Stability of Cell-Specific Gene Regulatory Networks Inferred by ScReNI

Why Out-of-Bag Accuracy Falls Short and Edge Weight Variance Shows Promise

Eduard Cimpean

Supervisors: Marcel Reinders, Bram Pronk, Timo Verlaan

EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 21, 2026

Name of the student: Eduard Cimpean
Final project course: CSE3000 Research Project
Thesis committee: Marcel Reinders, Sicco Verwer, Bram Pronk, Timo Verlaan

Abstract

Gene regulatory networks (GRNs) inferred at single-cell resolution offer insight into regulatory mechanisms underlying complex diseases such as Alzheimer's disease, but the reliability of the GRNs produced by methods such as ScReNI is not fully explored. This work investigates the stability and reliability of GRNs inferred by a Python reimplementation of ScReNI applied to the Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD) and a mouse retinal dataset. Three experiments were conducted. First, gene-level prediction accuracy, measured via random forest out-of-bag (OOB) R^2, was found to vary across genes and to be driven primarily by expression level rather than network position. Second, two main candidate cell-level reliability metrics were evaluated against precision and recall: OOB R^2 showed no significant association with precision once network density and expression were statistically controlled for, whereas edge weight variance across repeated inference runs retained significant correlation with both, making it the more promising candidate metric identified. Third, comparing the GRNs of cells with high versus low edge weight variance within cell types revealed that such networks share similar topological structure but diverge substantially in edge weight magnitude. These findings indicate that out-of-bag accuracy alone is an insufficient proxy for GRN reliability, and that edge weight variance, while more informative, does not yet provide a fully cell-type-consistent reliability metric, motivating further work on alternative approaches to quantifying single-cell GRN stability.

1 Introduction

Gene regulatory networks (GRNs) are representations of the regulatory relationships between different genes, where the edges show the influence of transcription factors (TFs) on the expression levels of the target genes. Understanding these networks is essential for uncovering the molecular mechanisms underlying complex diseases such as Alzheimer's disease, which affects over 55 million people worldwide [8] and remains without a definitive cure. In order to infer these networks, the gene activity across cells must be measured, which is achieved through RNA sequencing (RNA-seq) technology. Recent technology allows for RNA sequencing at the single-cell resolution (scRNA-seq) [4], measuring gene expression at the individual cell level. This level of detail reveals information that the bulk methods cannot provide: scRNA-seq enables the identification of "complex and rare cell populations, uncover regulatory relationships between genes, and track the trajectories of distinct cell lineages in development" [5] [9].

Complementing scRNA-seq in this paper, single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq) measures chromatin accessibility. By studying which regions of the genome are physically open, we get a more direct view into which regulatory relationships are possible within a given cell. Methods that use both modalities can capture both the transcriptional output and the regulatory potential of a cell at the same time [6].

ScReNI [10] is a recently developed algorithm that infers cell-specific GRNs using both scRNA-seq and scATAC-seq at the individual cell level. It recovers a GRN for each cell by harboring information from transcriptionally similar cells using a random forest model (Figure 2).

Although ScReNI has proved to be a strong algorithm for single-cell GRN inference, the performance of the underlying model has not been extensively explored. To be more precise, it is not investigated whether predictive performance varies across genes and cells and how such behavior could propagate and influence the inferred GRNs.

This work investigates the relation between model random forest performance and the reliability of the resulting ScReNi-inferred GRNs. Since the edges are derived directly from the weights predicted by the random forest importance scores, model performance could be a factor that directly influences the inferred networks, therefore, the initial direction to explore is whether some genes have different predictive accuracy than others, and how this impacts their role within the GRN. Subsequently, the work also investigates defining a metric that quantifies the reliability of a produced GRN and relating it back to the cell metadata followed by the comparison of GRNs from cells with high and low respective metric scores. By answering these questions, the robustness of ScReNi-inferred cell-specific GRNs will be better understood, while also providing guidance in interpreting the networks.

2 Methodology

This work investigates the stability of the GRNs inferred by ScReNI based on Alzheimer's Disease data using a series of experiments (Figure 1). The analysis is structured around three experiments, each targeting a specific aspect of GRN stability: per-gene random forest accuracy, cell-level network reliability and differences in GRNs of low and high accuracy genes. All experiments, as well as the ScReNI algorithm itself are implemented in Python.

