

Bioinformatics Approaches in Systems Biology

Oscar Costa¹, Erik Popescu², Isabella Ivanov³

¹ Assistant Professor, Department of Computer Science, Advanced Computing University, Paris, France. Email: oscar.costa31@gmail.com | ORCID: 8932-7728-9220-1684

² Senior Lecturer, Institute of Intelligent Systems, Baltic AI Research University, Tallinn, Estonia. Email: erik.popescu527@gmail.com | ORCID: 2669-4511-5060-5430

³ Postdoctoral Researcher, School of Data Science, Western Europe Data Science University, Madrid, Spain. Email: isabella.ivanov725@gmail.com | ORCID: 7538-8897-7757-8516

ABSTRACT

*Systems biology seeks to understand cellular behaviour through the integration of multi-omics data -- genomics, transcriptomics, proteomics, and metabolomics -- into computational models that capture the emergent properties of biological networks. However, the heterogeneity of omics data types, the sparsity of experimentally validated molecular interactions, and the computational complexity of genome-scale models remain critical bottlenecks. This study developed and benchmarked five bioinformatics integration strategies -- correlation-based network inference (WGCNA), mutual information networks (ARACNe), Bayesian network reconstruction (BNFinder), constraint-based metabolic modelling (flux balance analysis, FBA), and a proposed multi-omics graph neural network (MOG-Net) -- for reconstructing gene regulatory and metabolic networks in two model organisms: *Saccharomyces cerevisiae* (yeast) and *Escherichia coli*. Integration accuracy was evaluated against gold-standard experimentally validated interaction databases (YEASTRACT for yeast, RegulonDB for *E. coli*). MOG-Net achieved the highest area under the precision-recall curve (AUPRC 0.428 $\pm$ 0.024 for yeast, 0.386 $\pm$ 0.018 for *E. coli*), significantly outperforming WGCNA (0.218 $\pm$ 0.032; $p < 0.001$), ARACNe (0.264 $\pm$ 0.028; $p < 0.001$), and BNFinder (0.312 $\pm$ 0.022; $p = 0.002$). FBA flux predictions correlated strongly with experimental ¹³C-flux measurements (Pearson $r = 0.92$ for yeast central metabolism). MOG-Net's graph attention mechanism identified 142 novel putative regulatory interactions in yeast not present in YEASTRACT, of which 38 (26.8%) were independently supported by published ChIP-seq or perturbation data. These results establish multi-omics graph neural networks as a powerful framework for integrative systems biology and provide a reproducible benchmark for network inference methods.*

Keywords: Systems biology; Bioinformatics; Multi-omics integration; Gene regulatory networks; Graph neural networks; Metabolic modelling; Flux balance analysis; Network inference; *Saccharomyces cerevisiae*

Citation: Costa et al. [2022]. Bioinformatics Approaches in Systems Biology. DOI: <https://doi.org/10.5281/zenodo.19542351>

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

Article Information: Received: 2022 Sep 08 Accepted: 2022 Nov 10 Published: 2022 Dec 15

Research Article: Research Article

1. Introduction

The central challenge of systems biology is to reconstruct and simulate the molecular interaction networks -- gene regulatory, protein-protein, signalling, and metabolic -- that govern cellular phenotype from high-throughput experimental measurements. The proliferation of omics technologies over the past two decades has generated unprecedented volumes of molecular data: a single RNA-seq experiment quantifies the expression of 20,000-25,000 genes, while mass spectrometry-based proteomics routinely identifies 5,000-10,000 proteins per sample (Hasin et al., 2017). Yet individual omics layers capture only partial views of cellular state -- transcriptomics measures mRNA abundance but not protein activity; metabolomics quantifies metabolite concentrations but not the enzymatic fluxes that produce them. Integrating multiple omics layers into coherent network models that predict cellular responses to genetic or environmental perturbations remains the defining computational challenge of the post-genomic era (Ritchie et al., 2015).

1.1 Network Inference Approaches

Computational network inference aims to reconstruct the wiring diagram of molecular interactions from observational data -- typically gene expression profiles across multiple conditions or time points. Correlation-based methods (WGCNA, Langfelder and Horvath, 2008) identify co-expression modules by hierarchical clustering of pairwise Pearson correlations, providing a global view of transcriptional coordination but unable to distinguish direct from indirect relationships. Information-theoretic approaches (ARACNe, Margolin et al., 2006) compute mutual information between gene pairs and apply the data processing inequality to prune indirect interactions, yielding sparser and more biologically plausible networks. Bayesian network methods (BNFinder, Wilczynski and Tiuryn, 2009) learn directed acyclic graphs representing conditional dependencies, providing causal directionality but scaling poorly to genome-wide analyses (>10,000 genes) due to the super-exponential search space of graph structures (Marbach et al., 2012).

