

Article

Not peer-reviewed version

AENetMoX: Fast and Precise scRNA-seq Gene Regulatory Network Inference Through Multi-Modal Feature Integration in Brain Organoids

[Shirshak Aryal](#) *

Posted Date: 16 March 2026

doi: 10.20944/preprints202603.1191.v1

Keywords: Gene Regulatory Network (GRN) inference; single-cell RNA-sequencing (scRNA-seq); machine learning; transcription factors (tf); brain organoids



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

AENetMoX: Fast and Precise scRNA-seq Gene Regulatory Network Inference Through Multi-Modal Feature Integration in Brain Organoids

Shirshak Aryal

Independent Researcher, Kathmandu 44600, Nepal; shirshak.acad@gmail.com

Abstract

Motivation: Gene regulatory network (GRN) inference from single-cell RNA-seq (scRNA-seq) data remains hampered by technical noise, high false-positive rates, and extreme computational costs. Existing methods often require hours or days to process developmental datasets yet fail to capture the physical and topological constraints of regulatory interactions, essential for accurate regulatory mapping. **Results:** We developed AENetMoX, a multimodal autoencoder that integrates transcriptomic correlations with transcription factor (TF) binding motifs and protein-protein interaction (PPI) networks. We evaluated AENetMoX against SCENIC, GRNBoost2, and CLR across 48 independent configurations using three human brain organoid lineages. At $K = 100$, AENetMoX achieved $7.7 \pm 5.1\%$ precision (95.7% relative improvement over SCENIC; $p = 1.039 \times 10^{-3}$, one-sided Wilcoxon test; Vargha-Delaney $\hat{A} = 0.635$). ChIP-seq validation showed 51.6% precision (+9.5% over SCENIC, $p = 0.141$, $\hat{A} = 0.500$) and 168.4% improvement in F1 over SCENIC ($p = 2.883 \times 10^{-9}$, $\hat{A} = 0.854$). Ablation studies revealed PPI integration as the primary driver of performance, increasing precision by 492% ($\sim 5.9 \times$) and ChIP-seq recovery precision by 198% ($\sim 3 \times$) over its expression-only variant. Crucially, AENetMoX completes inference in under 5 minutes, a 24-fold speedup over SCENIC (~ 2 hours) and significantly outperforming CLR (> 12 hours). Analysis of novel predictions identified 334 unique regulatory edges, including temporally persistent *SMAD3* and *SOX2* hubs that remain stable across multiple developmental stages. **Availability:** Source code is available at <https://github.com/Shirshak52/AENetMoX>. All datasets and databases used are available at OSF (Project DOI: <https://doi.org/10.17605/OSF.IO/K6EHW>).

Keywords: Gene Regulatory Network (GRN) inference; single-cell RNA-sequencing (scRNA-seq); machine learning; transcription factors (tf); brain organoids

1. INTRODUCTION

Gene regulatory networks (GRNs) define transcriptional programs controlling cellular differentiation. Inferring these networks from single-cell RNA-sequencing (scRNA-seq) data remains challenging due to technical noise, gene dropout, and limited ground-truth validation (Pratapa *et al.*, 2020). These issues are particularly acute in developmental contexts like human brain organoids, where regulatory dynamics are complex and cell-type-specific interactions are poorly characterized (Quadrato *et al.*, 2017; Birtele *et al.*, 2025).

Current GRN inference methods generally fall into three categories. Statistical approaches like Context Likelihood of Relatedness (CLR) (Faith *et al.*, 2007) use mutual information but require extensive computation (often > 12 hours per dataset). Ensemble methods such as GRNBoost2 (Moerman *et al.*, 2019) and SCENIC (Aibar *et al.*, 2017) improve scalability by applying gradient boosting with optional motif-based filtering yet frequently achieve low precision at high-confidence thresholds. Deep learning approaches have shown promise, but most existing architectures for GRN inference either require paired chromatin accessibility data or struggle with the extreme sparsity of

developmental scRNA-seq datasets (Yuan and Bar-Joseph, 2019; Kartha *et al.*, 2022; Chen and Liu, 2022).

A key limitation of existing methods is their treatment of transcription factors (TFs) as independent regulators. TFs frequently function as multi-protein complexes. For example, neuronal differentiation involves heterodimerization of bHLH factors (Dennis *et al.*, 2019) and cooperative binding by SWI/SNF chromatin remodelers (Sokpor *et al.*, 2017). Protein-protein interaction (PPI) networks encode this cooperativity, yet few GRN inference methods incorporate PPI topology beyond post-hoc filtering.

We developed AENetMoX (AutoEncoder for Network, Motif, and eXpression features), a multimodal approach that integrates three data types: (i) gene expression correlations, (ii) TF binding motif strengths from position weight matrices derived from JASPAR2024 non-redundant position frequency matrices (Rauluseviciute *et al.*, 2024), and (iii) TF-TF PPI network features from STRING (Szkarczyk *et al.*, 2025a). By encoding PPI topology as features within an autoencoder, AENetMoX captures cooperative regulatory mechanisms that expression-only methods fail to resolve.

We evaluated AENetMoX on three independent human brain organoid datasets (cortical, ventral forebrain, ventral midbrain) spanning 12 developmental timepoints. Compared to SCENIC (the highest-performing baseline) at $K = 100$, AENetMoX achieved a 95.7% relative improvement in precision ($7.7 \pm 5.1\%$ vs. $3.9 \pm 3.0\%$) and superior recovery of ChIP-seq validated TF-DNA interactions (F1 score improvement of 168.4%; $p = 2.883 \times 10^{-9}$). Ablation studies comparing the full model against its expression-only (AEX) and expression-plus-motif (AEMoX) variants demonstrated that PPI integration is the critical driver of accuracy. AENetMoX yielded a 5.9-fold increase in precision and a 3.0-fold increase in ChIP-seq recovery precision compared to the AEX ablation variant. All inferences were completed in under 5 minutes per dataset on consumer-grade hardware, enabling high-throughput GRN inference.