2.1 Datasets

SEA-AD

This work uses the Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD) [1], a large-scale single-cell multiomics dataset of human brain tissue provided by the Allen Institute for Brain Science. The dataset contains paired scRNA-seq and scATAC-seq measurements from postmortem brain tissue across 28 donors, comprising approximately 138,000 cells spanning 24 distinct cell subclasses. The paired nature of the data, where RNA and ATAC measurements originate from the same cells, makes it particularly well suited for ScReNI, which benefits from paired modalities for neighborhood construction. For computational feasibility, experiments are conducted on a default subsample of 1,200 cells, consisting of 50 cells per cell subclass across all 24 subclasses. Gene expression data is represented by 500 highly variable genes (HVGs)

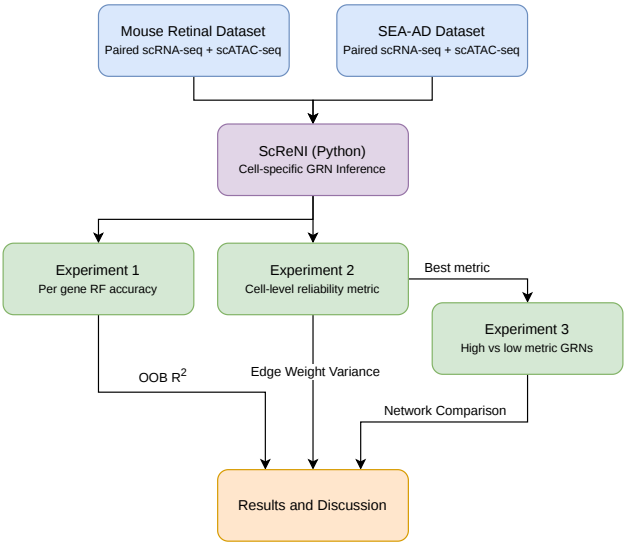


Figure 1: General approach of the work conducted in this paper.

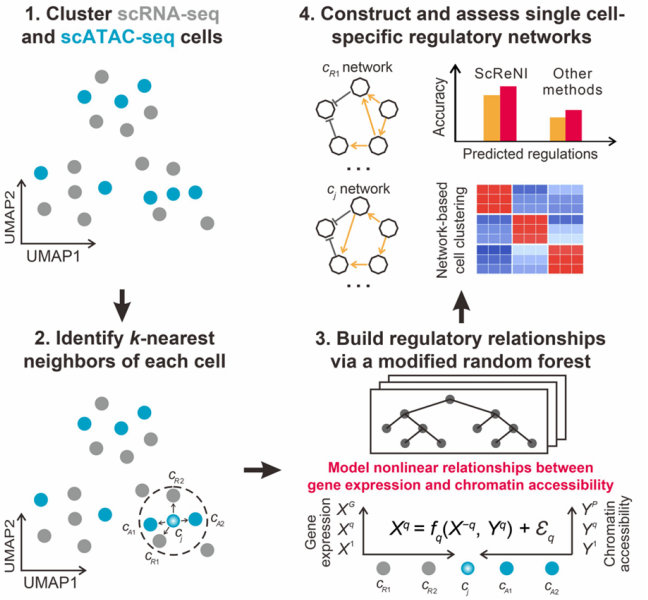


Figure 2: Overview of ScReNI from the original work [10].

and chromatin accessibility by 10,000 highly variable peaks (HVPs).

Mouse Retinal

The mouse retinal dataset is a paired single-cell dataset that contains both scRNA-seq and scATAC-seq measurements from mouse retinal cells and was obtained from the original paper's Seurat exports. After preprocessing, the data used contains 400 cells, consisting of 100 cells per cell subclass, with 500 highly variable genes and 217 chromatin accessibility peaks. Due to its smaller size, simpler biological context and availability of ground truth, it is a perfect tool to validate both the reimplementation of ScReNI (subsection 2.3) and some of the experiments performed in this paper.

2.2 ScReNI

In order to infer cell-specific GRNs, ScReNI first integrates scRNA-seq and scATAC-seq data in a shared space and then, for each cell, it identifies the k neighbors that will be used as a training set. A separate random forest is trained for each gene of a cell, using the expressions of all other genes and the accessibility of associated chromatin peaks as input features. The importance scores that result from the random forest are used as weights between genes to infer the cell-specific GRN.

2.3 Reimplementation of ScReNI

The algorithm used in this paper is a Python version that was rewritten based on the original R code, specifically the network inference methods (wScReNI and kScReNI) and the precision and recall evaluator. The methods have been reproduced as faithfully as possible to the source material, with some unavoidable underlying differences in internal computation ways of certain methods like the Random Forest, which works slightly different in the *scikit-learn* Python package compared to R, in particular, the random forest implementation in *scikit-learn* differs slightly from the *randomForest* R package in its tree-building procedure regarding feature selection at each split. The fidelity of the reimplementation was validated by reproducing Figure 2C from the original paper [10], namely the precision and recall when used on the original dataset of mouse retinal data. The Python version produced precision and recall values within one percent of the reported values, confirming sufficient fidelity for use in subsequent experiments.

2.4 Experimental Setup

Gene Accuracy

A key property of the random forest relevant to this work is the out-of-bag (OOB) score. This is an internal estimate of model accuracy computed on samples that are not used in training each tree. It provides an accuracy estimate without the need of separating a test set, making OOB R^2 a suitable metric for evaluating how well the model captures different gene relationships. This property is fundamental to the stability analysis conducted in this work. For the inference of networks from both datasets, the wScReNi inference method had seed = 42, nthread = 8, max_cell_per_batch = 10.

