What is hidden in the darkness??

(scATAC in non-coding genome)

Daria, Philipp, Sara, Alexander

Organism: Mus musculus
Genome assembly: mm10
Tissue: Cerebral cortex, adult mice p50
Dataset: 10x scATAC

Regions of interest:

- Long non-coding RNAs — Gencode gtf file

- LINEs (long interspersed nuclear elements)
- SINES (short interspersed nuclear elements)
- LTR (long terminal repeats)
- Retroposons
- Simple repeats
- Satellites

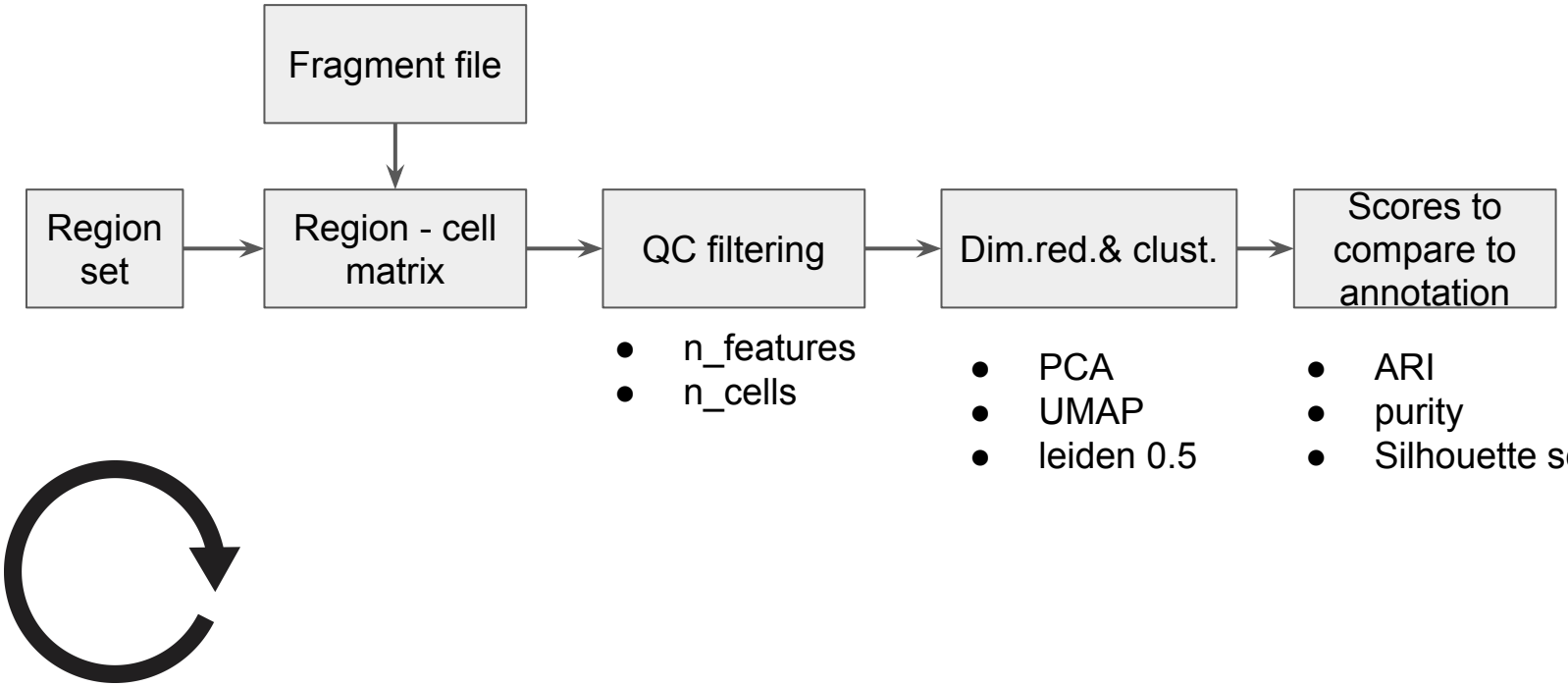
— Repeatmasker mm10

- Enhancers
- Promoters
- Protein coding

— SCOG 2022 files

https://www.10xgenomics.com/resources/datasets/fresh-cortex-from-adult-mouse-brain-p-50-1-standard-1-2-0

