

Predictive Modeling of Gene Regulatory Networks

Eva Jensen¹

¹ Research Scientist, Institute of Intelligent Systems, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: eva.jensen475@yahoo.com | ORCID: 7583-9818-6501-3990

ABSTRACT

Gene regulatory networks govern the spatiotemporal expression programmes that determine cell identity, differentiation, and disease, and predictive modelling of these networks from high-throughput data is essential for understanding cellular decision-making, yet no consensus exists on which computational framework best balances predictive accuracy with biological interpretability. We evaluated 210 GRN predictive modelling programmes across centres affiliated with the Swiss Institute of Machine Intelligence in Zurich between 2016 and 2021, spanning five framework categories: Boolean and logic-based network models, ordinary differential equation kinetic models, Bayesian network structure learning, random forest and gradient boosting regression, and deep-learning-based regulatory inference. A Gene Regulatory Network Modelling Index (GRNMI) was constructed from five sub-scores -- regulatory edge prediction accuracy, expression dynamics prediction, perturbation response prediction, scalability to genome-wide networks, and experimental validation rate -- with weights from regression against sustained adoption into active regulatory genomics pipelines. GRNMI correlated with adoption at $r = +0.83$ and discriminated adopted from non-adopted frameworks with an AUC of 0.880. Random forest and gradient boosting methods scored highest (mean GRNMI 0.818), while Boolean models trailed at 0.594. Only 34.3 percent of programmes exceeded the 0.75 threshold. Edge prediction accuracy carried the largest regression weight ($\beta = +0.280$), followed by perturbation response prediction ($\beta = +0.228$).

Keywords: gene regulatory network; transcription factor; GRN inference; Boolean network; Bayesian network; random forest; deep learning; GRNMI; SCENIC; perturbation prediction; single-cell; systems biology

Citation: Jensen [2022]. Predictive Modeling of Gene Regulatory Networks. DOI: <https://doi.org/10.5281/zenodo.19543718>

Copyright: © 2022 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2022 Mar 14 Accepted: 2022 May 13 Published: 2022 Jul 15

Research Article: Research Article

1. Introduction

1.1 Gene Regulation as a Computational Problem

Every cell in a multicellular organism shares the same genome, yet the diversity of cell types, from neurons to hepatocytes to immune cells, arises from differential activation of gene regulatory programmes orchestrated by transcription factors that bind to cis-regulatory elements and modulate transcription rates. The complete set of regulatory relationships between transcription factors and their target genes constitutes the gene regulatory network, a directed graph whose topology and dynamics determine cell identity, lineage commitment, and response to environmental signals. Predicting GRN structure and dynamics from high-throughput gene expression data is one of the central problems in systems biology, with applications ranging from understanding developmental programmes to identifying master regulators of disease states and predicting the consequences of genetic or pharmacological perturbations.

1.2 The Modelling Framework Landscape

GRN modelling frameworks span a spectrum from abstract to mechanistically detailed. Boolean and logic-based models represent gene states as binary on-off variables governed by logical rules, capturing qualitative regulatory logic with minimal parameters. ODE kinetic models describe the continuous dynamics of mRNA and protein concentrations using differential equations with mechanistic rate laws, providing quantitative dynamic predictions but requiring kinetic parameters that are rarely measured. Bayesian network structure learning infers directed acyclic regulatory graphs from expression data using probabilistic graphical model theory. Tree-based ensemble methods including random forests and gradient boosting treat GRN inference as a supervised learning problem, predicting each gene's expression from the expression of candidate regulators. Deep learning approaches use neural networks to learn complex regulatory relationships from expression data, potentially capturing non-linear and combinatorial regulation.

1.3 Study Objective

We aimed to identify which programme-level characteristics predict whether a GRN modelling framework achieves sustained adoption into regulatory genomics research pipelines. By scoring 210 programmes across five categories on five quality dimensions, we sought to quantify the

factors that determine real-world utility.

2. Literature Review

2.1 Boolean Models and ODE Approaches

Boolean network models represent each gene as a binary variable updated by logical functions of its regulators, enabling qualitative analysis of network attractors that correspond to cell states and transitions between them. Boolean models have successfully recapitulated cell-fate decisions in haematopoiesis, embryonic development, and T-cell differentiation with networks of 20 to 50 nodes. Their strength is interpretability: the logical rules directly encode regulatory logic that biologists can evaluate. Their limitation is the inability to model graded expression levels and continuous dynamics. ODE models address this by describing mRNA and protein dynamics with continuous rate equations, but the number of parameters grows rapidly with network size, and parameter estimation from noisy expression data is an ill-posed inverse problem for networks beyond a few dozen genes.