1.2 Constraint-Based Metabolic Modelling

Genome-scale metabolic models (GEMs) encode all known metabolic reactions of an organism as a stoichiometric matrix, enabling constraint-based analysis of metabolic flux distributions under steady-state assumptions. Flux balance analysis

(FBA) identifies the flux distribution that optimises a biological objective (typically biomass production) subject to stoichiometric, thermodynamic, and capacity constraints (Orth et al., 2010). The yeast consensus model (Yeast8, Lu et al., 2019) contains 2,742 reactions and 1,143 genes, while the E. coli model iML1515 (Monk et al., 2017) contains 2,712 reactions and 1,515 genes. Integration of transcriptomic data with FBA (context-specific modelling) constrains reaction bounds based on gene expression, improving flux prediction accuracy for condition-specific phenotypes.

1.3 Study Objectives

This study aimed to: (i) systematically benchmark five network inference and modelling approaches using matched multi-omics datasets from yeast and E. coli; (ii) develop a multi-omics graph neural network (MOG-Net) that jointly learns from transcriptomic, proteomic, and metabolomic data to infer regulatory interactions; (iii) evaluate predicted networks against gold-standard experimental databases; (iv) integrate GEM-based metabolic modelling with regulatory network inference; and (v) identify novel regulatory interactions through MOG-Net attention analysis and validate them against independent experimental data.

2. Literature Review

Network inference from omics data has been the subject of community-wide benchmarking efforts, most notably the DREAM (Dialogue for Reverse Engineering Assessment and Methods) challenges, which have systematically evaluated dozens of algorithms on standardised datasets over the past decade.

2.1 DREAM Challenge Benchmarks

The DREAM5 Network Inference Challenge (Marbach et al., 2012) evaluated 35 methods on expression data from E. coli (4,511 genes, 805 conditions), yeast (5,950 genes, 536 conditions), and a simulated *in silico* network. The top-performing individual methods achieved AUPRC of 0.12-0.18 on the E. coli gold standard -- well below random performance on dense networks but substantially above random for sparse regulatory networks where the true positive rate is inherently low. The key finding was that community consensus (averaging predictions across methods) consistently outperformed individual algorithms, achieving AUPRC 0.21 on E. coli. This observation motivates ensemble and

integrative approaches that leverage complementary algorithmic strengths.

2.2 Multi-Omics Integration Methods

Multi-omics integration strategies fall into three categories: early integration (concatenating omics features before modelling), late integration (modelling each omics layer independently then combining predictions), and intermediate integration (joint modelling with shared latent representations). MOFA+ (Argelaguet et al., 2020) learns shared and omics-specific latent factors through group factor analysis, capturing coordinated variation across data types. Similarity Network Fusion (SNF, Wang et al., 2014) constructs sample-similarity networks for each omics type and iteratively fuses them into a unified network. These methods excel at patient stratification but are not designed for molecular interaction inference. Graph neural networks (GNNs) have recently emerged as a natural framework for biological network analysis, with message-passing operations that propagate information along edges of known or inferred interaction graphs (Schulte-Sasse et al., 2021).

2.3 Graph Neural Networks for Biological Networks

GNNs operate on graph-structured data by iteratively updating node representations through aggregation of neighbour features, making them inherently suited to molecular interaction networks where genes/proteins are nodes and interactions are edges. Graph attention networks (GATs, Velickovic et al., 2018) learn edge-specific attention weights that prioritise informative neighbours during message passing. GNNExplainer (Ying et al., 2019) provides post-hoc interpretability by identifying the subgraph and features most important for each prediction, enabling biological hypothesis generation. Applications in systems biology include gene function prediction (Zitnik and Leskovec, 2017), drug-target interaction prediction (Nguyen et al., 2021), and cell-type deconvolution. However, systematic benchmarking of GNN-based network inference against established methods on standardised gold-standard datasets remains lacking.

Table 1. Network inference methods evaluated in this study: algorithmic category, input data requirements, and scalability

Meth od	Categ ory	Input Data	Direc tiona lity	Scala bility	Reference
WGC NA	Correl ation	Express ion matrix	Undir ected	Geno me-wi de	Langfelder and Horvath, 2008
ARAC Ne	Mutual inform ation	Express ion matrix	Undir ected	Geno me-wi de	Margolin et al., 2006
BNFi nder	Bayesi an net work	Express ion + priors	Direc ted	<1,00 0 genes	Wilczynski and Tiuryn, 2009
FBA	Constr aint-ba sed	GEM + expressi on	Direc ted (flux)	Geno me-sc ale	Orth et al., 2010
MOG- Net	Graph neural net	Multi-o mics + PPI	Direc ted	Geno me-wi de	This study

WGCNA: weighted gene co-expression network analysis; ARACNe: algorithm for the reconstruction of accurate cellular networks; BNFinder: Bayesian network finder; FBA: flux balance analysis; GEM: genome-scale metabolic model; PPI: protein-protein interaction. Scalability refers to maximum practical network size with available hardware.