2. METHODS

2.1. Datasets and Preprocessing

We analyzed three publicly available scRNA-seq datasets representing distinct human brain organoid lineages. Cortical Organoids: Timepoints at weeks 3, 5, 8, and 10 were retained. Weeks 15 and 24 were excluded due to low cell counts ($< 5,000$ cells). The final dataset comprised approximately 87,000 cells following quality control (QC) and subsampling. Ventral Forebrain Organoids: Data included timepoints at months 1, 2, and 4, totaling approximately 64,000 cells post-QC. Ventral Midbrain Organoids: Timepoints at days 15, 30, 60, 90, and 120 were included, comprising approximately 117,000 cells. No subsampling was required for this lineage.

Quality control: Genes detected in fewer than 10 cells were removed. We excluded mitochondrial and ribosomal genes, as well as cells with < 200 or $> 5,000$ detected genes or $> 15\%$ mitochondrial expression.

Subsampling: To ensure computational efficiency while preserving cellular heterogeneity, datasets exceeding 150,000 cells post-QC were subsampled to 30% of the total cells per timepoint.

Normalization: Raw UMI counts were normalized to 10,000 total counts per cell and log1p transformed using Scanpy (v1.11.5; Wolf *et al.*, 2018). Raw counts were preserved in the `AnnData.layers['raw_counts']` for downstream processing.

Feature selection: For each dataset, we identified the top 3,000 highly variable genes (HVGs) using the Seurat v3 variance-stabilizing transformation (Stuart *et al.*, 2019). The final working gene set for each model comprised the union of all transcription factors (TFs) listed in AnimalTFDB 3.0 (Hu *et al.*, 2019) and the identified 3,000 HVGs.

All preprocessing steps were applied uniformly to ensure fair comparison across methods.

Prior knowledge integration: Transcriptional regulation was guided by motifs from JASPAR2024 (Rauluseviciute *et al.*, 2024). Promoter sequences were retrieved from EPDnew (Meylan *et al.*, 2019) using a window of -499 to $+100$ bp relative to the transcription start site (TSS) in the

hg38 reference genome. Protein interaction topology was derived from STRING v12.0 (Szkarczyk *et al.*, 2025b).

Performance benchmarks: Model performance was evaluated against a combined set of TRRUST v2 (Han *et al.*, 2018), KnockTF 2.0 (Feng *et al.*, 2024), and DoRothEA (Garcia-Alonso *et al.*, 2019), with DoRothEA IDs mapped via UniProt (Bateman *et al.*, 2025). ChIP-seq evidence was sourced from ReMap2022 (Hammal *et al.*, 2022) and mapped to gene names using GENCODE release 49 (Mudge *et al.*, 2025).

Biological validation: Via Enrichr (Kuleshov *et al.*, 2016) libraries, functional coherence was assessed using GO Biological Process 2025, and pathway enrichment was conducted using GO Biological Process 2025, GO Molecular Function 2025, KEGG 2021 Human, Reactome Pathways 2024, and WikiPathways 2024 Human (Aleksander *et al.*, 2025; Kanehisa *et al.*, 2021; Milacic *et al.*, 2024; Agrawal *et al.*, 2024).

2.2. AENetMoX Architecture

AENetMoX integrates expression, motif, and PPI features through three sequential modules:

Module I: Candidate Edge Selection

For each timepoint, we computed Pearson correlations (ρ) between all gene pairs in the working set. Candidate edges were retained if at least one gene was identified as a transcription factor (TF) and $|\rho| \geq 0.2$. The correlation threshold is a lenient prefilter to reduce the search space, while consistent with established GRN preprocessing workflows such as SCENIC (Aibar *et al.*, 2017). For each TF, we then extracted six expression statistics: mean, standard deviation, coefficient of variation (CV), skewness, percentage of expressing cells, and maximum expression.

Module II: Motif Scoring and Edge Prioritization

We constructed log-odds position weight matrices (PWMs) from JASPAR 2024 (Rauluseviciute *et al.*, 2024) position frequency matrices (PFMs) using a uniform background probability ($b = 0.25$) and pseudocount ($\epsilon = 10^{-6}$):

$$PWM_{i,j} = \log_2 \left(\frac{PPM_{i,j} + \epsilon}{0.25} \right)$$

For each candidate TF-target pair, we scanned the target gene's proximal promoter (-499 to $+100$ bp relative to TSS; Meylan *et al.*, 2019) with the TF's PWM, using a sliding window. The raw motif score (S_{motif_i}) at position i was:

$$S_{motif} = \sum_{j=1}^L PWM[base_{i+j}, j]$$

where L is motif length. The motif strength (S_{motif}) was the maximum score across all windows, min-max normalized per TF ($S_{motif} \in [0,1]$).

We computed a plausibility score (S_{plaus}) and a final dynamic score (S_{dyn}) to prioritize edges:

$$\begin{aligned} S_{plaus} &= 0.6 \times |\rho| + 0.4 \times S_{motif} \\ S_{dyn} &= 0.5 \times S_{plaus} + 0.5 \times CV_{norm} \end{aligned}$$

where CV_{norm} was the min-max normalized CV of the TF. To balance network coverage with computational efficiency, we applied adaptive filtering. If candidate edges exceeded 5,000 after correlation filtering ($|\rho| \geq 0.2$), we ranked then by S_{dyn} and applied a TF diversity constraint, limiting each TF to $\max(5, \frac{N}{100})$ target edges (where $N = 5,000$). This prevents individual "superhub" TFs from dominating the network while ensuring representation for all active regulators. If filtering reduced edges below 1,000, we ranked them by S_{dyn} and retained all of them to maintain sufficient signal for autoencoder training.