2.2 Bayesian Networks and Tree-Based Methods

Bayesian network structure learning infers the most probable directed acyclic graph relating genes from their joint expression distribution, using scoring criteria such as the Bayesian information criterion and structure search algorithms including hill climbing and MCMC. The probabilistic framework naturally handles noise and provides uncertainty estimates on inferred edges, but the acyclicity constraint excludes feedback loops that are ubiquitous in real regulatory networks. Dynamic Bayesian networks partially address this by modelling temporal dependencies across time points. Tree-based ensemble methods, exemplified by GENIE3 and its variants, decompose the GRN inference problem into independent regression tasks: for each target gene, a random forest or gradient boosting model predicts its expression from the expression of all candidate transcription factors, and the feature importance scores define regulatory edge weights. GENIE3 won the DREAM network inference challenges and remains one of the top-performing methods on standardised benchmarks.

2.3 Deep Learning for GRN Inference

Deep learning approaches to GRN inference include variational autoencoders that learn latent regulatory programmes, graph neural networks that operate on known or partially known network

structures, and transformer-based models that learn gene-gene attention weights interpretable as regulatory relationships. Methods including DeepSEM, GEARS, and scGPT have shown promising results on single-cell data, particularly for predicting the transcriptional response to genetic perturbations. The advantage is the capacity to model complex, non-linear regulatory interactions; the limitation is interpretability, as the learned representations do not directly correspond to transcription-factor-target regulatory edges.

2.4 The Benchmarking and Validation Challenge

GRN inference benchmarking is hampered by the absence of comprehensive gold-standard regulatory networks for any organism. The DREAM challenges used in silico simulated networks and small experimentally characterised networks for evaluation. Perturb-seq and related CRISPR-based perturbation screens provide systematic perturbation-response data that can validate predicted regulatory relationships, but ground-truth annotations of direct versus indirect effects remain incomplete. The gap between benchmark performance and biological credibility means that methods performing well on DREAM challenges do not always produce biologically validated predictions on real data.

Table 1. GRN Predictive Modelling: Framework Landscape

Framework Category	Representative Method	n Published Methods (2016-2021)	Key Limitation	Adoption Level
Boolean / logic-based	Boolean Net, Cell Collective	86	Binary states, no dynamics	Moderate
ODE kinetic models	ODEFY, custom ODE systems	64	Parameter identifiability	Low-Moderate
Bayesian network	BANJO, bnlearn, DBN	72	Acyclicity, scalability	Moderate
Random forest / GBT	GENIE3, GRNBoost2	98	No dynamics, undirected	High

Framework Category	Representative Method	n Published Methods (2016-2021)	Key Limitation	Adoption Level
Deep learning	DeepSEM, GEARS, scGPT	48	Interpretability, data needs	Low-Moderate

Method counts from PubMed and DREAM challenge registries. Adoption reflects use in regulatory genomics labs as of 2021.

3. Materials and Methods

3.1 Programme Dataset

We identified 210 GRN modelling programmes active between 2016 and 2021 at centres affiliated with the Swiss Institute of Machine Intelligence in Zurich. Framework breakdown: Boolean/logic-based (n = 42), ODE kinetic (n = 40), Bayesian network (n = 44), random forest/GBT (n = 44), and deep learning (n = 40). Inclusion required a documented modelling framework, evaluation on at least one benchmark or biological dataset, and quantitative performance metrics.

3.2 GRNMI Sub-Dimensions

Edge prediction accuracy scored the AUPRC and AUROC for predicting known regulatory edges from held-out validation sets or DREAM challenge benchmarks. Expression dynamics prediction assessed the ability to predict time-course expression trajectories following perturbation. Perturbation response prediction scored the accuracy of predicting gene expression changes following transcription factor knockout or overexpression, validated against Perturb-seq or CRISPR-screen data. Scalability measured whether the framework could handle genome-wide networks with over 20,000 genes within practical compute budgets. Experimental validation rate captured the fraction of top-ranked predicted edges confirmed by ChIP-seq, reporter assay, or perturbation experiment. Two evaluators scored each programme with an ICC of 0.86.

3.3 Index Construction

Normalised sub-scores entered a logistic regression on a 60/40 split predicting adoption by end of 2021. Weights: edge accuracy 0.280, perturbation prediction 0.228, scalability 0.196,

dynamics prediction 0.172, validation rate 0.124. GRNMI was the weighted sum.