To analyze how random forest accuracy relates to a gene's properties within the network, the per-gene OOB R^2 was correlated against a set of properties. These were computed on the top 500 edges with highest weights of each network, and then averaged across cells resulting in a single value per gene. This edge restriction was introduced to reduce noise from less relevant near-zero weights, with 500 also being a threshold used in the original paper [10]. Among the used properties, two that are worth mentioning in more detail are PageRank and betweenness centrality. PageRank [3] is an algorithm originally designed by Google that applied to this work measures a gene's overall influence within the network, accounting for the influence of its regulators and betweenness centrality [2] measures the frequency with which a gene lies on the

shortest regulatory path between other gene pairs, reflecting its role as a network intermediary.

Stability Metric

The first proposed metric is the mean OOB R^2 of each cell across all genes. The second is the edge weight variance across multiple runs of the network inference step. This has been done by running the inference of the networks five times (seeds: 892, 7739, 6545, 4388, 4330), computing the variance of each gene X gene pair and averaging this per cell across genes. The seeds were generated using a fixed random number generator (NumPy *default_rng*) initialized with seed 42. Five integer seeds were then sampled uniformly from the range [0, 10000). To determine whether either candidate metric reflects genuine network reliability rather than simply network density or cell expression, partial Spearman correlations were additionally computed between each metric and precision/recall, controlling separately for network density (number of non-zero edges per cell) and mean expression level (mean raw expression across a cell's own genes, prior to KNN pooling). To control for density, each metric was first linearly regressed on density and expression, respectively and partial correlations were obtained by correlating the residuals of these regressions.

GRN changes with high vs low metric

To investigate how the GRNs of more and less reliable cells differ structurally, cells within each cell type were ranked by their mean edge weight variance and split into a high and a low reliability group, defined as the top and bottom 10% of cells within that type. Three comparisons were performed between these groups.

Edge consistency was assessed by computing the Pearson correlation between every pair of individual cell networks within a group (prior to any averaging), yielding a distribution of pairwise correlations for the high group and for the low group, which were compared using a one-sided Mann-Whitney U test.

Transcription factor activity was assessed by first computing the element-wise mean network across all cells in each group, then counting, for each TF, the number of non-zero outgoing edges (out-degree) in this group-level average network. The difference in out-degree between the high and low groups was computed for each TF.

Network overlap between the two group-level average networks was assessed using both binary and weighted Jaccard similarity. Binary Jaccard was computed by selecting the top K edges by weight from each network and measuring the proportion of shared edges relative to their union, for a range of K values up to the number of non-zero edges in the sparser of the two networks. Weighted Jaccard extends this to account for edge magnitude, computed as $\sum \min(w_A, w_B) / \sum \max(w_A, w_B)$ over the union of the top-K edges from each network, allowing assessment of agreement in edge strength in addition to edge presence.

2.5 Evaluation

The second experiment requires the precision and recall as baselines to ensure the proposed metric is correct in indicating a more trustworthy GRN. In order to get this baseline of biological plausibility in an inferred GRN, its edges are compared against a ground truth obtained from ChIP-Atlas [7]. This ground truth information is only available for the mouse retinal data, as processing the human data would have exceeded the deadlines of this work.

2.6 Code Availability

The Python implementation of ScReNI and the data preprocessing code is publicly accessible on GitHub (https://github.com/DelftBioinformaticsLab/bsc-screni). In this same repository you will find all of the experiments' code on the *accuracy* branch of the repository (https://github.com/DelftBioinformaticsLab/bsc-screni/tree/accuracy). All dependencies are located in the *pixi.toml* file. The original R implementation can also be accessed publicly on GitHub (https://github.com/Xuxl2020/ScReNI)