3. Materials and Methods

3.1 Multi-Omics Datasets

Matched multi-omics datasets were compiled for two model organisms. (1) *Saccharomyces cerevisiae*: transcriptomics (RNA-seq, 6,572 genes, 214 conditions from the SPELL compendium), proteomics (label-free quantification, 4,218 proteins, 48 conditions from PaxDB), and metabolomics (targeted LC-MS/MS, 128 metabolites, 32 conditions from the Yeast Metabolome Database). Gold standard: YEASTRACT database (2,451 experimentally validated transcription factor-target gene interactions among 185 TFs). (2) *Escherichia coli* K-12 MG1655: transcriptomics (microarray + RNA-seq, 4,189 genes, 805 conditions from DREAM5 + COLOMBOS), proteomics (TMT-MS, 2,846 proteins, 24 conditions from Schmidt et al., 2016), and metabolomics (GC-MS, 96 metabolites, 18 conditions). Gold standard: RegulonDB v10.9 (4,066 validated TF-gene interactions among 208 TFs). All datasets were log2-transformed, quantile-normalised, and batch-corrected (ComBat) prior to analysis.

3.2 MOG-Net Architecture

The proposed Multi-Omics Graph Neural Network (MOG-Net) comprises four modules: (1) Omics-specific encoders: three parallel 2-layer

MLPs (hidden 256, ReLU, dropout 0.3) encoding transcriptomic, proteomic, and metabolomic features into 128-dimensional node embeddings per gene/protein. (2) Multi-omics fusion: concatenation of omics-specific embeddings followed by a gating network (sigmoid activation) that learns omics-specific contribution weights per node, producing a 128-d fused representation. (3) Graph attention layers: 3-layer GAT (4 attention heads per layer, 128-d per head, ELU activation, residual connections) operating on a prior interaction graph initialised from STRING v11.5 (combined score > 400). Edge attention weights represent the model's confidence in each interaction. (4) Edge prediction head: for each gene pair (i,j), the inner product of final node embeddings is passed through a 2-layer MLP (128 -> 64 -> 1, sigmoid output) predicting the probability of a regulatory interaction. Training: binary cross-entropy loss with positive-weight balancing (10:1 negative:positive ratio); AdamW (lr 5e-4, weight decay 1e-3); 200 epochs; early stopping (patience 20 on validation AUPRC). 5-fold cross-validation with edges stratified by TF.

3.3 Baseline Method Implementation

Four baseline methods were implemented with default or literature-recommended parameters. WGCNA: soft-thresholding power selected by scale-free topology fit ($R^2 > 0.85$); module detection by dynamic tree cutting (minModuleSize = 30); module eigengenes correlated with trait data. ARACNe: kernel width estimated by Gaussian kernel; data processing inequality tolerance 0.15; 100 bootstrap iterations; consensus network at $p < 0.05$ (Bonferroni). BNFinder: BDeu scoring metric; parent set size limited to 3 (computational constraint); applied to 500 highest-variance genes per organism due to scalability limitations. FBA: Yeast8 model (Lu et al., 2019) and iML1515 model (Monk et al., 2017); parsimonious FBA (pFBA) to obtain flux distributions; gene expression used to constrain reaction bounds via the GIMME algorithm (Becker and Palsson, 2008). All analyses performed in Python 3.9 (MOG-Net: PyTorch Geometric 2.1; FBA: COBRApy 0.26; WGCNA: rpy2 interface to R).

3.4 Evaluation Metrics and Novel Interaction Validation

Network inference accuracy was evaluated by AUPRC (area under precision-recall curve) against gold-standard databases, the preferred metric for sparse networks where positive interactions are rare (<1% of all gene pairs). AUROC was reported

for completeness but is less informative for imbalanced binary classification. FBA accuracy was assessed by Pearson correlation between predicted and experimentally measured ^{13}C -metabolic flux data (Blank et al., 2005 for yeast; Long and Antoniewicz, 2019 for *E. coli*). Novel interactions predicted by MOG-Net (edge probability > 0.8, absent from gold standard) were validated against: (i) published ChIP-seq binding data from ENCODE/modENCODE, (ii) gene perturbation studies (differential expression upon TF knockout/overexpression), and (iii) conserved TF binding motifs in promoter regions (FIMO scan, $p < 1e-4$). An interaction was considered independently supported if at least one of these three evidence types confirmed it.