Module III: PPI Feature Integration

TFs were mapped to the STRING PPI network (Szklarczyk et al., 2025) using a high-confidence threshold (*combined_score* ≥ 700). For each TF, we extracted 17 topological and neighborhood features:

- **Centrality (8):** Degree, weighted degree, betweenness, closeness, eigenvector, PageRank, clustering coefficient, and average neighbor degree.
- **Neighborhood aggregates (6):** The six expression statistics of neighbor TFs.
- **Contrasts (3):** TF variance vs. neighborhood variance (activity weighted degree, expression contrast, and stability contrast)

Of the 1,665 TFs in AnimalTFDB 3.0, 900 (54.1%) had high-confidence interactions in STRING. JASPAR provided motif matrices for 615 TFs (74.3% overlap with AnimalTFDB). The final integrated feature set (expression + motif + PPI) was available for 461 TFs (27.7%). The final feature vector (26 dimensions) comprised 8 basic features ($|\rho|$, S_{motif} , and 6 TF statistics), 17 PPI features, and 1 binary indicator marking network presence. For TFs lacking PPI data, graph features were set to 0, enabling graceful degradation of the model.

Autoencoder Training

The model uses a symmetric bottleneck architecture: 26 → 64 → 32 → 16 → 8 → 16 → 32 → 64 → 26.

It includes batch normalization and dropout (0.2) to prevent overfitting. We trained for 100 epochs using Adam optimizer (learning rate = 0.001) with a Mean Squared Error (MSE) loss function. The final edge importance score (I) was calculated as the L_2 norm of the 8-dimensional bottleneck vector (Z):

$$I = \sqrt{\sum_{i=1}^8 z_i^2}$$

2.3. Experimental Setup

2.3.1. Timepoint Partitioning

To preserve temporal specificity of regulatory states, scRNA-seq expression matrices were partitioned by timepoint. Each resulting subset was processed independently across all evaluated models, including AENetMoX, baseline methods, and ablation variants. This approach avoids the "averaging effect" common in aggregate analysis and allows for the detection of transient regulatory events.

2.3.2. Baseline Methods

We compared AENetMoX against three established GRN inference frameworks:

- **CLR** (Faith et al., 2007): A statistical method based on Mutual Information with context likelihood normalization.
- **GRNBoost2** (Moerman et al., 2019): A regression-based method using gradient boosting.
- **SCENIC** (Aibar et al., 2017): An ensemble framework that combines GRNBoost2 with motif-based pruning,
-motifs-v10nr_clust-nr.hgnc-m0.001-o0.0.tbl

SCENIC (pySCENIC, commit 06bafba, retrieved from the Aertslab GitHub repository) was executed following the standard workflow. GRNBoost2 was run with a fixed seed and default Arboreto settings. Co-expression modules were derived using modules_from_adjacencies with thresholds of 0.75 and 0.90, top_n_targets=50, and min_genes=20.

Motif pruning (cisTarget) utilized the following specific databases:

1. hg38_refseq-r80_10kb_up_and_down_tss.mc9nr.genes_vs_motifs.rankings.feather
2. hg38_refseq-r80_500bp_up_and_100bp_down_tss.mc9nr.genes_vs_motifs.rankings.feather
3. motifs-v10nr_clust-nr.hgnc-m0.001-o0.0.tbl

Regulons were generated with df2regulons using a Rank threshold of 1500 and a Normalized Enrichment Score (NES) threshold of 3.0. Finally, AUCCell scores were computed with auc_threshold=0.05 to quantify regulon activity across cells.

2.3.3. Ablation Studies

To isolate the contributions of different data modalities, we trained two internal variants using the same symmetric autoencoder architecture:

AEX (Expression-only): Utilizes 7 input dimensions ($|\rho|$ and 6 TF expression statistics).

AEMoX (Expression + Motif): Utilizes 8 input dimensions (AEX features plus the S_{motif} score).

Both variants used training procedures and hyperparameters identical to the full AENetMoX model to ensure that performance gains were attributable strictly to feature integration.

Hub TF Identification

We implemented an adaptive hub-detection algorithm based on out-degree elbow detection to identify key regulators. For each inferred network, TFs were classified as hubs if their out-degree (D_{out}) was ≥ 3 and exceeded 15% of the maximum out-degree in the network ($D_{out} \geq \max(3, \max(D_{out}) \cdot 0.15)$). To ensure mathematical significance, we applied an elbow-detection step that prioritized the top 5 – 12 regulators based on the greatest consecutive degree differences. This procedure was used exclusively for network topology analysis and the generation of GRN diagrams.

Algorithm

The AENetMoX inference pipeline is summarized in the procedure below.

Algorithm 1: AENetMoX Regulatory Inference

Input: CP10K normalized expression matrix X , TF annotations (AnimalTFDB 3.0), log-odds PWMs (JASPAR2024), promoter sequences (EPDnew), PPI network (STRING v12.0).

Output: Ranked Regulatory Network $G = \{(TF, Target, I)\}$.

Feature Extraction

Identify TFs in X via AnimalTFDB

Define working gene set: TFs $\cup$ top 3,000 HVGs

Candidate Edge Selection (Module I)

Compute Pearson correlations ρ for all gene pairs.

Retain edges where: (i) ≥ 1 gene is a TF, (ii) $|\rho| \geq 0.2$.

For each TF, extract 6 expression statistics (mean, SD, CV, skew, % expressing, max).

Motif Scoring and Edge Filtering (Module II)

For each candidate edge (TF, target):

Scan target promoter for S_{motif} (max score, min-max normalized).