3 Results

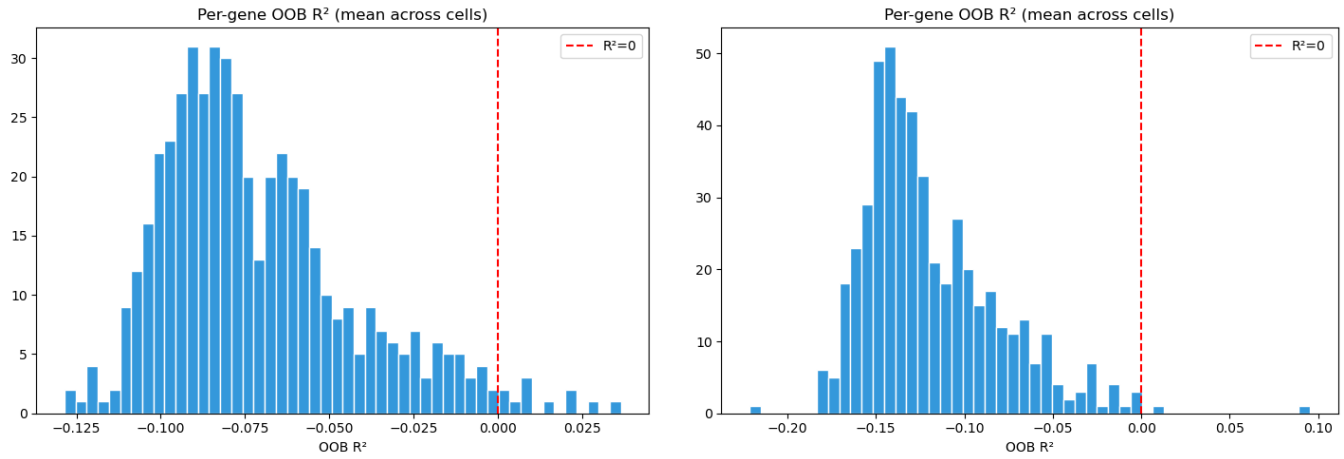
3.1 Gene Accuracy

The first experiment analyzes whether random forest accuracy varies across genes and how this variation is tied to the inferred GRN. For each gene, the random forest that was fitted during the network inference is evaluated using the out-of-bag R^2 (OOB R^2) score, which provides an accuracy estimate without the need of defining a separate test set, as described in section 2.4. The accuracy score of a gene is defined as the mean OOB R^2 across cells in which the gene was expressed. Genes with zero expression across all cells in a given KNN window were skipped and assigned a NaN score. Low and high accuracy genes are also mapped onto the network to determine if they occupy certain positions like hubs or peripheral nodes.

The distribution of mean OOB R^2 score displayed in Figure 3 shows a mostly negative trend. For the SEA-AD subset, out of the 500 genes in the cells, only 11 have achieved a positive OOB R^2 score, while 487 have a negative OOB R^2 score and 2 are entirely NaN (Figure 3a). The mean OOB R^2 across all genes is -0.071 (std=0.028). For the mouse retinal subset, out of the 500 genes, only two have a positive OOB R^2, with the other 498 having a negative score (Figure 3b). The mean OOB R^2 across all genes in this case is -0.119 (std=0.038). This shows that for the majority of genes, for both subsets, the random forest model performed worse than the trivial mean predictor.

In order to understand how the different accuracy scores of genes impact their properties within the network, the OOB R^2 was correlated against a set of network and expression properties, computed on the top 500 edges of each cell-specific network and then averaged across cells, with the results summarized in Table 1.

The gene predictability is most tightly correlated to mean expression and expression variance for both datasets (SEA-AD: r = 0.673, r = 0.592 and mouse: r = 0.725, r = 0.720), surpassing any network topology measure. The mean outgoing and incoming edge weight follow as the next strongest correlations (r = 0.44 - 0.52 for both datasets). The in-degree, showing how regulated a gene is, and the PageRank, describing how influential that gene is respective to the entire net-



(a) (b)

Figure 3: Histograms showing the distribution of OOB R^2 for each gene for the (a) SEA-AD and (b) mouse retinal datasets.

work, are the last moderately correlated properties (r $\approx$ 0.4 for both datasets). Betweenness centrality, measuring how often a gene lies on the path between different genes' regulatory routes, has a weak but consistent positive association in both datasets. Out-degree is the only tested property that shows a discrepancy between datasets, a significant negative association in mouse retinal data, but no significant association in SEA-AD.

Table 1: Spearman correlation between gene-level OOB R^2 and network/expression properties, computed within the top-500 edge GRN, for both datasets. All reported correlations are significant at $p < 0.001$ unless otherwise noted in the text.

Property	**SEA-AD**	**Mouse retinal**
Mean expression	0.673	0.725
Expression variance	0.592	0.720
Mean outgoing edge weight	0.520	0.437
Mean incoming edge weight	0.442	0.518
In-degree	0.412	0.428
PageRank	0.409	0.393
Betweenness centrality	0.287	0.234
Out-degree	0.032†	-0.234

† Not significant ($p = 0.475$).

3.2 Stability Metric

The second experiment aims to define a stability metric that can be provided along with the produced GRN. This experiment computes and compares the proposed metric to the baseline precision and recall that are computed against a ground truth only available to the mouse retinal dataset. The metric will be considered valid as long as it aligns with the readings of the baseline that cannot be trivially explained by other correlations.

After having investigated the per-gene accuracy in the first experiment, it is first hypothesized that a per-cell accuracy score could be a potential candidate for our metric as a well equipped random forest is expected to produce more accurate results. As such, the first proposed metric was the cell-specific mean OOB R^2 score across all genes. The score mean was -0.112 (std=0.023), with all cells having a negative score. In order to evaluate whether this metric suggestion correctly reflects network quality, it was correlated with the precision (Figure 4a) and recall (Figure 4b). The OOB R^2 shows no relevant correlation with precision (Spearman r = 0.034, p = 0.5), but a strong association with recall (Spearman r = 0.399, p <0.001). The recall correlation is consistent across all four cell types (MG, RPC1, RPC2, RPC3, r = 0.47-0.61), while the precision association is only present in the MG type (r = 0.249).