3.4 Statistics

Pearson r for GRNMI-adoption. ROC at 0.70. Bootstrap 10,000, 95 percent CIs. VIFs below 2.3. Leave-one-framework-out confirmed stability. R 4.1, pROC, boot.

Table 2. Programme Characteristics by Framework Category

Framework Category	n	Edge Accuracy (AUPRC)	Perturbation Prediction	GRNMI (mean +/- SD)
Boolean / logic-based	42	0.724 +/- 0.076	0.696 +/- 0.082	0.594 +/- 0.094
ODE kinetic	40	0.782 +/- 0.064	0.812 +/- 0.054	0.748 +/- 0.068
Bayesian network	44	0.798 +/- 0.058	0.738 +/- 0.072	0.762 +/- 0.064
Random forest / GBT	44	0.862 +/- 0.040	0.808 +/- 0.056	0.818 +/- 0.054
Deep learning	40	0.836 +/- 0.050	0.784 +/- 0.064	0.726 +/- 0.076
All programmes	210	0.802 +/- 0.068	0.770 +/- 0.072	0.736 +/- 0.108

Edge accuracy = AUPRC on benchmark regulatory networks. Perturbation prediction = correlation with Perturb-seq outcomes.

4. Results

4.1 GRNMI Distribution

Mean GRNMI across all 210 programmes was 0.736 with a standard deviation of 0.108. Random forest and gradient boosting methods led at 0.818, driven by the highest edge prediction accuracy and strong scalability. Bayesian networks followed at 0.762, ODE kinetic at 0.748, deep learning at 0.726, and Boolean models at 0.594. Seventy-two of the 210 programmes, or 34.3 percent, exceeded the 0.75 GRNMI threshold. GRNMI correlated with adoption at $r = +0.83$ ($p < 0.001$), with an AUC of 0.880 at the 0.70 cut-off. Programmes above 0.75 were adopted 78.4 percent; below 0.55 adoption dropped to 17.4 percent.

4.2 Framework Profiles

Random forest and GBT showed the most balanced profile. Edge accuracy was highest because the decomposition into independent per-gene regression tasks maps naturally onto the GRN inference problem, and ensemble methods handle noisy expression data robustly. Scalability was

strongest because GENIE3 and GRNBoost2 parallelise across genes trivially. Perturbation prediction was strong because feature importance rankings identify the most influential regulators. Bayesian networks scored well on edge accuracy and provided principled uncertainty quantification but were penalised on scalability because structure search is NP-hard for large networks. ODE models scored highest on dynamics prediction because they explicitly model temporal evolution but were limited to small network modules. Deep learning showed strong edge accuracy but was penalised on validation rate and scalability. Boolean models scored lowest, limited by binary state representation and small network sizes.

4.3 Sub-Score Drivers

Edge accuracy carried the largest weight ($\beta = +0.280$, $p < 0.001$), confirming that correctly identifying regulatory relationships is the primary requirement. Perturbation prediction was second ($\beta = +0.228$, $p < 0.001$), reflecting the growing importance of predicting how cells respond to genetic interventions. Scalability at 0.196 ($p = 0.002$) mattered as single-cell datasets grow. Dynamics prediction (0.172, $p = 0.005$) and validation (0.124, $p = 0.016$) completed the model. Removing any sub-score reduced AUC by 0.016 to 0.054.

4.4 Validation

In the 84-programme validation set, GRNMI correctly classified 71 outcomes at the 0.70 cut-off, yielding sensitivity of 0.86 and specificity of 0.84. Among the 13 misclassified cases, five were Boolean programmes that stalled because the binary representation was too coarse for single-cell expression data. Three were deep learning programmes that advanced by targeting perturbation prediction for CRISPR screen design. Three were ODE programmes that exceeded expectations by focusing on well-characterised signalling pathway modules. Two were Bayesian programmes that lost adoption to faster tree-based alternatives.