Compute plausibility: $S_{plaus} = 0.6 \cdot |\rho| + 0.4 \cdot S_{motif}$

Compute dynamic score: $S_{dyn} = 0.5 \cdot S_{plaus} + 0.5 \cdot CV_{norm}$

Adaptive filtering:

If ($(|candidate\ edges\ with\ |\rho| \geq 0.2|) > 5000$):

Sort by S_{dyn} descending

Iterate through sorted list

Retain edges only if: (a) Total selected edges < 5000 . (b) Edges for the specific TF $< \max(5, 50)$

If ($|remaining\ candidate\ edges| == 0$):

Retain all edges from global pool

Else: Retain all *candidate edges* with $|\rho| \geq 0.2$

PPI Feature Integration (Module III)

Map TFs to STRING (*combined_score* > 700)

For each TF, extract 17 graph features

Set features to 0 if TF not in network
Add binary presence indicator
Feature assembly
Concatenate features per edge into a 26-D vector: [$|\rho|$, S_{motif} , 6 TF stats, 17 PPI features, 1 binary]
Autoencoder Training
Train symmetric MLP: $26 \rightarrow 64 \rightarrow 32 \rightarrow 16 \rightarrow 8 \rightarrow 16 \rightarrow 32 \rightarrow 64 \rightarrow 26$
Optimizer: Adam ($lr = 0.001$), Loss: MSE, Epochs: 100
Edge Ranking
Compute importance: $I = \|z\|_2$ (L_2 norm of the 8D bottleneck)
Rank final edge list by $|I|$.
Implementation
AENetMoX is implemented in Python using the PyTorch framework.
Regularization: The architecture utilizes Batch Normalization and a Dropout rate of 0.2 to maintain robustness against the inherent noise and sparsity of single-cell transcripts.
Optimization: Efficiency is achieved via pre-processed JASPAR and STRING assets and the adaptive scaffold filter. By capping TF targets and retaining the top 5,000 edges by S_{dyn} , the model reduces dimensionality while preserving regulatory diversity. This architecture enables a full inference cycle for an entire dataset in < 5 minutes on consumer-grade hardware (NVIDIA RTX 4050, 6GB VRAM).

4. RESULTS AND DISCUSSION

We developed AENetMoX, a multimodal autoencoder that integrates transcription factor protein-protein interaction (PPI) network features with expression correlations and DNA-binding motif strengths for gene regulatory network (GRN) inference from scRNA-seq data
To ensure a robust assessment, we constructed multi-tiered validation sets. The unified validation set integrated 121,319 unique regulatory edges from TRRUST v2 (Han et al., 2018) (8,403 edges), KnockTF 2.0 (Feng et al., 2024) (101,887 edges), and DoRothEA (Garcia-Alonso et al., 2019) (15,117 edges). Physical binding evidence was derived from ReMap2022 (Hammal et al., 2022), encompassing 7,807,522 edges across 1,209 unique TFs.
Precision and AUPRC were calculated on the unified validation set comprised of TRRUST v2, KnockTF 2.0. and DoRothEA (mapped to gene names using UniProt database, Bateman et al., 2025). ChIP-seq metrics were calculated using the ReMap2022 mapped to gene names with GENCODE (release 49) annotations (Mudge et al., 2025).
Evaluated across 12 developmental timepoints across four different network densities ($K = 10, 50, 100, 500$) from three independent human brain organoid datasets (cortical, ventral forebrain, ventral midbrain), AENetMoX achieved superior precision compared to established methods while maintaining computational efficiency, supported by one-sided paired Wilcoxon signed rank tests, $n = 48$ (12 timepoints $\times$ 4 K values) – See Supplementary Table S1 for the full Wilcoxon test metrics.

4.1. Performance Summary

Table 1 shows performances at $K = 100$ (mean $\pm$ standard deviation) and the relative change vs the best-performing baseline, with p -values and Vargha-Delaney Effect Sizes ($\hat{A}$) from the Wilcoxon tests.

Table 1. Performance at $K = 100$ (12 timepoints).

Metric	AENetMoX	Best Baseline	Δ vs Best Baseline	p-value	$\hat{A}$
Precision	$7.7 \pm 5.1\%$	SCENIC:	+95.7%	1.039 $\times 10^{-3}$	0.635
		$3.9 \pm 3.0\%$			
AUPRC	0.046 ± 0.024	SCENIC:	+21.8%	0.013	0.583
		0.038 ± 0.017			
ChIP-seq	$51.6 \pm 39.8\%$	SCENIC:	+9.5%	0.141	0.500
Precision		$47.1 \pm 43.7\%$			
ChIP-seq F1	0.00102 ± 0.00177	CLR:	−20.0%	0.845	0.542
		$0.00128 \pm$			
		0.00292			
ChIP-seq F1	0.00102	0.00038	+168.4%	2.883 $\times 10^{-9}$	0.854
vs SCENIC	± 0.00177	± 0.00073			

AENetMoX achieved $7.7 \pm 5.1\%$ precision compared to SCENIC's $3.9 \pm 3.0\%$ (Figure 1), representing a 95.7% relative improvement ($p = 1.039 \times 10^{-3}$, $\hat{A} = 0.635$). AUPRC gains remained superior: 0.046 ± 0.024 vs. SCENIC's 0.038 ± 0.017 (21.8% improvement, $p = 0.013$, $\hat{A} = 0.583$).

