



Generation of Realistic Gene Regulatory Networks by Enriching for Feed-Forward Loops

Erik K. Zhivkopoulos^{1†}, Oleg Vavulov^{2†}, Thomas Hillerton¹ and Erik L. L. Sonnhammer^{1*}

¹Department of Biochemistry and Biophysics, Science for Life Laboratory, Stockholm University, Solna, Sweden, ²Bioinformatics Institute, St. Petersburg, Russia

OPEN ACCESS

Edited by:

Marco Pellegrini,
Italian National Research Council, Italy

Reviewed by:

Vitaly Gursky,
Ioffe Institute (RAS), Russia
Michael Mayo,
US Army Corps of Engineers,
United States

*Correspondence:

Erik L. L. Sonnhammer
erik.sonnhammer@scilifelab.se

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

Received: 15 November 2021

Accepted: 13 January 2022

Published: 10 February 2022

Citation:

Zhivkopoulos EK, Vavulov O, Hillerton T
and Sonnhammer ELL (2022)
Generation of Realistic Gene
Regulatory Networks by Enriching for
Feed-Forward Loops.
Front. Genet. 13:815692.
doi: 10.3389/fgene.2022.815692

The regulatory relationships between genes and proteins in a cell form a gene regulatory network (GRN) that controls the cellular response to changes in the environment. A number of inference methods to reverse engineer the original GRN from large-scale expression data have recently been developed. However, the absence of ground-truth GRNs when evaluating the performance makes realistic simulations of GRNs necessary. One aspect of this is that local network motif analysis of real GRNs indicates that the feed-forward loop (FFL) is significantly enriched. To simulate this properly, we developed a novel motif-based preferential attachment algorithm, FFLatt, which outperformed the popular GeneNetWeaver network generation tool in reproducing the FFL motif occurrence observed in literature-based biological GRNs. It also preserves important topological properties such as scale-free topology, sparsity, and average in/out-degree per node. We conclude that FFLatt is well-suited as a network generation module for a benchmarking framework with the aim to provide fair and robust performance evaluation of GRN inference methods.

Keywords: network biology, gene regulatory networks, gene-gene interaction, network motif structure, network generation, network simulation, benchmarking

INTRODUCTION

Understanding large-scale biological relationships between genes and the proteins they encode remains a great challenge in systems biology. The wide availability of system-level expression datasets has given rise to a variety of reverse engineering methods that aim to reconstruct the hidden regulatory gene-gene and gene-protein relationships. Such relationships form a gene regulatory network (GRN) that regulates developmental processes in organisms and controls adaptation to changes in the environment (Davidson, 2010). By contrast with other networks in biological systems, GRNs are harder to validate as the interactions that occur between genes usually involve indirect interactions through biological molecules making the interaction hard to detect and quantify. The incompleteness and scarcity of ground-truth networks results in problems when evaluating the performance of methods that seek to infer GRNs from large-scale expression data (Emmert-Streib and Dehmer, 2018).

The problem of inferring a gene regulatory network from gene expression data has received significant attention. A variety of GRN inference methods are commonly used (Margolin et al., 2006; Faith et al., 2007; Friedman et al., 2010; Huynh-Thu et al., 2010; Zavlanos et al., 2011) to tackle this problem. It was also the focus of four separate Dialogue for Reverse Engineering Assessments and Methods (DREAM) challenges, with DREAM5 being the most recent one (Marbach et al., 2012).

Newer, more advanced algorithms require not only expression data but also utilize additional information such as experimentally validated interactions and Gene Ontology terms (Chouvardas et al., 2016), structures of genomic datasets and network topology (Siahpirani and Roy, 2017), DNA binding domains of transcription factors, and promoter sequences of its putative targets (Kang et al., 2018), or use the iterative kernel PCR model (Iglesias-Martinez et al., 2021). Despite this, for most methods the performance on real experimental datasets remains modest (Marbach et al., 2012; Chen and March 2018; Pratapa et al., 2020).

