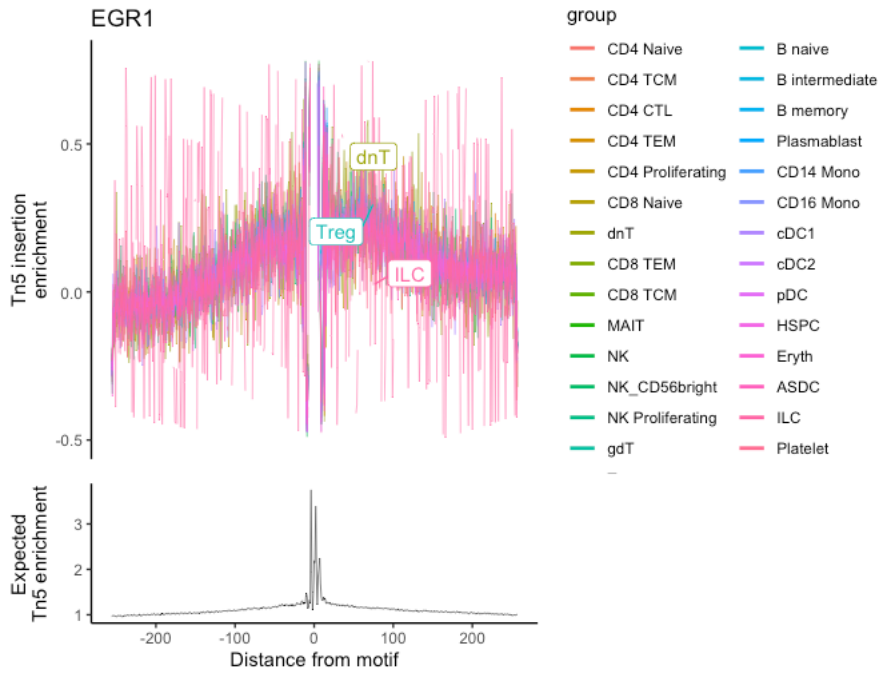
FOOTPRINT analysis

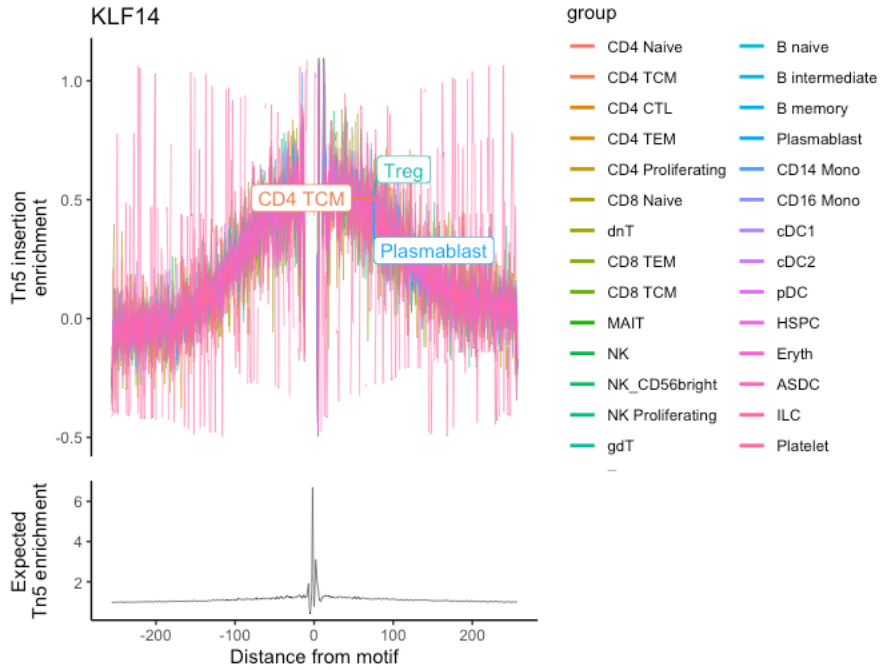




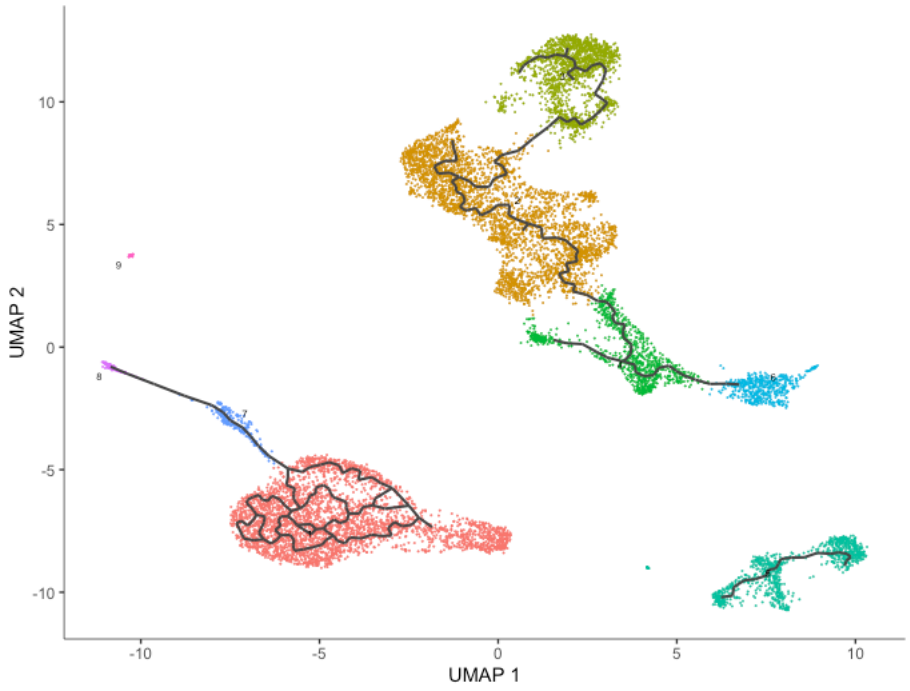

EGR3
Tn5 insertion enrichment
0.5
0.0
-0.5
ILC
Treg
Plasmablast
Expected Tn5 enrichment
4
3
2
1
-200
-100
0
100
200
Distance from motif
group
CD4 Naive
CD4 TCM
CD4 CTL
CD4 TEM
CD4 Proliferating
CD8 Naive
dnT
CD8 TEM
CD8 TCM
MAIT
NK
NK_CD56bright
NK Proliferating
gdT
B naive
B intermediate
B memory
Plasmablast
CD14 Mono
CD16 Mono
cDC1
cDC2
pDC
HSPC
Eryth
ASDC
ILC
Platelet