On ChIP-seq validation (where both TF and target had ReMap2022 binding evidence), AENetMoX achieved $51.6 \pm 39.8\%$ precision versus SCENIC's $47.1 \pm 43.7\%$ (+9.5%, $p = 0.141$, $\hat{A} = 0.500$, *ns*). While F1 scores showed higher variance, AENetMoX achieved a 168.4% improvement over SCENIC (0.00102 ± 0.00177 vs 0.00038 ± 0.00073 , $p = 2.883 \times 10^{-9}$, $\hat{A} = 0.854$). Although CLR achieved a higher mean F1 score (0.00128 ± 0.003), its performance was characterized by extreme instability (*Coeff. of Variation* = 228%, $\hat{A} = 0.542$ vs. AENetMoX, $p = 0.845$, *ns*). These results indicate that AENetMoX provides a superior and more stable balance between precision and recall when recovering experimentally validated physical interactions compared to traditional ensemble or statistical methods.

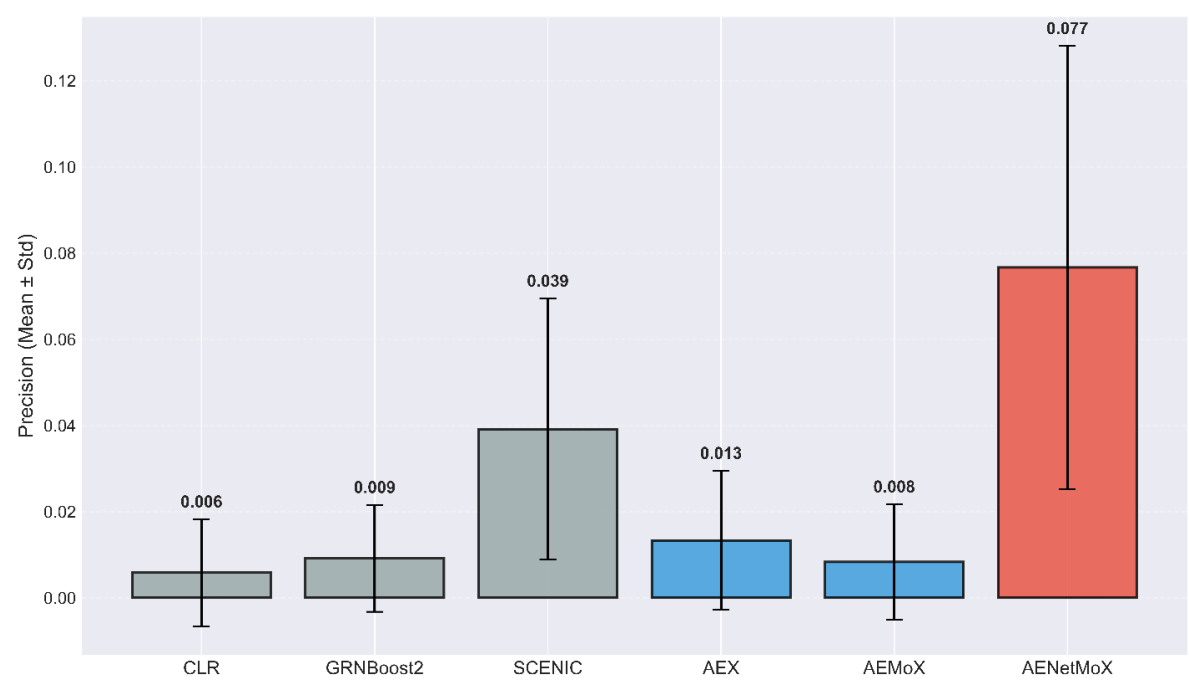


Figure 1: Precision Comparison at K=100

Additional performance metric visualizations are provided in Supplementary Figures 1.

4.2.. Ablation Study Reveals PPI Contribution

To isolate feature contributions, we compared AENetMoX against two ablated variants: AEX (expression-only) and AEMoX (expression + motif). Performance results at $K = 100$ (mean $\pm$ standard deviation) are summarized in Table 2.

Table 2. Ablation Study ($K = 100$, 12 timepoints).

Method	Features	Precision (Δ vs AEX)	AUPRC	ChIP-seq	ChIP-seq F1 (Δ vs AEX)
			(Δ vs AEX)	Precision (Δ vs AEX)	
AENetMoX	Expression +	$7.7 \pm 5.1\%$	0.046 ± 0.024	$51.6 \pm 39.8\%$	0.00102 ± 0.002
	Motif + PPI	(+492%)	(+318.2%)	(+198.27%)	(+121.74%)

AEMoX			0.012		
	Expression +	0.8 ± 1.3%		17.7 ± 32.2%	0.00049 ± 0.002
			± 0.008		
	Motif	(−38.5%)		(+2.31%)	(+6.52%)
			(+9.1%)		
AEX			0.011		
	Expression only	1.3 ± 1.6%		17.3 ± 30.3%	0.00046 ± 0.002
			± 0.007		

Motif features provided negligible benefit in isolation, decreasing precision by 38.5% (AEMoX vs. AEX: 0.8% vs. 1.3%) while slightly improving AUPRC (+9.1%). In contrast, integrating PPI features (AENetMoX vs. AEX) resulted in substantial gains across all metrics:

- Precision: +492% (1.3% → 7.7%, $p = 2.594 \times 10^{-8}$)
- AUPRC: +318.2% (0.011 → 0.046, $p = 7.918 \times 10^{-10}$)
- ChIP-seq Precision: +198.27% (17.3% → 51.6%, $p = 8.358 \times 10^{-9}$)
- ChIP-seq F1: +121.74% (0.00046 → 0.00102, $p = 4.706 \times 10^{-6}$)

Relative to motifs only (AEMoX), AENetMoX improves precision by 862%, with significant gains in AUPRC (+283%), ChIP-seq precision (+191%), and ChIP-seq F1 (+108%). Compared to expression only (AEX), the gains represent increases of 5.9 ×, 4.2 ×, 3.0 ×, and 2.2 ×, respectively. These results confirm that the PPI module is not merely additive but is the primary driver of high-confidence regulatory inference.

