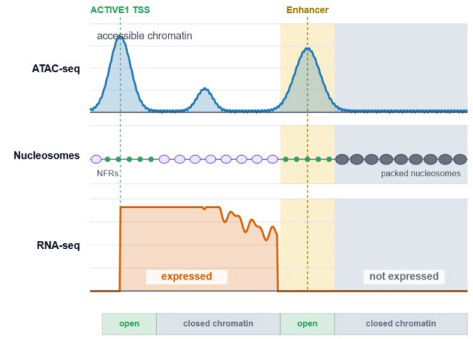
Assessing the Added Value of scATAC-seq Data in Cell-Specific Gene Regulatory Network Inference for Alzheimer's Disease

Leo Lin
chenchenlin@tudelft.nl
Professor: Marcel Reinders
Supervisor: Bram Pronk, Timo Verlaan

1. Background

Alzheimer's disease (AD) is a progressive neurodegenerative disorder and the most common cause of dementia.
Gene regulatory networks (GRNs) represent how genes influence each other as directed graphs.
Single-cell technologies (sc) measure gene activity in individual cells, enabling cell-specific GRNs.

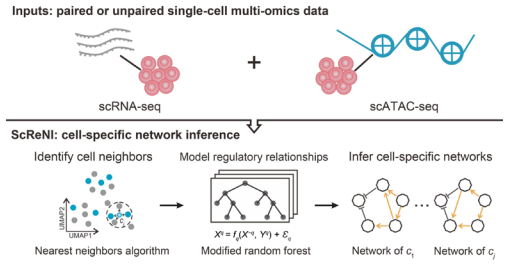
ScReNI is a machine learning method that integrates two data types:
- **scRNA-seq** measures gene expression
- **scATAC-seq** measures chromatin accessibility



2. Research Question

Does adding chromatin accessibility improve cell-specific GRN inference beyond RNA-only ScReNI?

3. ScReNI Method



Adapted from Xu et al., 2025, ScReNI workflow schematic.

4. Research Method

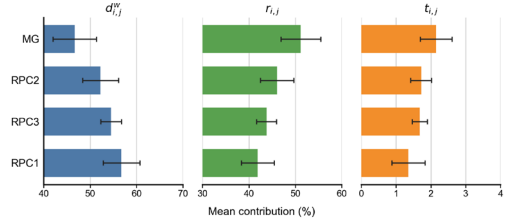
ScReNI has two network variants:
- **kScReNI** only uses RNA-seq
- **wScReNI** adds ATAC-seq peak accessibility features

During network construction, ScReNI converts random-forest feature importance scores into directed regulatory edge weights. Each edge represents the inferred influence of regulator gene j on target gene i:

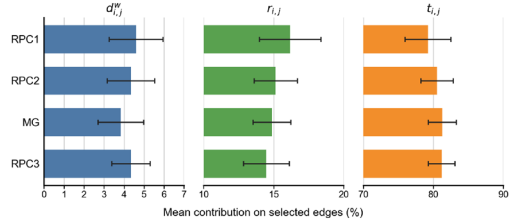
$$\boldsymbol{w_{i,j} = d^{w}_{i,j} + r_{i,j} + t_{i,j}}$$

$\boldsymbol{d^{w}_{i,j}}$ = expression importance of regulator j when predicting target i
$\boldsymbol{r_{i,j}}$= accessibility signal near regulator j (broad regulator-locus support)
$\boldsymbol{t_{i,j}}$= motif-gated target-peak attribution (sparse target-specific chromatin support)

In this research we have evaluated four different variations to this formula.

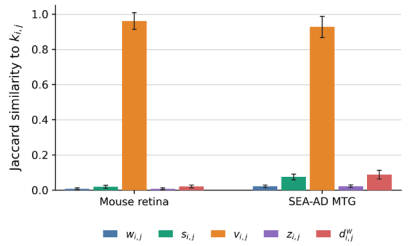


Expression importance and regulator-locus accessibility both contribute strongly to total wScReNI weight, whereas the target-peak attribution is sparse globally.



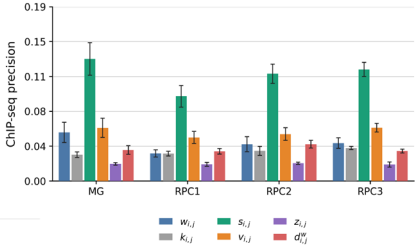
After restricting to edges with active motif-gated target-peak attribution, the target-peak term dominates selected-edge weight.

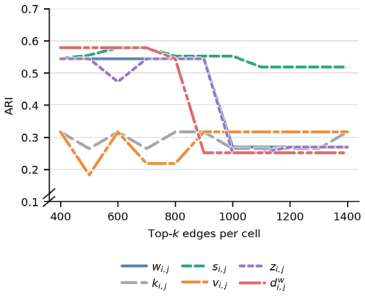
5. Results



Top-500 edge-set Jaccard similarity between each formula variant and RNA-only kScReNI.
Only the mixed diagnostric variant which uses the RNA-only base weights stays close to kScReNI. This shows that wScReNI and all formula variants replace the RNA-only topology instead of small chromatin correction to kScReNI.

Mean ChIP-Atlas precision at top-500 predicted edges.
The target-specific variant $s_{i,j}$ is highest in every retinal cell type, indicating that its top-ranked edges have the strongest overlap with external ChIP-supported binding evidence.





Network-based clustering ARI across top-k edge thresholds.
The high ARI score for the target-specific formula $s_{i,j}$ shows that removing the regulator-locus term does not remove the network-level cell-type signal.

SEA-AD supports component-pattern analysis, not yet strong AD regulatory-discovery claims.
The microglia-specific HVG experiment shows that the checked AD-relevant regulator genes are removed before the GRN inference step. Future AD-focused ScReNI pipelines should force-retain disease-relevant TFs and validate against matched human brain TF-binding evidence.

6. Conclusion

The target-specific ATAC attribution formula variant

$$\boldsymbol{s_{i,j} = d^{w}_{i,j} + t^{*}_{i,j}}$$

improved mouse retina ChIP-supported precision by 2.8x over published wScReNI and 3.6x over RNA-only kScReNI, while preserving cell-type network structure. For AD use, feature selection and matched validation remain the key bottlenecks.