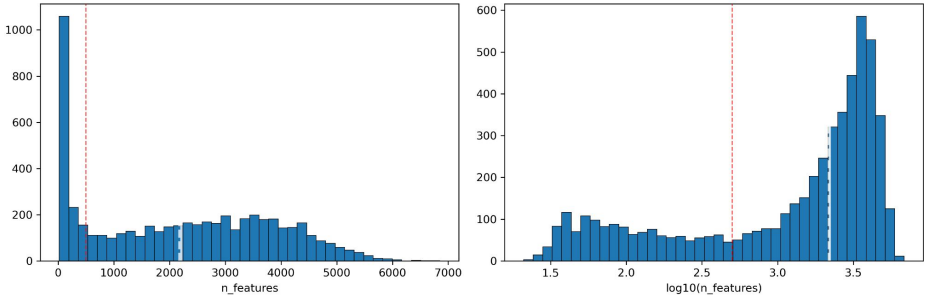
Data processing



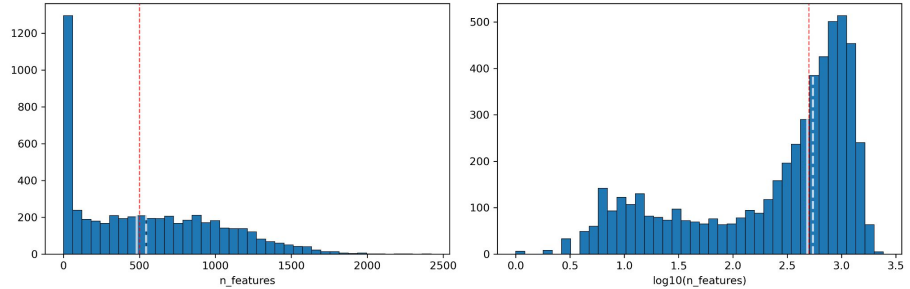
https://github.com/dariarom94/SCOG_noncoding

QC Distributions. Nr of peaks in features.

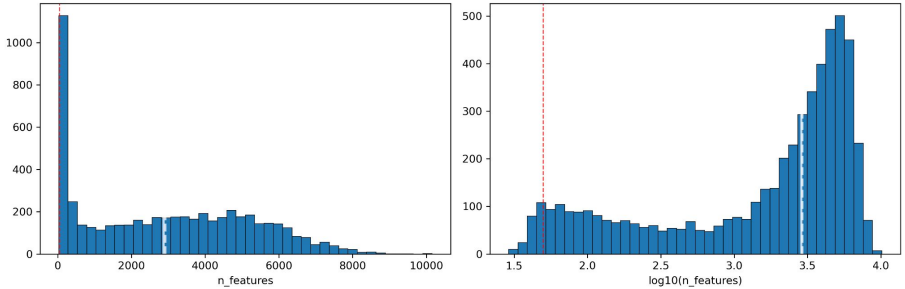
Protein coding (threshold 500)



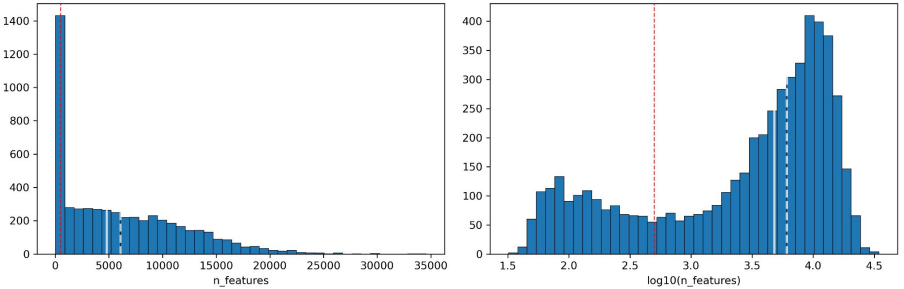
lncRNAs (threshold 500)



Promoters (threshold 500)