The Precision-Coherence Trade-Off

Our results reveal a precision-coherence trade-off in GRN inference (Table 3; complete data in Supplementary Table S2). Motif-based methods such as SCENIC and AEMoX achieve high functional coherence (91.9% and 92.2% of TFs with GO-enriched targets, respectively) and pathway enrichment (65.5% and 51.5%) but produce fragmented networks with low connectivity (largest components of 30.9% and 40.9%). In contrast, AENetMoX prioritizes experimentally validated interactions, achieving 51.6% ChIP-seq precision and 68.3% network connectivity.

Table 3. Biological Validation Metrics ($K = 100$).

Method	Coherence (%)	Enrichment	ChIP-seq	Hub	Largest
	TFs)	(%)	Precision (%)	TFs (%)	Component (%)
CLR	83.6 ± 23.5	65.4 ± 29.3	23.3 ± 30.2	4.7 ± 1.4	52.0 ± 18.3
GRNBoost2	79.9 ± 24.6	50.7 ± 33.6	27.3 ± 38.3	7.5 ± 3.4	66.4 ± 19.4
SCENIC	91.9 ± 10.7	65.5 ± 10.9	47.1 ± 43.7	5.2 ± 1.7	30.9 ± 16.3
AEX	82.7 ± 25.6	48.8 ± 26.9	17.3 ± 30.3	3.0 ± 2.0	33.7 ± 22.1

AEMoX	92.2 ± 19.0	51.5 ± 25.1	17.7 ± 32.2	2.6 ± 0.8	40.9 ± 25.9
AENetMoX	88.6 ± 18.0	46.2 ± 19.1	51.6 ± 39.8	3.1 ± 1.0	68.3 ± 21.6

This trade-off reflects biological reality: transcription factors regulate targets across multiple pathways to coordinate complex cellular programs (Lambert et al., 2018; Sokpor et al., 2017). By incorporating PPI topology, AENetMoX captures these cross-pathway regulatory relationships that GO enrichment metrics may penalize as "incoherent." Despite a slight reduction in enrichment scores (46.2% vs. SCENIC's 65.5%), AENetMoX maintains coherence levels (88.6%) higher than expression-only baselines like CLR (83.6%) and GRNBoost2 (79.9%).

Crucially, the 68.3% network connectivity indicates that AENetMoX recovers the interconnected structure characteristic of real biological networks, whereas motif-heavy methods (SCENIC: 30.9%; AEMoX: 40.9% in largest component) fragment the GRN into isolated functional modules. Furthermore, AENetMoX achieves nearly 3-fold higher ChIP-seq precision than AEMoX (51.6% vs. 17.7%), demonstrating that PPI integration is essential for recovering physically validated interactions, even when they span multiple biological processes.

Biological validation visualizations are provided in Supplementary Figures 2.

Dataset-Specific Performance and Discovery of Novel Regulatory Edges

Precision at $K = 100$ varied across neural contexts, peaking in ventral forebrain (14.3%), followed by cortical organoids (9.1%) and ventral midbrain (2.6%). This gradient likely reflects disparities in ChIP-seq database coverage: ventral forebrain GABAergic interneuron development involves well-characterized regulators (e.g., DLX1/2, NKX2-1), whereas midbrain dopaminergic differentiation remains underrepresented in curated validation databases.

Beyond recovering known biology, AENetMoX identified a substantial repertoire of high-confidence novel interactions. We analyzed the top 100 predictions ($K = 100$) across 12 timepoints (1,200 total predictions) across all datasets. Of these, 507 edges (42.2%) were identified as "Novel" (ChIP-seq validated in ReMap2022 but absent from TRRUST v2, KnockTF 2.0, or DoRothEA). After deduplication by removing timepoint context, this yielded 334 unique novel TF-target interactions (40.3% recovery rate). At these precision levels, approximately 1 in 7 ventral forebrain predictions is expected to validate experimentally, compared to 1 in 11 for cortical and 1 in 38 for midbrain models. Visualizations are included in Supplementary Figures 2.

Analysis of the temporal distribution of these predictions reveals regulatory persistence that offsets lower dataset precision. In the ventral midbrain, despite a 2.6% precision rate, AENetMoX consistently identified the SMAD3 → TRIM9 and SMAD3 → CIB3 interactions across four independent timepoints (Day 30, 60, 90, and 120). The high frequency of these "novel" hits across independent data subsets suggests a stable regulatory signal for midbrain axonal guidance and maturation that is systematically omitted from curated databases.

In the cortical and forebrain lineages, AENetMoX identified a dominant regulatory hub centered on JUN and SOX2. The top-ranked novel edges demonstrate high biological plausibility (Supplementary Tables S3b and S4b for top 10 per dataset; S3a and S4a for full lists), including AP-1 complex signatures (JUN → JUNB, JUND) and critical neurodevelopmental links (JUN → SOX2, SOX2 → POU3F2). Notably, the model captures a maturation-dependent shift in SOX2 targets: early-stage regulation of progenitor factors (HES5, NKX2-1) transitions to glial and extracellular matrix regulators (BCAN, METRN) in later stages. By providing these ranked prediction lists, AENetMoX narrows the experimental search space from thousands of potential interactions to a discrete set of high-confidence candidates for perturbational validation.

Computational Efficiency

AENetMoX completes the full inference cycle for an entire developmental dataset in < 5 minutes, representing a significant improvement over SCENIC's ~2 hour execution time (a ~24 × speedup). This efficiency is primarily driven by the dual-filtering strategy described in the methods: retaining the top 5,000 edges by dynamic score (S_{dyn}) to reduce the autoencoder's input dimensionality, while simultaneously enforcing an adaptive TF-diversity cap.