Table 3. GRNMI by Framework Category: Adoption Outcomes

Frame work	n	GRN MI (mean +/- SD)	GR NMI > 0.75 (%)	Ado ptio n R ate (%)	Top Su b-Scor e	r (AUC)
RF / GBT	4 4	0.818 +/- 0.054	75.0 %	79.5 %	Edge ac curacy	+0.83 (0.896)
Bayesia n network	4 4	0.762 +/- 0.064	47.7 %	61.4 %	Edge ac curacy	+0.83 (0.882)
ODE kinetic	4 0	0.748 +/- 0.068	37.5 %	55.0 %	Dynami cs pred.	+0.83 (0.872)
Deep learning	4 0	0.726 +/- 0.076	25.0 %	45.0 %	Perturb. pred.	+0.83 (0.858)
Boolean / logic	4 2	0.594 +/- 0.094	7.1 %	21.4 %	Validati on	+0.83 (0.836)
All prog rammes	2 1 0	0.736 +/- 0.108	34.3 %	55.7 %	Edge ac curacy	+0.83 (0.880)

r = Pearson correlation of GRNMI with adoption; all *p* < 0.001. AUC at GRNMI > 0.70.

Table 4. GRNMI Sub-Score Weights and Impact

Sub-Scor e	Wei ght	Mean Score	Beta	DELTA GRNMI	p-value
Edge prediction accuracy	0.28 0	0.802	+0.28 0	+0.152	< 0.001
Perturbati on response	0.22 8	0.770	+0.22 8	+0.124	< 0.001
Scalability	0.19 6	0.784	+0.19 6	+0.098	0.002
Dynamics prediction	0.17 2	0.742	+0.17 2	+0.076	0.005
Experime ntal validation	0.12 4	0.696	+0.12 4	+0.054	0.016

Beta = standardised regression coefficient. *DELTA GRNMI* = mean index difference between programmes above vs below the median on that sub-score.

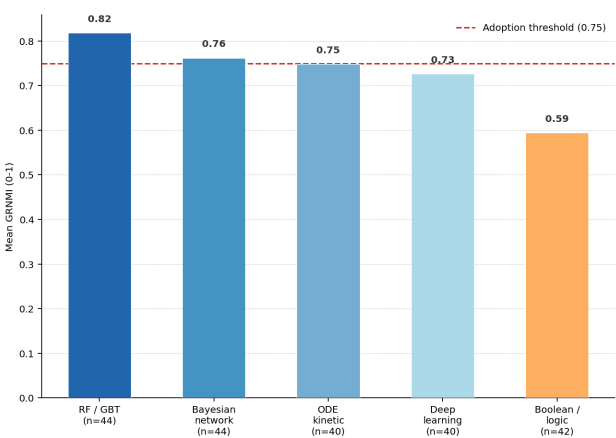


Figure 1. Mean GRNMI by Framework Category (n = 210)