For an alternative candidate metric, the network inference step was repeated five times using different fixed random seeds in order to get the weight variance of each edge, averaging them into a single score for each cell. The hypothesis was that a cell with low edge weight variance would produce more accurate results because the random forest is more convinced the predicted weights are correct by giving similar importance scores across all runs. This metric shows substantially stronger correlation to both precision (Figure 4c, r = 0.516, p <0.001) and recall (Figure 4d, r = 0.663, p <0.001) compared to the previous metric, but is as well inconsistent across different cell types. It has strong association to precision and recall in MG (r = 0.615, r = 0.693) and RPC3 (r = 0.390, r = 0.304), but non significant to precision for RPC1 and RPC2. The initial results are going against the hypothesis, where we were expecting a negative correlation (when variance increases, precision and recall were expected to decrease). At this point a theory is that this is caused by the expression level of each cell, where highly expressed cells get more variance in edge weights, compared to unexpressed cells, where most values are close to zero.

With earlier findings of the per-gene analysis pointing to-

wards a pretty strong connection between OOB R^2 and expression and the unexpected results of edge weight variance, expression was also investigated at the cell-level metrics proposed. Both OOB R^2 and edge weight variance were correlated with the mean expression level of the cell and the network density, defined as the number of non-zero edges predicted by the method, and both metrics showed strong associations (OOB R^2: r=0.406 , r=0.549 and edge variance: r = 0.234, r = 0.598, respectively). With both metrics having such high correlation with network density, it was investigated on its own and showed that it almost completely explained recall by itself (r = 0.859). This raised the possibility of the metrics to just act as a proxy for network density rather than actual reliability signal. To properly investigate whether the metrics held actual informative signal, partial Spearman correlation were computed, controlling separately for density and mean expression. All correlations investigated are summarized in Table 2, where OOB R^2 gets really weak associations to precision and recall after controlling for density and expresion, meaning that this metric was mostly driven by these factors and did not carry extra stability signal. In contrast, edge weight variance managed to keep a strong association to both precision (r = 0.470) and recall (r = 0.364), and showed almost no change at all when controlling for expression (r = 0.543, r = 0.657, respectively).

3.3 GRN Comparison

With a potentially meaningful metric in place, the GRNs of more reliable cells will be compared to those of less reliable cells in an attempt to see how these may differ structurally. Since the metric was only able to be validated against the mouse dataset, that will be used for this experiment as well. The experiment is conducted on each of the four cell types in order to separate biological differences from computational ones. Cells within each of the four cell types (MG, RPC1, RPC2, RPC3) were split into a high and low edge weight variance group, taking the top and bottom 10% (10 cells per group). The average network was constructed for each group by taking the element-wise mean of the networks and three comparisons were performed. Because of the highest correlation with precision and recall from subsection 3.2, only the MG type results will be highlighted.

First comparison is edge consistency, measured as the distribution of pairwise correlations between networks within a group. The hypothesis was that cells with a better metric score would "agree" more on the same important edges. In practice, edge consistency did not differ significantly between high and low variance groups (p = 0.29). The high edge variance groups consistently had more non-zero edges than low edge variance groups across all four cell types (for MG: 89740 versus 50069 non-zero edges). This reflects the high density correlation of the metric reported in subsection 3.2.

After seeing how much cells from within the same group agree with each other, the next direction to investigate is the overlap between the two groups. The network overlap between high and low variance group-level networks was assessed using both binary and weighted Jaccard similarity, computed over multiple top ranked edges K. While binary Jaccard increases with K (Figure 5a), indicating the 2 groups start sharing more and more edges as the network expands, weighted Jaccard decreases with K (Figure 5b), showing the two groups disagree more as the network is expanded. The trend is consistent across al four types of cells. This means that while largely agreeing on what edges are present, the two groups disagree more and more on the strength of these connections as weaker edges are added.

Lastly, the transcription factor out-degree activity, measuring how many other genes it regulates, was consistently higher in the high edge weight variance group across all four cell types. The pattern is not present across all TFs, with some having the same number of outgoing edges and in some cases, the low variance group had some TFs with more activity Figure 6. This effect is likely explained in part by the high correlation to density found in subsection 3.2, rather than reflecting a TF-specific regulatory effect.

4 Responsible Research

4.1 Dataset description and personally identifiable information

The SEA-AD dataset is freely available to be used in research under the data use agreement set by the Allen Institute for Brain Science. While the dataset contain donor data such as age, sex and disease progression, no analysis in this paper aims to identify individual donors and all experiments happen at the cell level, where gene expressions cannot be traced back to the original donor. The mouse retinal dataset involves no human subjects and therefore has no personally identifiable information.

4.2 Ethics

One potential ethical concern is that resulted inferred GRNs could be misinterpreted as having direct clinical implications. In reality, the regulatory edges inferred by ScReNI are statistical associations derived from a subsample of cells and require heavy biological validation before influencing a diagnosis or treatment.

4.3 Reproducibility

To support reproducibility, the full implementation of ScReNI and all experiment code is publicly available (subsection 2.6), including the data preprocessing pipeline and dependency specification. All dataset-specific parameters required to reproduce the experiments, including subsample sizes, the number of highly variable genes and peaks, and the k-nearest-neighbor graph configuration, are reported in subsection 2.1. The random seeds used for the repeated network inference runs in the stability metric experiment are listed in subsection 2.4. Software dependencies and their versions are pinned in the repository's pixi.toml file, ensuring that the computational environment can be reconstructed exactly.