By limiting per-TF targets to 1% of the total network (50 targets per 5,000 edges), AENetMoX prevents "superhub" TFs from monopolizing the feature space. This ensures a broad representation of the regulatory landscape without the computational cost associated with dense, all-to-all gene correlations. Consequently, this optimized architecture enables high-fidelity GRN construction on consumer-grade GPUs, such as the NVIDIA RTX 4050 (6 GB VRAM), removing the necessity for specialized high-performance computing (HPC) clusters or extensive cloud resources.

Limitations and Future Directions

AENetMoX has four primary limitations. First, requiring motif matches during candidate edge selection may exclude non-canonical binding events, such as tethered recruitment where TFs bind DNA indirectly via protein-protein interactions rather than direct sequence recognition. Second, STRING PPI coverage remains incomplete: only 54.1% of human TFs (900 in STRING v12.0 out of 1665 in AnimalTFDB 3.0) have high-confidence interactions (*combined_score* ≥ 700), limiting PPI feature utility for poorly characterized regulators. TFs absent from STRING receive zero-valued PPI features, potentially underestimating their regulatory importance. While the ablation study demonstrates that partial PPI coverage still yields substantial gains (198% vs AEX, 191% vs AEMoX on ChIP-seq precision), full integration of expression, motif, and PPI features is available for only 27.7% of TFs (461/1,665), with the remaining relying on subset feature combinations.

Third, JASPAR motif coverage (74.3% of AnimalTFDB TFs) means ~25% of TFs lack sequence-based priors, forcing the model to rely solely on expression correlations for those regulators. Fourth, ChIP-seq validation from ReMap2022 uses bulk or mixed cell-type datasets, not single-cell resolution, introducing false negatives when cell-type-specific regulatory interactions are unmarked in reference databases.

Future work should integrate chromatin accessibility (scATAC-seq) to directly measure regulatory element activity, reducing dependency on JASPAR coverage by observing actual open chromatin states. Incorporating 3D genome organization data (Hi-C or promoter-capture Hi-C) could improve distal enhancer-promoter linkage predictions, capturing long-range regulatory interactions. Furthermore, extending PPI databases with protein structure predictions (AlphaFold-Multimer, (Omid et al., 2024)) may increase TF-TF interaction coverage beyond STRING's experimental evidence. Adapting AENetMoX to time-series scRNA-seq data (e.g. RNA velocity or metabolic labeling) may enable inference of causal regulatory dynamics. Finally, experimental validation of the 334 novel predictions, such as via pooled CRISPR interference (CRISPRi) screens (Meng et al., 2025), would provide orthogonal confirmation of the model's predictive accuracy across different organoid contexts.

Generalizability Beyond Brain Organoids

While we evaluated AENetMoX exclusively on brain organoid datasets, the method contains no brain-specific components. The three data types integrated, i.e. gene expression correlations, TF binding motifs (JASPAR2024), and PPI networks (STRING v12.0), are universal resources applicable to any tissue or organism with scRNA-seq data. We restricted evaluation to neural contexts to enable rigorous cross-dataset validation within a coherent biological domain, but the underlying architecture imposes no tissue-specific constraints.

Future work should assess AENetMoX performance on non-neural developmental systems (cardiac organoids, gastrulation, hematopoiesis) and across species (mouse, zebrafish). The method's reliance on general biological priors rather than tissue-specific training suggests it may function as a universal GRN inference framework. Initial tests on publicly available non-brain datasets would

clarify whether the precision gains observed here (+95.7% over SCENIC) extend beyond neurodevelopment or reflect specific properties of neurodevelopmental datasets.

CONCLUSION

By integrating protein-protein interaction networks with sequence motifs and expression data, AENetMoX achieves superior precision for high-confidence GRN predictions from scRNA-seq data while maintaining computational efficiency (~24 × faster than SCENIC). Our ablation study reveals that PPI features increase ChIP-seq validation precision by +191% (~3 ×) relative to motif-based approaches, albeit at the cost of functional coherence. This trade-off challenges the field's reliance on GO enrichment as a primary validation metric, demonstrating that ChIP-seq recovery, not functional coherence, should define ground truth for regulatory network accuracy.

We provide ranked prediction lists for three human brain organoid types with dataset-specific expected validation success rates (ventral forebrain: 1 in 7, cortical: 1 in 11, ventral midbrain: 1 in 38), enabling targeted experimental prioritization. The 334 novel TF-target interactions identified by AENetMoX represent experimentally validated edges supported by ReMap2022 ChIP-seq data but absent from curated databases, offering immediate candidates for perturbational studies in neural development. Source code is publicly available at <https://github.com/Shirshak52/AENetMoX> under MIT license.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/doi/s1>.

Funding: This work received no external funding.

Acknowledgments: We thank the Aerts lab for developing and openly sharing SCENIC and pySCENIC, the developers of GRNBoost2 and Arboreto, and the authors of CLR for making their code publicly available. We also acknowledge the ReMap consortium for curating and providing the comprehensive ChIP-seq binding data used for validation.

Conflicts of Interest: None declared.

Availability and Implementation: The AENetMoX source code is freely available at <https://github.com/Shirshak52/AENetMoX> under the MIT license.

All datasets, prior knowledge databases, validation databases, and SCENIC resources (raw and processed) are publicly available at OSF (DOI: <https://doi.org/10.17605/OSF.IO/K6EHW>, licensed under CC BY 4.0).