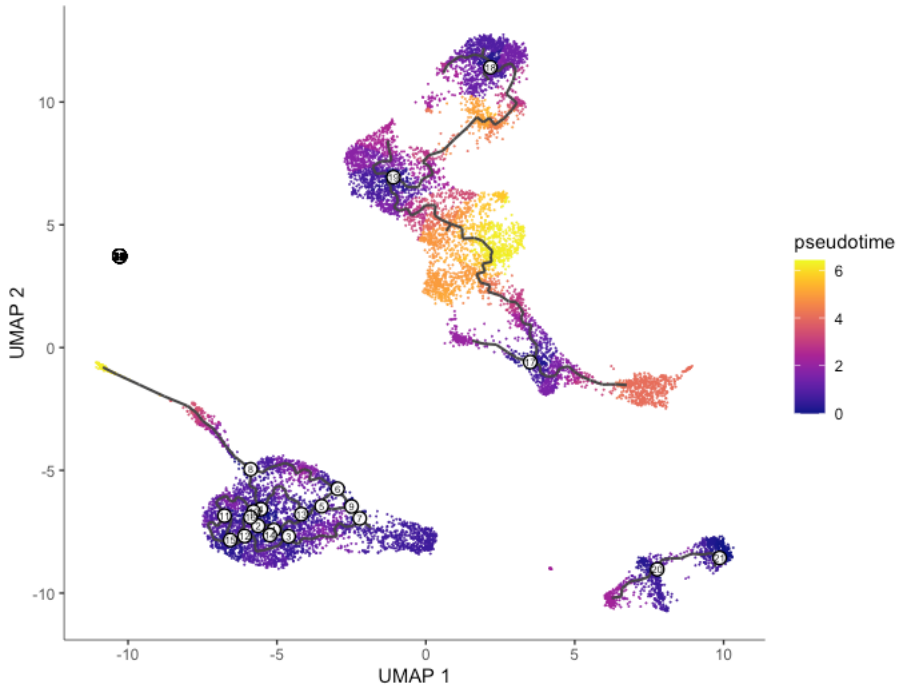

KLF15
Tn5 insertion enrichment
0.5
0.0
-0.5
Treg
Plasmablast
CD4 TCM
Expected Tn5 enrichment
6
4
2
-200
-100
0
100
200
Distance from motif
group
CD4 Naive
CD4 TCM
CD4 CTL
CD4 TEM
CD4 Proliferating
CD8 Naive
dnT
CD8 TEM
CD8 TCM
MAIT
NK
NK_CD56bright
NK Proliferating
gdT
B naive
B intermediate
B memory
Plasmablast
CD14 Mono
CD16 Mono
cDC1
cDC2
pDC
HSPC
Eryth
ASDC
ILC
Platelet


EGR1
Tn5 insertion enrichment
0.5
0.0
-0.5
dnT
Treg
ILC
Expected Tn5 enrichment
3
2
1
-200
-100
0
100
200
Distance from motif
group
CD4 Naive
CD4 TCM
CD4 CTL
CD4 TEM
CD4 Proliferating
CD8 Naive
dnT
CD8 TEM
CD8 TCM
MAIT
NK
NK_CD56bright
NK Proliferating
gdT
B naive
B intermediate
B memory
Plasmablast
CD14 Mono
CD16 Mono
cDC1
cDC2
pDC
HSPC
Eryth
ASDC
ILC
Platelet


KLF14
Tn5 insertion enrichment
1.0
0.5
0.0
-0.5
Treg
CD4 TCM
Plasmablast
Expected Tn5 enrichment
6
4
2
-200
-100
0
100
200
Distance from motif
group
CD4 Naive
CD4 TCM
CD4 CTL
CD4 TEM
CD4 Proliferating
CD8 Naive
dnT
CD8 TEM
CD8 TCM
MAIT
NK
NK_CD56bright
NK Proliferating
gdT
B naive
B intermediate
B memory
Plasmablast
CD14 Mono
CD16 Mono
cDC1
cDC2
pDC
HSPC
Eryth
ASDC
ILC
Platelet

PSEUDOTIME analysis





References

[1] Tim Stuart and Avi Srivastava and Caleb Lareau and Rahul Satija. **Multimodal single-cell chromatin analysis with Signac**. bioRxiv. 2020. https://doi.org/10.1101/2020.11.09.373613

[2] Trapnell, C., Cacchiarelli, D., Grimsby, J. et al. **The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells.** Nat Biotechnol 32, 381–386 (2014). https://doi.org/10.1038/nbt.2859

[3] Butler, A., Hoffman, P., Smibert, P. et al. **Integrating single-cell transcriptomic data across different conditions, technologies, and species.** Nat Biotechnol 36, 411–420 (2018). https://doi.org/10.1038/nbt.4096

[4] **Cell RangerTM R Kit Tutorial: Secondary Analysis on 10x GenomicsTM Single Cell 3' RNAseq PBMC Data. (n.d.). 10x Genomics.** Retrieved July 11, 2017, from http://cf.10xgenomics.com/supp/cell-exp/cellrangerrkit-PBMC-vignette-knitr-1.1.0.pdf

[5] **What is Cell Ranger (2016, November 21).** Retrieved July 11, 2017, from https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/whatis-cell-ranger

[6] **The authors acknowledge the Minnesota Supercomputing Institute (MSI) at the University of Minnesota for providing resources that contributed to the research results reported within this paper.** URL: http://www.msi.umn.edu