4.4 Generative AI

Generative AI tools were used as supporting instruments during the process of this work. Their use was limited to improving the overall quality and presentation of the work. This includes brainstorming ideas, generating example code for detailed figures and plots, generating and formatting tables for

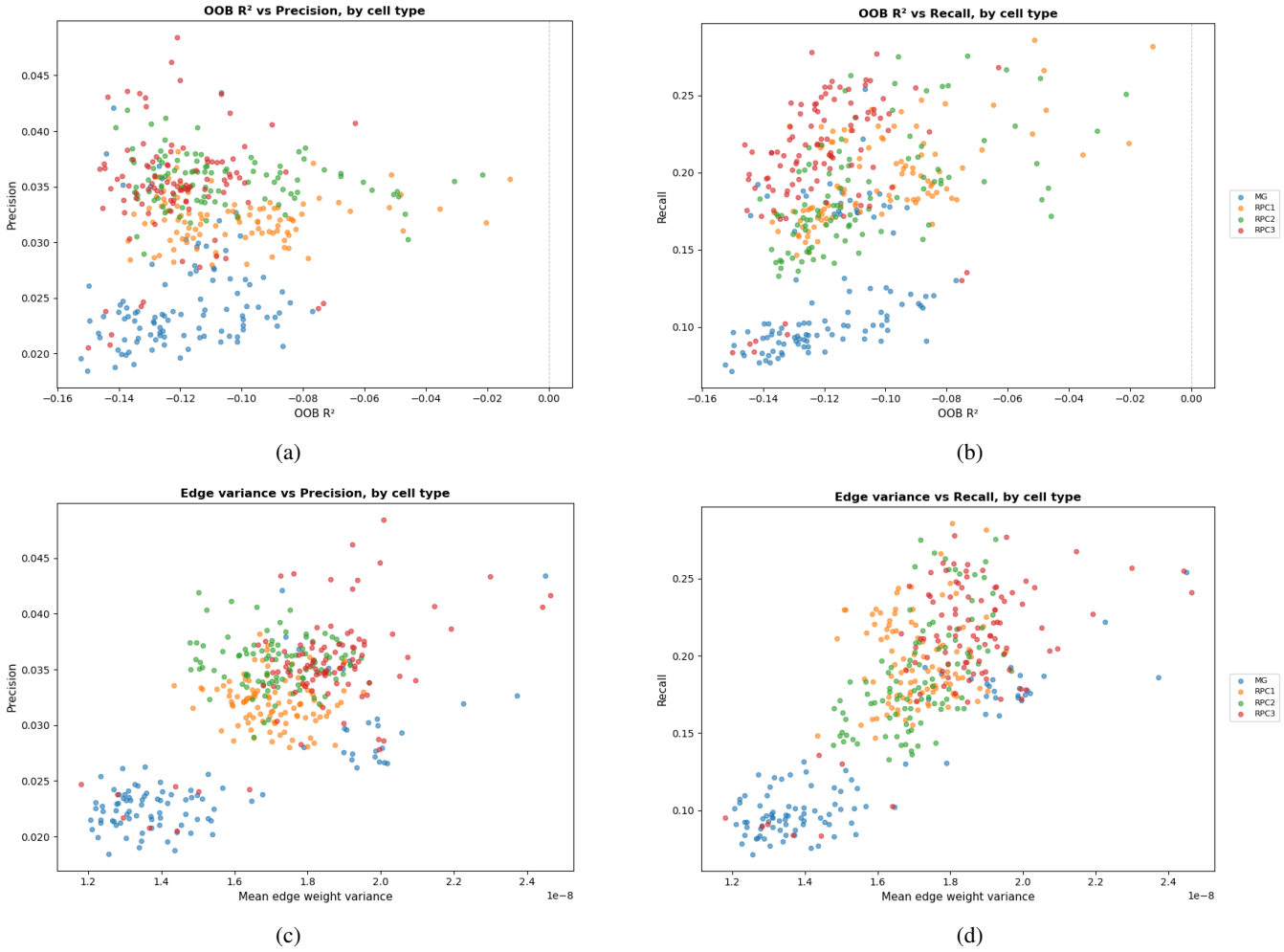


Figure 4: Correlation with precision and recall of OOB R^2 (a), (b) and edge weight variance (c), (d), respectively.

LaTeX integration, assisting with grammar correction, language refinement, and rephrasing author-written text to offer better readability and clarity. All research, analyses, interpretation and results were developed by the author, who still has full responsibility for the accuracy, originality and integrity of the final work. No sensitive information was provided to AI systems.

5 Discussion

The mostly negative distribution of random forest accuracy showed not only that genes have different predictability, but also that most of the time, the random forest struggled to explain the features (the genes). The dominance of expression level and variance when correlated to the random forest accuracy, visible at similar levels in both datasets, indicates that this accuracy at the gene resolution is more a statistical artifact of detectability rather than a signal of biological importance or regulatory complexity. A highly expressed gene provides the random forest with more signal to work with, independent of their specific role within the regulatory network.