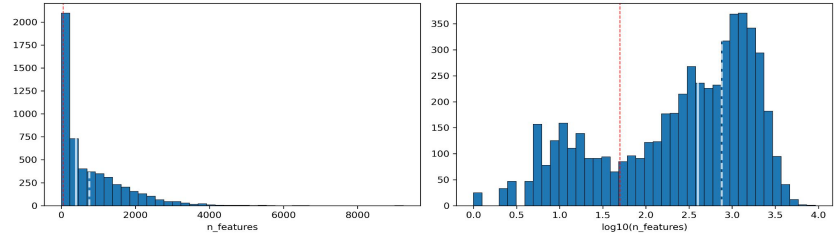


Enhancers (threshold 500)

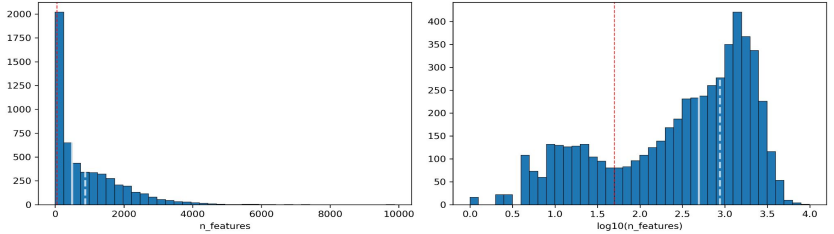


QC Distributions. Nr of peaks in features.

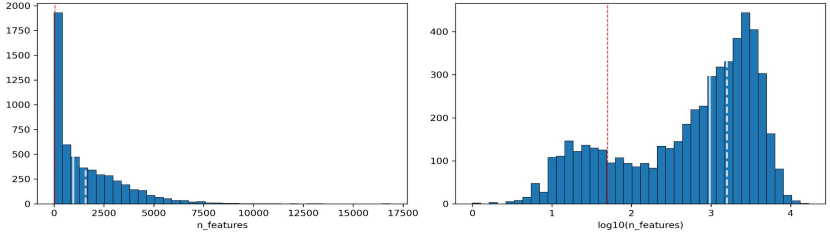
LINEs (threshold 50)



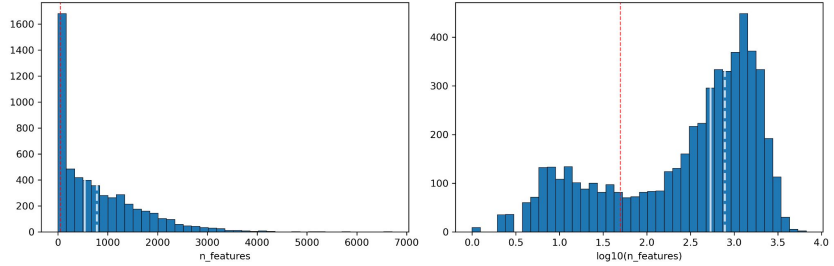
LTRs (threshold 50)



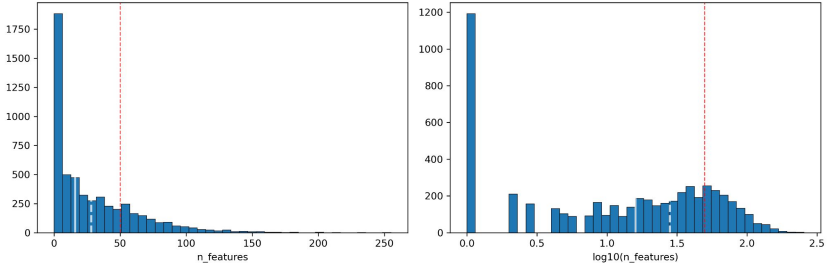
SINEs (threshold 50)



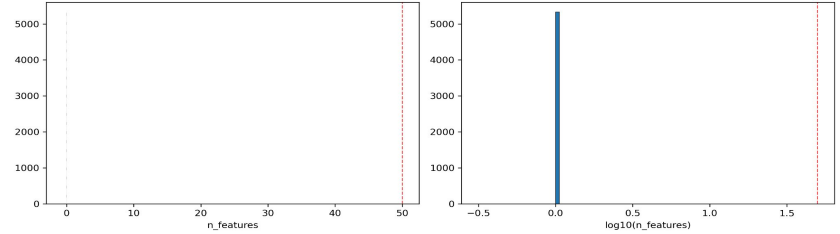
Simple repeats (threshold 50)