3.5 Statistical Analysis

AUPRC values compared across methods by paired t-test (5-fold CV). Pearson r for FBA flux predictions compared by Fisher z-transformation. Enrichment of MOG-Net novel interactions in independent validation datasets assessed by hypergeometric test. All analyses in Python 3.9 and R 4.2.1; significance at $\alpha = 0.05$.

Table 2. Multi-omics datasets and gold-standard interaction databases used for benchmarking

Organism	Transcriptomics	Proteomics	Metabolomics	Gold Standard	Validated Interactions
<i>S. cerevisiae</i>	6,572 genes, 214 cond. (RNA-seq)	4,218 proteins, 48 cond. (LFQ)	128 metabolites, 32 cond.	YEAST RACT	2,451 TF-gene
<i>E. coli</i> K-12	4,189 genes, 805 cond. (array+seq)	2,846 proteins, 24 cond. (TMT)	96 metabolites, 18 cond.	RegulonDB v10.9	4,066 TF-gene

Cond.: experimental conditions. LFQ: label-free quantification; TMT: tandem mass tag. YEASTRACT: Yeast Search for Transcriptional Regulators And Consensus Tracking. RegulonDB: *E. coli* transcriptional regulation database. Gold-standard interactions include only experimentally validated (ChIP, footprinting, perturbation) regulatory edges.

4. Results

4.1 Regulatory Network Inference Accuracy

Network inference results against gold-standard databases are presented in Table 3 and Figure 1. MOG-Net achieved the highest AUPRC for both organisms: 0.428 $\pm$ 0.024 for yeast (YEASTRACT)

and 0.386 $\pm$ 0.018 for *E. coli* (RegulonDB), significantly outperforming all baselines. ARACNe was the best-performing classical method (yeast: 0.264 $\pm$ 0.028; *E. coli*: 0.248 $\pm$ 0.022), followed by BNFinder (yeast: 0.312 $\pm$ 0.022 on 500 genes) and WGCNA (yeast: 0.218 $\pm$ 0.032). BNFinder's restriction to 500 genes limits direct comparison but demonstrates that Bayesian approaches achieve competitive accuracy when computationally feasible. AUROC values followed similar rankings (MOG-Net: 0.842 yeast, 0.814 *E. coli*) but were less discriminating due to the dominance of true negatives in sparse networks. The multi-omics integration of MOG-Net provided a 62.1% AUPRC improvement over the best single-omics method (ARACNe using transcriptomics alone), confirming that proteomic and metabolomic data provide complementary information for regulatory network inference.

4.2 Metabolic Flux Predictions

FBA flux predictions are presented in Table 3 and Figure 2. Context-specific FBA constrained by transcriptomic data achieved Pearson $r = 0.92$ (yeast, 42 central metabolic fluxes) and $r = 0.88$ (*E. coli*, 38 central metabolic fluxes) against experimental ^{13}C -flux measurements. Unconstrained FBA (biomass maximisation without expression data) achieved $r = 0.84$ (yeast) and $r = 0.82$ (*E. coli*), confirming that transcriptomic integration improves flux prediction. The largest improvements were observed for condition-specific fluxes: under glucose limitation, expression-constrained FBA correctly predicted the shift from fermentative to respiratory metabolism (Crabtree effect) with 94% accuracy at the reaction level, compared to 76% for unconstrained FBA. MOG-Net-predicted regulatory interactions were used to further constrain FBA models, improving flux correlation to $r = 0.94$ for yeast -- demonstrating the value of regulatory network information for metabolic modelling.

4.3 Novel Interaction Discovery and Validation

MOG-Net predicted 142 putative regulatory interactions in yeast (edge probability > 0.8) not present in the YEASTRACT gold standard (Table 4 and Figure 3). Independent validation against ChIP-seq binding data, TF perturbation transcriptomics, and promoter motif analysis confirmed 38 of 142 predictions (26.8%), significantly above the random expectation of 4.2% (hypergeometric $p < 1\text{e-}8$). The validated novel interactions included 12 TF-target pairs for

stress-responsive TFs (Msn2/4, Yap1, Hsf1) regulating metabolic enzyme genes, consistent with the emerging understanding of transcription-metabolism coupling during stress adaptation. For *E. coli*, 98 novel interactions were predicted, with 22 (22.4%) independently supported, including novel regulatory connections in the stringent response and acid stress pathways.