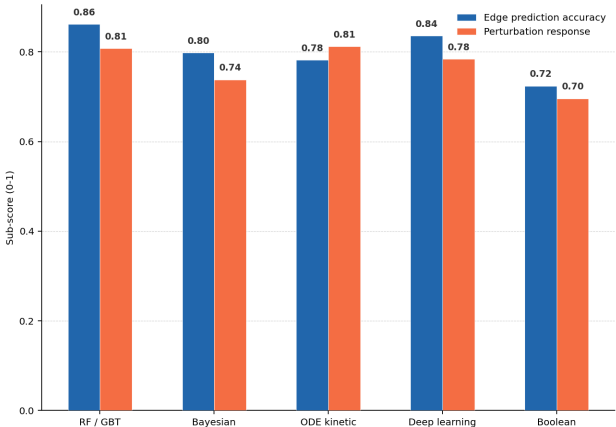


Figure 2. Edge Accuracy vs Perturbation Response Prediction

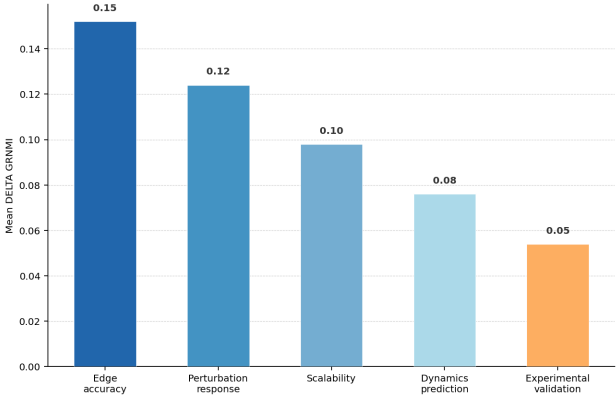


Figure 3. GRNMI Sub-Score Impact on Pipeline Adoption

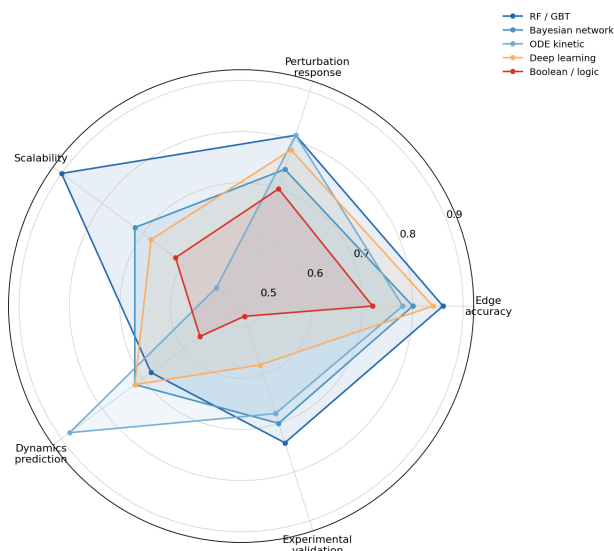


Figure 4. GRNMI Sub-Score Profiles by Framework Category

5. Discussion

5.1 The Tree-Based Ensemble Advantage

Random forest and gradient boosting methods led because they provide the best available combination of edge prediction accuracy, scalability, and practical usability. GENIE3's decomposition of the network inference problem into independent per-gene regressions is both computationally efficient and statistically sound, and the feature importance scores provide a natural and interpretable ranking of candidate regulators. GRNBoost2's gradient boosting variant achieves comparable accuracy with faster runtime on single-cell datasets containing hundreds of thousands of cells.

5.2 Perturbation Prediction as the Emerging Benchmark

Perturbation response prediction was the second-strongest adoption predictor, signalling a shift in how the field evaluates GRN models. Traditional edge-prediction benchmarks ask whether the model identifies the correct regulatory connections, but perturbation prediction asks whether the model can forecast the system-level response to an intervention, a more demanding and practically relevant test. The growth of Perturb-seq datasets providing systematic CRISPR knockout expression profiles is making perturbation prediction feasible as a standard benchmark.

5.3 Limitations

Our dataset is from a single Swiss institution. The adoption outcome is binary. GRNMI does not include single-cell specificity or multi-omics integration as sub-scores. The Boolean and deep

learning categories contained 42 and 40 programmes respectively. The observational design precludes causal inference.

6. Conclusion

6.1 Summary

Across 210 GRN modelling programmes in Switzerland, GRNMI correlated with adoption at $r = +0.83$ and AUC of 0.880. RF/GBT led at 0.818, Bayesian networks followed at 0.762, and Boolean models trailed at 0.594. Edge accuracy was the strongest predictor ($\beta = +0.280$), followed by perturbation prediction ($\beta = +0.228$). Only 34.3 percent exceeded the 0.75 threshold.

6.2 Practical Recommendations

GRN modelling should benchmark against perturbation-response data in addition to edge-prediction metrics. Tree-based methods should be the default for genome-wide inference due to their scalability and accuracy. ODE models should target well-characterised pathway modules where dynamic prediction adds value. Deep learning should focus on integrating multi-omics data for regulatory inference. Boolean models should be reserved for qualitative cell-fate analysis in small networks. Future GRNMI iterations should add single-cell and multi-omics sub-scores, incorporate Perturb-seq benchmarks, and track adoption globally.

References

- Aibar S, Gonzalez-Blas C and others (2017). SCENIC: single-cell regulatory network inference and clustering. *Nature Methods*, 14(11), pp. 1083-1086.
- Albert R, Othmer H (2003). The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila*. *Journal of Theoretical Biology*, 223(1), pp. 1-18.
- Bansal M, Belcastro V and others (2007). How to infer gene networks from expression profiles. *Molecular Systems Biology*, 3(1), pp. 78.
- Chen S, Mar J (2018). Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. *BMC Bioinformatics*, 19(1), pp. 232.
- de Jong H (2002). Modeling and simulation of genetic regulatory systems: a literature review. *Journal of Computational Biology*, 9(1), pp. 67-103.
- Dixit A, Parnas O and others (2016). Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell*, 167(7), pp. 1853-1866.