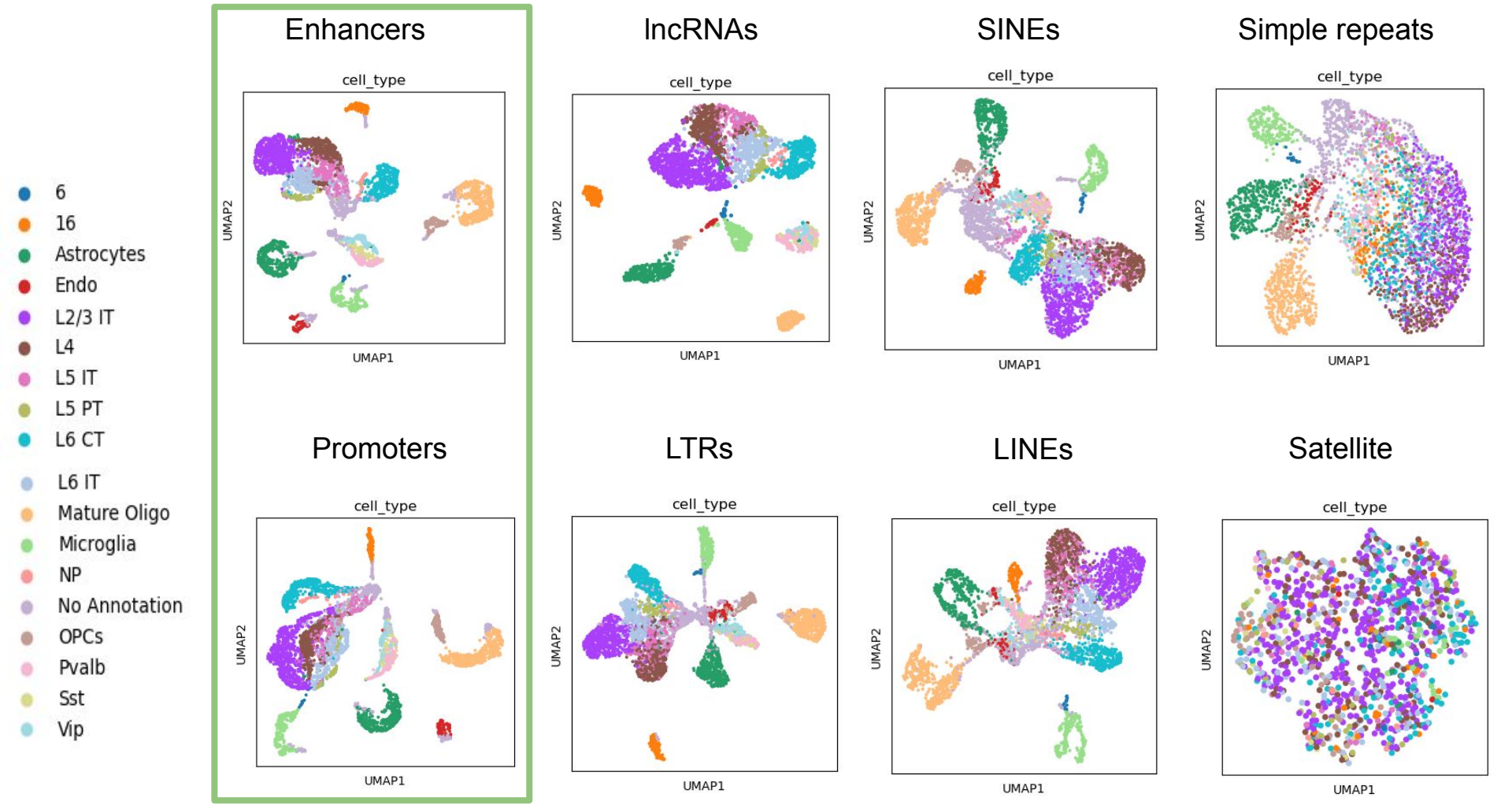


Satellite (threshold 50)