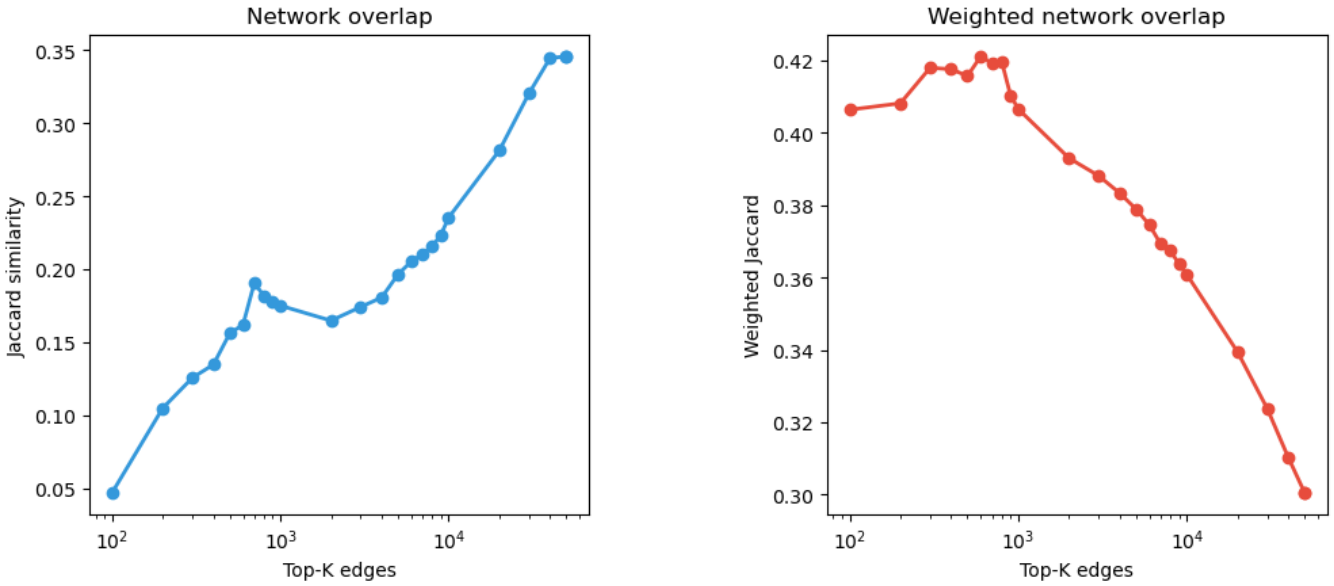
The similar correlation of in-degree and PageRank across both datasets is a more meaningful finding: genes that are heavily regulated by other genes and genes with a high overall network influence tend to be better predicted by the model. This could indicate that such genes are part of denser local regulatory neighborhoods, giving the random forest more correlated predictive features.

The discrepancy in the out-degree association is a limitation to acknowledge. In the mouse retinal data, genes that were well predicted regulated fewer genes, suggesting they place downstream in the regulatory hierarchy. This pattern did not replicate in SEA-AD, where the association was not significant. Given this inconsistency, claim about out-degree being connected to gene predictability should be treated as dataset specific, rather than generalizable with this finding being reported as a limitation rather than a concrete conclusion.

Random forest accuracy did not perform better at the cell-level, despite being directly tied to the accuracy of the model that predicts the GRNs, it is not suited to be a reliability metric for said GRNs. Its correlation to recall, while strong at first sight, proved to be mostly driven by its association to network density, showing that OOB R^2 functions primarily as a proxy

Table 2: Comparison of metrics' correlation to Precision, Recall Density and Expression under different circumstances.

Metric	Precision	Recall	Density	Expression	Precision (ctrl density)	Recall (ctrl density)	Precision (ctrl expression)	Recall (ctrl expression)
OOB R^2	0.034	0.399	0.549	0.406	-0.130	-0.170	0.059	0.230
Edge Variance	0.516	0.663	0.598	0.234	0.470	0.364	0.543	0.657



(a) Binary Jaccard

(b) Weighted Jaccard

Figure 5: Network overlap between top and bottom 10 % edge weight variance cells of subtype MG.

for how dense and well-expressed a cell is, rather than capturing whether the inferred edges are correct. The complete lack of association to precision in three of the four cell types further shows that high random forest accuracy does not imply a more biologically faithful network, contrary to initial expectation.

Edge weight variance shows more promise as a metric. Unlike OOB R^2, it keeps its correlations high after controlling for density and expression, indicating that it captures information about network reliability that is not simply inherited by the cell's expression levels or dense connections. This is the strongest positive evidence found for a viable cell-specific reliability metric. That being said, its inconsistency across some cell types means it should not be treated as a generally validated metric, but rather as the most promising direction identified, warranting further research across a larger range of cell types and datasets.

A limitation of this experiment is that the edge weight variance analysis was only conducted on the mouse retinal dataset, due to the lack of satisfactory ground truth for the SEA-AD dataset. Moreover, this metric requires five times the computation time of OOB R^2, making it far more expensive, depending on the dataset size. A practical tradeoff that should be taken into account when selecting a metric for large-scale analysis.