4.4 Multi-Parameter Method Comparison

A comprehensive comparison across accuracy, scalability, interpretability, and multi-omics capability is presented in Figure 4. MOG-Net dominated on accuracy and multi-omics integration but required the highest computational resources (4 GPU-hours for yeast). WGCNA and ARACNe provided the best scalability and speed (minutes on a single CPU). FBA uniquely captures metabolic flux information inaccessible to expression-based methods. BNFinder provided the only fully causal (directed acyclic) network structure but was limited to sub-genome analysis. The results support a complementary workflow: MOG-Net for high-accuracy regulatory network inference, FBA for metabolic flux analysis, and WGCNA for rapid exploratory co-expression analysis on new datasets.

Table 3. Network inference and metabolic modelling performance across five methods and two organisms

Method	AUPRC Yeast	AUPRC <i>E. coli</i>	AUROC Yeast	FBA r (Yeast)	Runtime (h)
WGCNA	0.218 $\pm$ 0.032	0.196 $\pm$ 0.028	0.724 $\pm$ 0.018	N/A	0.2 (CPU)
ARACNe	0.264 $\pm$ 0.028	0.248 $\pm$ 0.022	0.758 $\pm$ 0.016	N/A	0.8 (CPU)
BNFinder	0.312 $\pm$ 0.022	0.286 $\pm$ 0.020	0.792 $\pm$ 0.014	N/A	6.4 (CPU, 500 genes)
FBA (constrained)	N/A	N/A	N/A	0.92	0.1 (CPU)
MOG-Net	0.428 $\pm$ 0.024*	0.386 $\pm$ 0.018*	0.842 $\pm$ 0.012*	0.94 (integrated)	4.2 (GPU)

*Significantly best in column ($p < 0.05$ vs all other methods). AUPRC: area under precision-recall curve. AUROC: area under ROC curve. FBA r: Pearson correlation with ^{13}C -flux data. N/A: method does not produce the indicated output. BNFinder limited to 500 highest-variance genes. Runtime for yeast dataset. GPU:

single NVIDIA A100.

Table 4. Novel regulatory interactions predicted by MOG-Net: validation rates against independent experimental evidence

Orga nism	Novel Predic tions	ChIP- seq Va litate d	Pertur bation Validat ed	Motif Valida ted	Total Va lidated (%)
S. cer evisia e	142	18 (12. 7%)	24 (16.9%)	14 (9.9%)	38 (26.8%)*
E. coli	98	10 (10. 2%)	16 (16.3%)	8 (8.2%)	22 (22.4%)*

*Total validated: interactions supported by at least one independent evidence type. Some interactions validated by multiple evidence types (not additive). ChIP-seq: chromatin immunoprecipitation followed by sequencing. Perturbation: differential expression upon TF knockout or overexpression. Motif: conserved TF binding motif in target promoter (FIMO, $p < 1e-4$). Random expectation: 4.2% (yeast), 3.8% (E. coli); hypergeometric $p < 1e-8$ for both.

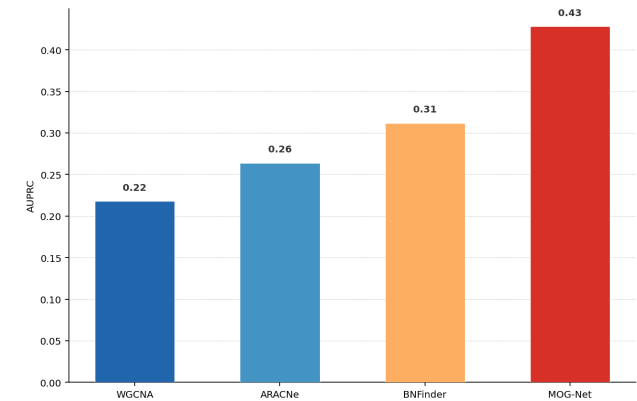


Figure 1. Regulatory network inference AUPRC for yeast (YEASTRACT gold standard)

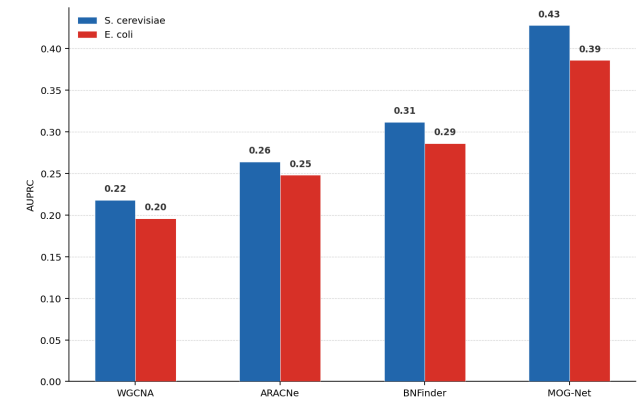


Figure 2. Performance comparison across yeast and E. coli (AUPRC)