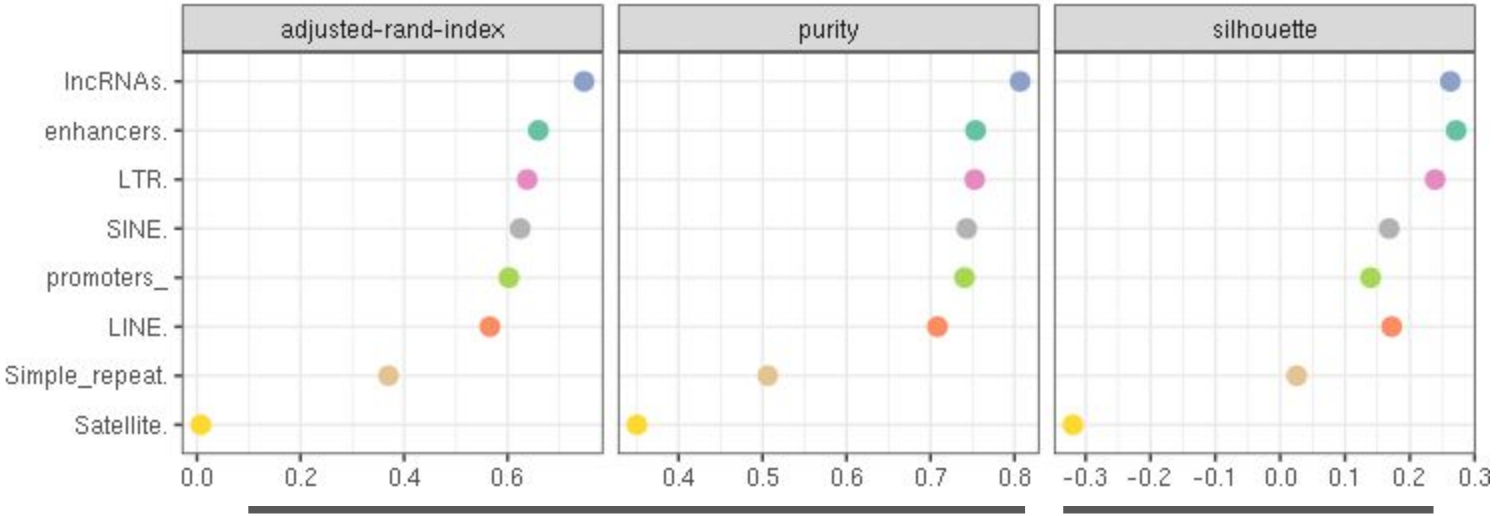


Retroposons (threshold 0)




Enhancers
lncRNAs
SINEs
Simple repeats
Promoters
LTRs
LINEs
Satellite
cell_type
UMAP1
UMAP2
6
16
Astrocytes
Endo
L2/3 IT
L4
L5 IT
L5 PT
L6 CT
L6 IT
Mature Oligo
Microglia
NP
No Annotation
OPCs
Pvalb
Sst
Vip