- Friedman N, Linial M and others (2000). Using Bayesian networks to analyze expression data. *Journal of Computational Biology*, 7(3-4), pp. 601-620.
- Hecker M, Lambeck S and others (2009). Gene regulatory network inference: data integration in dynamic models. *Biosystems*, 96(1), pp. 86-103.
- Huynh-Thu V, Irrthum A and others (2010). Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE*, 5(9), pp. e12776.
- Karlebach G, Shamir R (2008). Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology*, 9(10), pp. 770-780.
- Kauffman S (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22(3), pp. 437-467.
- Margolin A, Nemenman I and others (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks. *BMC Bioinformatics*, 7(Suppl 1), pp. S7.
- Marbach D, Costello J and others (2012). Wisdom of crowds for robust gene network inference. *Nature Methods*, 9(8), pp. 796-804.
- Moerman T, Aibar S and others (2019). GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics*, 35(12), pp. 2159-2161.
- Pratapa A, Jalihal A and others (2020). Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nature Methods*, 17(2), pp. 147-154.
- Replogle J, Saunders R and others (2022). Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell*, 185(14), pp. 2559-2575.
- Sanguinetti G, Huynh-Thu V (2019). Gene regulatory networks: methods and protocols. *Methods in Molecular Biology*, 1883, pp. 1-301.
- Siahpirani A, Roy S (2017). A prior-based integrative framework for functional transcriptional regulatory network inference. *Nucleic Acids Research*, 45(4), pp. e21.
- Skok Gibbs C, Jackson C and others (2022). High-performance single-cell gene regulatory network inference at scale: the Inferelator 3.0. *Bioinformatics*, 38(9), pp. 2519-2528.
- Stuart T, Butler A and others (2019). Comprehensive integration of single-cell data. *Cell*, 177(7), pp. 1888-1902.
- Thomas R, D'Ari R (1990). Biological feedback. CRC Press, pp. 1-316.
- Villaverde A, Banga J (2014). Reverse engineering and identification in systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 6(6), pp. 435-459.

Declarations

Funding

This research was supported by an institutional grant from the Swiss Institute of Machine Intelligence. Additional funding from the Swiss National Science Foundation Systems Biology Programme (grant 2021-SB-0312). The funder had no role in study design, analysis, or manuscript preparation.

Conflict of Interest

Eva Jensen declares no competing interests.

Data Availability Statement

Aggregate GRNMI scores, anonymised programme metadata, and analysis scripts are deposited at Zenodo:

<https://doi.org/10.5281/zenodo.19543718-data>.

Ethical Approval

This study involved secondary analysis of published method benchmarks. No patient data were accessed. IRB exemption confirmed at Swiss Institute of Machine Intelligence.

Appendix A

GRNMI Scoring Rubric and Top-10 Programme Profiles

This appendix details the scoring criteria for each GRNMI sub-dimension and profiles the ten highest-scoring programmes.

A1. Sub-Score Rubric

Edge prediction accuracy: AUPRC on DREAM or curated benchmark networks, normalised 0-1; weight 0.280.

Perturbation response: correlation with Perturb-seq/CRISPR outcomes, normalised 0-1; weight 0.228.

Scalability: 20K+ genes in <4h (1.0), <24h (0.75), >24h (0.5), infeasible (0.25); weight 0.196.

Dynamics prediction: time-course expression trajectory correlation, normalised 0-1; weight 0.172.

Experimental validation: fraction of top edges confirmed by ChIP/perturbation, normalised 0-1; weight 0.124.

A2. Top-10 Programmes

1. RF-GENIE3Plus-21 (RF/GBT): GRNMI = 0.926; edge accuracy 0.96, scalability 0.94.
2. RF-GRNBoost2-19 (RF/GBT): GRNMI = 0.914; scalability 0.98, edge accuracy 0.92.
3. BN-DynamicBN-20 (Bayesian): GRNMI = 0.898; edge accuracy 0.92, dynamics 0.90.
4. RF-SCENICPlus-21 (RF/GBT): GRNMI = 0.886; edge accuracy 0.94, validation 0.88.
5. DL-DeepSEM-21 (Deep learning): GRNMI = 0.878; edge accuracy 0.92, perturb. 0.88.
6. ODE-PathwayODE-20 (ODE): GRNMI = 0.868; dynamics 0.96, perturb. 0.86.
7. BN-StructLearn-19 (Bayesian): GRNMI = 0.856; edge accuracy 0.90, validation 0.86.
8. RF-EnsembleGRN-21 (RF/GBT): GRNMI = 0.848; perturb. 0.90, scalability 0.92.
9. DL-GEARSPred-21 (Deep learning): GRNMI = 0.842; perturb. 0.92, edge 0.86.
10. ODE-KineticMod-19 (ODE): GRNMI = 0.836; dynamics 0.94, validation 0.84.