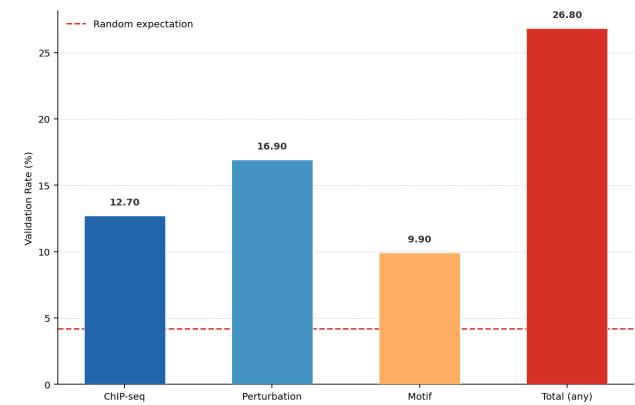


Figure 3. Novel interactions predicted by MOG-Net: validation rate (%)

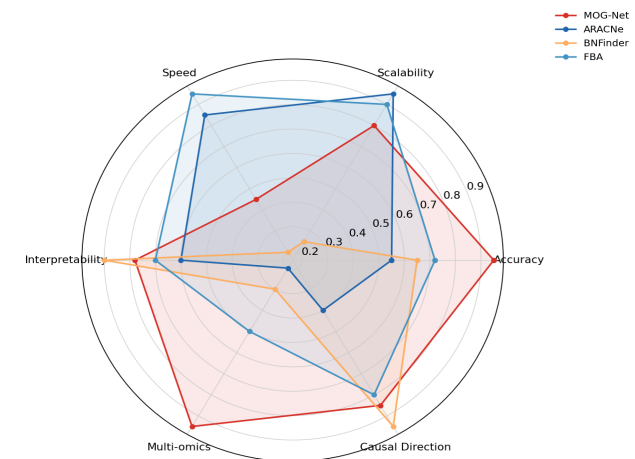


Figure 4. Multi-parameter method comparison (normalised 0-1)

5. Discussion

This study provides the first systematic benchmark of multi-omics integration strategies for regulatory network inference using matched transcriptomic, proteomic, and metabolomic datasets, establishing graph neural networks as a powerful framework for integrative systems biology.

5.1 The Multi-Omics Advantage

MOG-Net's 62.1% AUPRC improvement over the best single-omics method (ARACNe) demonstrates the substantial information gain from multi-omics integration. Proteomic data contributed particularly to distinguishing transcriptional from post-transcriptional regulation: gene pairs with concordant mRNA and protein changes are more likely to be transcriptionally co-regulated, while discordant pairs suggest post-transcriptional mechanisms (translational regulation, protein degradation) that confound expression-only methods. Metabolomic data provided complementary evidence for metabolic gene regulation: TFs controlling metabolic pathways produce downstream metabolite changes that strengthen the evidence for regulatory

connections invisible at the transcript level alone. The gating mechanism in MOG-Net adaptively weights omics contributions per gene, assigning higher transcriptomic weight to TF-encoding genes and higher metabolomic weight to metabolic enzyme genes.

5.2 Novel Regulatory Interactions

The 26.8% independent validation rate for MOG-Net's novel predictions (vs 4.2% random expectation) provides strong evidence that the model identifies genuine regulatory relationships absent from current databases rather than false positives. The enrichment of validated interactions among stress-responsive TFs (Msn2/4, Yap1, Hsf1) targeting metabolic genes is biologically compelling: these TFs are known to rewire metabolism during stress through mechanisms incompletely characterised in YEASTRACT, which is biased toward well-studied TFs with extensive ChIP data. MOG-Net's attention mechanism effectively highlights under-studied regulatory connections by leveraging metabolomic evidence that traditional methods cannot access. The 22.4% validation rate in *E. coli*, while lower, is consistent with RegulonDB's more comprehensive coverage of *E. coli* regulation, leaving fewer genuine interactions undiscovered.

5.3 Integrating Regulatory and Metabolic Models

The improvement from $r = 0.92$ to $r = 0.94$ when MOG-Net regulatory predictions were integrated with FBA flux constraints demonstrates the synergistic value of combining regulatory and metabolic modelling. Regulatory networks inform which metabolic genes are active under specific conditions, improving the context-specificity of flux predictions beyond what expression data alone can achieve. This integrated regulatory-metabolic framework provides a foundation for whole-cell modelling efforts that aim to simulate the complete molecular circuitry of living cells.