Original raw scRNA-seq datasets:

Cortical organoids: UCSC Cell Browser (<https://cells.ucsc.edu/organoidreportcard/organoids10X>)

Ventral forebrain organoids: Broad Institute Single Cell Portal (https://singlecell.broadinstitute.org/single_cell/study/SCP2892/scrna-seq-of-ventral-organoids?cluster=clustering.txt&spatialGroups=--&annotation=nFeature_RNA--numeric--study&subsample=all#study-download)

Ventral midbrain organoids: GEO Accession GSE168323 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE168323>)

Original database sources (detailed explanations and instructions in repository README):

AnimalTFDB 3.0: https://github.com/GuoBioinfoLab/AnimalTFDB3/blob/master/AnimalTFDB3/static/AnimalTFDB3/download/Homo_sapiens_TF

JASPAR2024: <https://jaspar.elixir.no/download/data/2024/CORE/>

EPDnew: <https://genome.ucsc.edu/cgi-bin/hgTables> (retrieval instructions in repository README)

STRING v12.0: <https://stringdb-downloads.org/download/protein.links.full.v12.0/>

SCENIC cisTarget databases: <https://resources.aertslab.org/cistarget/>

TRRUST v2: <https://www.grnpedia.org/trust/downloadnetwork.php>

KnockTF2.0:
<https://maayanlab.cloud/Harmonizome/dataset/KnockTF+Gene+Expression+Profiles+with+Transcription+Factor+Perturbations>
DoRothEA: Provided in repository (originally from OmniPath)
UniProt: https://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/idmapping/by_organism/
ReMap2022: https://remap.univ-amu.fr/download_page
GENCODE release 49: <https://www.genecodegenes.org/human/>

References

1. Agrawal,A. *et al.* (2024) WikiPathways 2024: next generation pathway database. *Nucleic Acids Res*, 52, D679–D689.
2. Aibar,S. *et al.* (2017) SCENIC: single-cell regulatory network inference and clustering. *Nat Methods*, 14, 1083–1086.
3. Aleksander,S.A. *et al.* (2025) The Gene Ontology knowledgebase in 2026. *Nucleic Acids Res*.
4. Bateman,A. *et al.* (2025) UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res*, 53, D609–D617.
5. Birtele,M. *et al.* (2025) Modelling human brain development and disease with organoids. *Nat Rev Mol Cell Biol*, 26, 389–412.
6. Chen,G. and Liu,Z.-P. (2022) Graph attention network for link prediction of gene regulations from single-cell RNA-sequencing data. *Bioinformatics*, 38, 4522–4529.
7. Dennis,D.J. *et al.* (2019) bHLH transcription factors in neural development, disease, and reprogramming. *Brain Res*, 1705, 48–65.
8. Faith,J.J. *et al.* (2007) Large-Scale Mapping and Validation of Escherichia coli Transcriptional Regulation from a Compendium of Expression Profiles. *PLoS Biol*, 5, e8–e8.
9. Feng,C. *et al.* (2024) KnockTF 2.0: a comprehensive gene expression profile database with knockdown/knockout of transcription (co-)factors in multiple species. *Nucleic Acids Res*, 52, D183–D193.
10. Garcia-Alonso,L. *et al.* (2019) Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res*, 29, 1363–1375.
11. Hammal,F. *et al.* (2022) ReMap 2022: a database of Human, Mouse, Drosophila and Arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments. *Nucleic Acids Res*, 50, D316–D325.
12. Han,H. *et al.* (2018) TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res*, 46, D380–D386.
13. Hu,H. *et al.* (2019) AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res*, 47, D33–D38.
14. Kanehisa,M. *et al.* (2021) KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res*, 49, D545–D551.
15. Kartha,V.K. *et al.* (2022) Functional inference of gene regulation using single-cell multi-omics. *Cell Genomics*, 2, 100166.
16. Kuleshov,M. V. *et al.* (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*, 44, W90–W97.
17. Lambert,S.A. *et al.* (2018) The Human Transcription Factors. *Cell*, 172, 650–665.
18. Meng,X. *et al.* (2025) CRISPR screens in human neural organoids and assembloids. *Nat Protoc*.
19. Meylan,P. *et al.* (2019) EPD in 2020: enhanced data visualization and extension to ncRNA promoters. *Nucleic Acids Res*.
20. Milacic,M. *et al.* (2024) The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Res*, 52, D672–D678.
21. Moerman,T. *et al.* (2019) GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics*, 35, 2159–2161.
22. Mudge,J.M. *et al.* (2025) GENCODE 2025: reference gene annotation for human and mouse. *Nucleic Acids Res*, 53, D966–D975.

23. Omid, A. *et al.* (2024) AlphaFold-Multimer accurately captures interactions and dynamics of intrinsically disordered protein regions. *Proceedings of the National Academy of Sciences*, 121.
24. Pratapa, A. *et al.* (2020) Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat Methods*, 17, 147–154.
25. Quadrato, G. *et al.* (2017) Cell diversity and network dynamics in photosensitive human brain organoids. *Nature*, 545, 48–53.
26. Rauluseviciute, I. *et al.* (2024) JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*, 52, D174–D182.
27. Sokpor, G. *et al.* (2017) Chromatin Remodeling BAF (SWI/SNF) Complexes in Neural Development and Disorders. *Front Mol Neurosci*, 10.
28. Stuart, T. *et al.* (2019) Comprehensive Integration of Single-Cell Data. *Cell*, 177, 1888–1902.e21.
29. Szklarczyk, D. *et al.* (2025) The STRING database in 2025: protein networks with directionality of regulation. *Nucleic Acids Res*, 53, D730–D737.
30. Wolf, F.A. *et al.* (2018) SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol*, 19, 15.
31. Yuan, Y. and Bar-Joseph, Z. (2019) Deep learning for inferring gene relationships from single-cell expression data. *Proceedings of the National Academy of Sciences*, 116, 27151–27158.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.