Now that a promising candidate metric was found, its impact on the networks was investigated next. The absence of significant difference in within-group ege consistency between the high and low edge variance groups shows that despite its association with precision and recall established in the second experiment, the edge weight variance does not correspond to a discrete "stable" and "unstable" distinction of groups. Cells grouped by this metric have the same coherence with one another as a result of the grouping. This shows the edge weight variance should be treated as a continuous signal rather than the basis of a binary classifier, at least for the group sizes used in this experiment. The large and consistent difference in network density between high and low edge variance cells makes any differences observed in TF-level activity not enough proof on their own to attribute genuine differences in regulatory behavior.

The opposite trends of the binary and weighted Jaccard similarity offers the most information out of this experiment. Binary Jaccard shows that high and low edge variance cells agree on which gene pairs are connected as we include more edges, their networks overlapping substantially when including all edges. Weighted Jaccard tells a different story: by taking edge strength into consideration, the agreement fails as weaker edges get introduced. In other words, the two groups built similar network frames, but assign very different confidence to the connections.

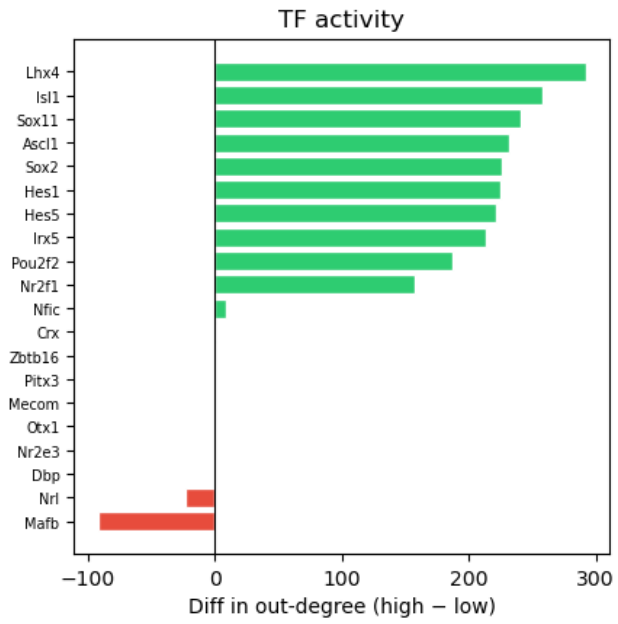


Figure 6: Difference in transcription factor activity between top and bottom 10 % edge weight variance cells of subtype MG.

References

[1] Allen Institute for Brain Science. About sea-ad. URL: https://brain-map.org/consortia/sea-ad/about-sea-ad.

[2] Linton C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, 1977. doi:10.2307/3033543.

[3] David F. Gleich. Pagerank beyond the web. *SIAM Review*, 57(3):321–363, 2015.

[4] Kongwei Huang, Yixue Xu, Tong Feng, Hong Lan, Fei Ling, Hai Xiang, and Qingyou Liu. The advancement and application of the single-cell transcriptome in biological and medical research. *Biology*, 2024. doi:10.3390/biology13060451.

[5] Byungjin Hwang, Ji Hyun Lee, and Duhee Bang. Single-cell rna sequencing technologies and bioinformatics pipelines. *Experimental Molecular Medicine*, 50, 2018. doi:10.1038/s12276-018-0071-8.

[6] Sai Ma, Bing Zhang, Lindsay M LaFave, Andrew S Earl, Zachary Chiang, Yan Hu, Jiarui Ding, Alison Brack, Vinay K Kartha, Tristan Tay, Travis Law, Caleb Lareau, Ya-Chieh Hsu, Aviv Regev, and Jason D Buenrostro. Chromatin potential identified by shared single-cell profiling of rna and chromatin. *Cell*, 183(4):1103–1116.e20, 2020. doi:10.1016/j.cell.2020.09.056.

[7] Shinya Oki, Tazro Ohta, Go Shioi, Hideki Hatanaka, Osamu Ogasawara, Yoshihiro Okuda, Hideya Kawaji, Ryo Nakaki, and Jun Sese Chikara Meno. Chip-atlas: a data-mining suite powered by full integration of public chip-seq data. *The EMBO Reports*, 19(12), 2018.

[8] World Health Organization. Dementia, 2025. URL: https://www.who.int/news-room/fact-sheets/detail/dementia.

[9] Cole Trapnell, Davide Cacchiarelli, Jonna Grimsby, Prapti Pokharel, Shuqiang Li, Michael Morse, Niall J Lennon, Kenneth J Livak, and Tarjei S Mikkelsen John L Rinn. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*, 32, 2014. doi:10.1038/nbt.2859.

[10] Xueli Xu, Yanran Liang, and Miaoxiu Tang. Screni single-cell regulatory network inference through integrating scrna-seq and scatac-seq data. *Genomics, Proteomics Bioinformatics*, 2025. doi:10.1093/gpbjnl/qzaf060.