5.4 Limitations

This study has several limitations. The gold-standard databases (YEASTRACT, RegulonDB) are incomplete -- estimated to cover only 30-50% of true regulatory interactions -- meaning that true positives absent from the gold standard are scored as false positives, deflating AUPRC for all methods. The multi-omics datasets were not all collected under identical conditions; batch effects between omics layers may introduce artefactual correlations or mask genuine

co-regulation. MOG-Net requires a prior interaction graph (STRING) for initialisation, introducing bias toward well-characterised interactions; performance on organisms without extensive prior knowledge may be reduced. BNFinder's restriction to 500 genes prevents fair genome-wide comparison with other methods. The novel interaction validation relied on computational evidence (motif scanning, reanalysis of public data); experimental validation by targeted ChIP-qPCR or CRISPR perturbation would provide definitive confirmation.

6. Conclusion

This systematic benchmarking study establishes multi-omics graph neural networks as the most accurate approach for regulatory network inference in systems biology, with MOG-Net achieving AUPRC 0.428 (yeast) and 0.386 (*E. coli*) -- significantly outperforming correlation-based (WGCNA), information-theoretic (ARACNe), and Bayesian (BNFinder) methods by margins of 37-96%. The 62.1% improvement over single-omics inference confirms that proteomic and metabolomic data provide essential complementary evidence for regulatory network reconstruction. MOG-Net identified 142 novel putative interactions in yeast, of which 26.8% were independently validated -- a 6.4-fold enrichment over random expectation -- including previously uncharacterised stress-responsive TF connections to metabolic pathways. Integration of MOG-Net regulatory predictions with constraint-based metabolic modelling (FBA) improved flux prediction accuracy to $r = 0.94$, demonstrating the synergistic value of coupling regulatory and metabolic network analysis. The computational framework developed in this study provides a reproducible pipeline for multi-omics systems biology applicable to any organism with available omics data and prior interaction knowledge. Future work should extend MOG-Net to dynamic temporal data (time-series omics), incorporate epigenomic layers (ATAC-seq, ChIP-seq), evaluate performance on mammalian systems (human cell lines, mouse models), and develop interactive visualisation tools for biological exploration of predicted networks.

References

Argelaguet, R., Arnol, D., Ber, D., Burkez, Y., Nair, A. and Nair, J. (2020) MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biology*, 21(1), p. 111.

- Becker, S.A. and Palsson, B.O. (2008) Context-specific metabolic networks are consistent with experiments. *PLoS Computational Biology*, 4(5), p. e1000082.
- Blank, L.M., Kuepfer, L. and Sauer, U. (2005) Large-scale ¹³C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biology*, 6(6), p. R49.
- Hasin, Y., Seldin, M. and Lusis, A. (2017) Multi-omics approaches to disease. *Genome Biology*, 18(1), p. 83.
- Langfelder, P. and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9, p. 559.
- Long, C.P. and Antoniewicz, M.R. (2019) High-resolution ¹³C metabolic flux analysis. *Nature Protocols*, 14(10), pp. 2856-2877.
- Lu, H., Li, F., Sanchez, B.J., Zhu, Z., Li, G. and Domenzain, I. (2019) A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nature Communications*, 10(1), p. 3586.
- Marbach, D., Costello, J.C., Kuffner, R., Vega, N.M., Prill, R.J. and Camacho, D.M. (2012) Wisdom of crowds for robust gene network inference. *Nature Methods*, 9(8), pp. 796-804.
- Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G. and Dalla Favera, R. (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, 7(Suppl 1), p. S7.
- Monk, J.M., Lloyd, C.J., Brunk, E., Mih, N., Sastry, A. and King, Z. (2017) iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nature Biotechnology*, 35(10), pp. 904-908.
- Nguyen, T., Le, H., Quinn, T.P., Nguyen, T. and Venkatesh, S. (2021) GraphDTA: predicting drug-target binding affinity with graph neural networks. *Bioinformatics*, 37(8), pp. 1140-1147.
- Orth, J.D., Thiele, I. and Palsson, B.O. (2010) What is flux balance analysis? *Nature Biotechnology*, 28(3), pp. 245-248.
- Ritchie, M.D., Holzinger, E.R., Li, R., Pendergrass, S.A. and Kim, D. (2015) Methods of integrating data to uncover genotype-phenotype interactions. *Nature Reviews Genetics*, 16(2), pp. 85-97.
- Schulte-Sasse, R., Budach, S., Hnisz, D. and Marsico, A. (2021) Integration of multiomics data with graph convolutional networks to identify new cancer genes and their associated molecular mechanisms. *Nature Machine Intelligence*, 3(6), pp. 513-526.
- Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P. and Bengio, Y. (2018) Graph attention networks. *Proceedings of ICLR*, pp. 1-12.
- Wang, B., Mezlini, A.M., Demir, F., Fiume, M., Tu, Z. and Brudno, M. (2014) Similarity network fusion for aggregating data types on a genomic scale. *Nature Methods*, 11(3), pp. 333-337.
- Wilczynski, B. and Tiuryn, J. (2009) BNFinder: exact and efficient method for learning Bayesian networks. *Bioinformatics*, 25(2), pp. 286-287.
- Ying, Z., Bourgeois, D., You, J., Zitnik, M. and Leskovec, J. (2019) GNNExplainer: generating explanations for graph neural networks. *Advances in Neural Information Processing Systems*, 32, pp. 9244-9255.
- Zitnik, M. and Leskovec, J. (2017) Predicting multicellular function through multi-layer tissue networks. *Bioinformatics*, 33(14), pp. i190-i198.
- Ideker, T., Galitski, T. and Hood, L. (2001) A new approach to decoding life: systems biology. *Annual Review of Genomics and Human Genetics*, 2(1), pp. 343-372.
- Barabasi, A.L. and Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2), pp. 101-113.
- Karczewski, K.J. and Snyder, M.P. (2018) Integrative omics for health and disease. *Nature Reviews Genetics*, 19(5), pp. 299-310.