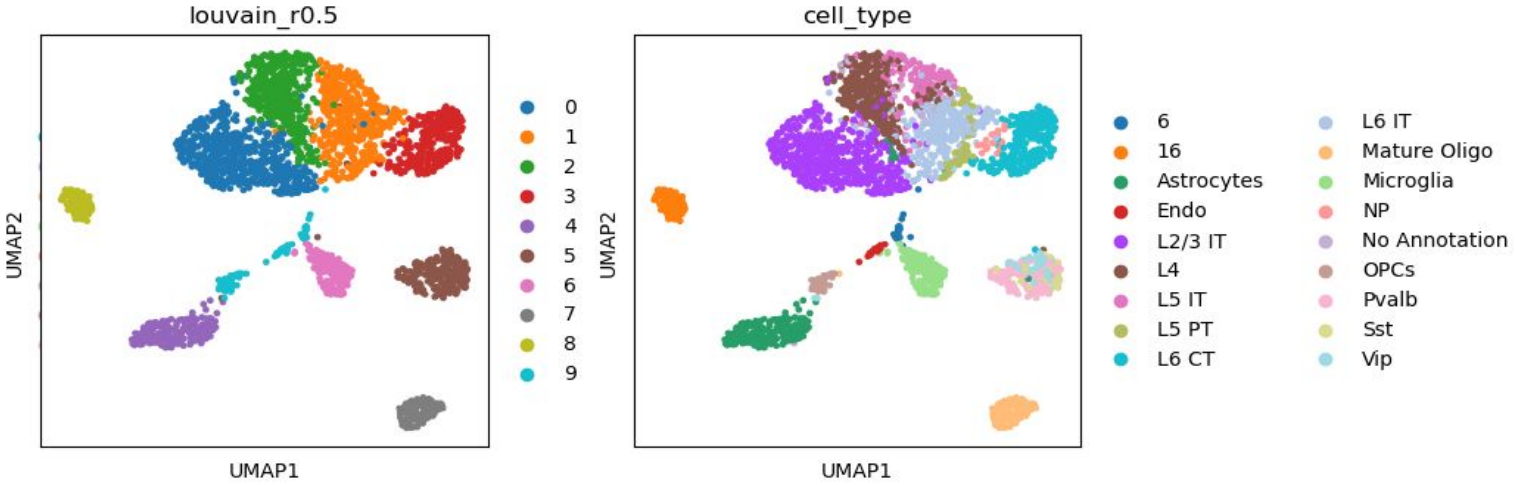
Evaluation metrics



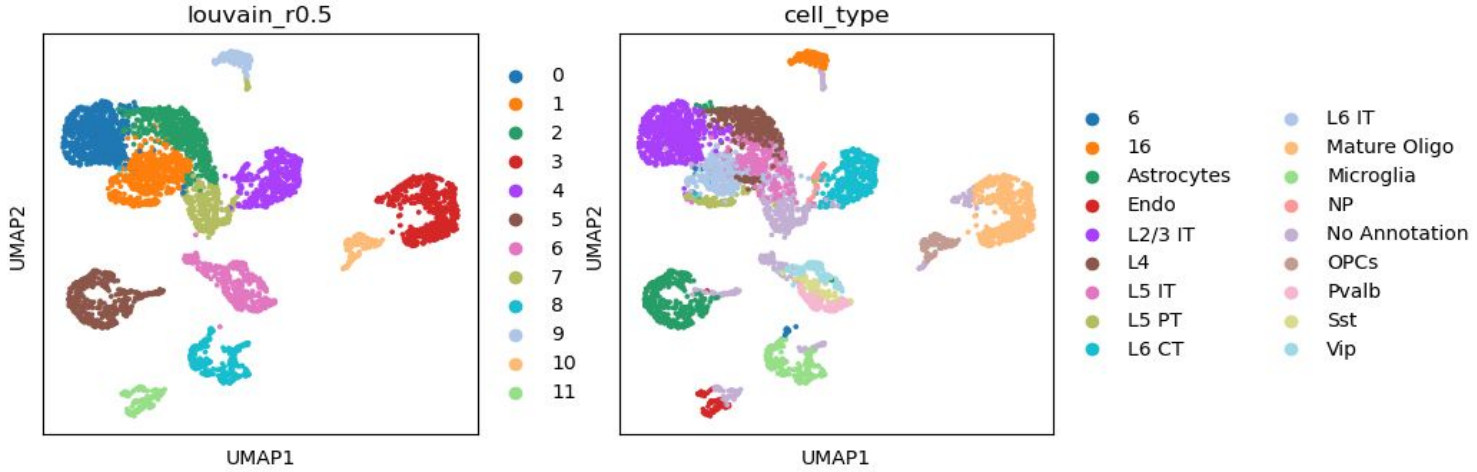
External evaluation:
Louvain clustering (res0.5)
vs cell types (ground truth)

Internal evaluation:
Cell types (ground truth)

lncRNAs



Enhancers



Summary

Chromatin accessibility in non coding region of the genome contains cell-type specific information.

- Enhancers and lncRNAs performs the best in separating two type of neurons and glial cells
- All repetitive elements are doing a good job in separating neuronal and glial cell types, but usually not able to separate glutamatergic and GABAergic neurons
- LTRs shows the best clustering for both neuronal and glial among all repetitive elements.
- Chromatin accessibility on satellite DNA does not seem to be cell type specific