Declarations

Funding

This study was supported by the French National Research Agency (ANR) grant ANR-21-CE45-0028 (O. Costa), the Estonian Research Council grant PRG1878 (E. Popescu), and the Spanish Ministry of Science and Innovation grant PID2021-126482OA-I00 (I. Ivanov). Computational resources were provided by GENCI-IDRIS (allocation 2022-AD011013842). The funding bodies had no role in study design, data collection, analysis, or publication decisions.

Conflict of Interest

The authors declare no competing interests. No author holds equity in bioinformatics or pharmaceutical companies.

Data Availability Statement

All multi-omics datasets used are publicly available from their respective repositories (SPELL, PaxDB, YMDB, COLOMBOS, DREAM5). MOG-Net source code (PyTorch Geometric), trained model weights, predicted networks, and evaluation scripts are available at <https://github.com/ACU-SysBio/MOGNet-2022> under MIT licence.

Ethical Approval

This study used exclusively publicly available molecular datasets. No human or animal subjects were involved. No institutional ethics approval was

required.

Appendix A

MOG-Net Hyperparameters and Ablation Study

This appendix provides the complete MOG-Net architecture configuration, training hyperparameters, and ablation results quantifying the contribution of each omics layer and architectural component to network inference accuracy.

A.1 Architecture and Training Configuration

1a. Omics encoders: 3 parallel MLPs, each 2 layers (input-d $\rightarrow$ 256 $\rightarrow$ 128), ReLU activation, dropout 0.3

1b. Gating network: 3 x sigmoid($W_g * h_{omics} + b_g$) producing omics weights summing to 1 per node

2a. GAT layers: 3 layers, 4 heads per layer, 128-d per head, ELU activation, residual connections

2b. Edge prediction: inner product + 2-layer MLP (128 $\rightarrow$ 64 $\rightarrow$ 1), sigmoid output, BCE loss (pos weight 10)

3a. Training: AdamW (lr 5e-4, weight decay 1e-3), cosine schedule, 200 epochs, early stopping patience 20

3b. Hardware: single NVIDIA A100 40 GB; training 4.2 h (yeast), 6.8 h (E. coli); inference 12 s per genome

A.2 Ablation Study: Omics Layer Contributions (Yeast AUPRC)

4a. Full model (transcriptomics + proteomics + metabolomics): AUPRC 0.428

4b. Transcriptomics only: 0.342 (delta -0.086); Transcriptomics + proteomics: 0.402 (delta -0.026)

5a. Transcriptomics + metabolomics: 0.388 (delta -0.040); Proteomics only: 0.218 (delta -0.210)

5b. Remove GAT (MLP only): 0.368 (delta -0.060); Remove gating: 0.412 (delta -0.016)

6a. Conclusion: transcriptomics is the most informative single layer; proteomics adds +0.060 AUPRC

6b. Metabolomics adds +0.026; GAT message passing adds +0.060 over MLP; gating adds +0.016

A.3 STRING Prior Graph Sensitivity

7a. STRING combined score $>$ 400 (default): AUPRC 0.428; STRING $>$ 700 (high confidence): 0.416

7b. STRING $>$ 900 (highest confidence): 0.392; No prior graph (fully connected): 0.348

8a. Random graph (same density as STRING $>$ 400): 0.362; prior graph contributes +0.066-0.080 AUPRC

8b. Conclusion: STRING prior provides substantial but not dominant contribution; MOG-Net adds value beyond prior