Regardless of the method used, it is important to fairly assess its performance with respect to other methods. As some methods can only predict Boolean networks, assessment should be done in terms of binary error classification such as the number of false positives and false negatives. In addition to this, experimental information about transcriptional interactions is usually only available in the binary form. Boolean networks can only be defined by their topology, which is why it is essential to understand the structure of GRN graphs. It is also worth pointing out that most GRN inference methods can only predict a static network structure, which implies that *in-silico* generated GRNs should also possess biological stability.

While the true structure of real GRNs is usually not known, they tend to share some topological features: the scale-free property (Barabasi and Albert, 1999), where the node degrees follow a power-law degree distribution, and often have the small world property (Watts and Strogatz, 1998), and where nodes form distinct clusters in which they are connected to each other in lattice rings. These properties are different from random graphs where node degrees are normal distributed across all nodes in the system. Some attempts to simulate GRNs have been made by implementing methods that generate random (Watts and Strogatz, 1998; Mendes et al., 2003) or scale-free (Barabasi and Albert, 1999) graphs with given sets of parameters, but eventually methods based on the idea of subnetwork-selection from biological networks gained more popularity (Van den Bulcke et al., 2006). One example of this is GeneNetWeaver (GNW) (Schaffter et al., 2011), which was used to generate *in silico* networks for the DREAM challenges.

The regulatory dynamics of GRNs is shaped by network patterns that are more frequent in GRNs than in other networks (Milo et al., 2002; Shen-Orr et al., 2002) and may carry information-processing functions. These local patterns, or motifs, and do not result in emergence of specific patterns in gene expression but rather determine dynamical boundaries of the phase space of the system (Ahnert and Fink, 2016). It was suggested that some motifs could be particularly important for network dynamics and therefore become overrepresented and drive the evolution of the networks (Prill et al., 2005). Examples of how feed-forward loops are involved in such dynamics are ample, including sign-sensitive delay elements (Mangan et al., 2003), bi-phase response generators (Kaplan et al., 2008), band-pass filters (Sohka et al., 2009), and decoders of oscillatory signals (Zhang et al., 2016). Due to this, simulating a network structure that preserves the overrepresentation of motifs is of utmost importance for capturing realistic dynamics of GRNs. The idea

of building gene regulatory networks by using motifs as building blocks was first introduced by Abdelzaher et al. (2015a) that hypothesized that this could be important for the evolution of GRN topology in *E. coli*.

Network inference methods aim to solve the problem of finding regulatory interactions within a set of genes. This, however, doesn't imply that all edges in a reconstructed network represent physical binding between transcription factors and their respective targets. Gardner and Faith (2005) describe two groups of reverse-engineering algorithms. The first group seeks to identify regulators that directly control mRNA expression, and the second one is focused on identification of general regulatory interactions between different genes that may be indirect. Regardless of interaction type, simulated data should allow for exploring a wide range of network properties to evaluate inference algorithms performance. It was shown that FFLs are significantly overrepresented in experimentally validated transcriptional regulation databases (Lee et al., 2002; Milo et al., 2002). FFLs were also found to be significantly overrepresented in other databases of microRNAs and their predicted targets (Krek et al., 2005; Lewis et al., 2005) with Z-score range between 1.39 and 6.03 (Shalgi et al., 2007). Other TF-microRNA studies demonstrated that in the circuitry of gene regulation via intermediate microRNAs, in mouse and human, and the FFL motif is also enriched (Tsang et al., 2007). This suggests that FFL is an important signature of real GRNs that represent either direct or indirect interactions between genes.

In the present study the significance of 3-node motifs in four directed GRNs based on experimentally verified transcriptional interaction databases were evaluated. In agreement with previous studies (Lee et al., 2002; Milo et al., 2002; Boyer et al., 2005), it was found that the feed-forward loop (FFL) is the only motif that is overrepresented. This motivated us to develop a novel motif-based preferential attachment algorithm called FFLatt for simulating realistic structures of GRNs that are enriched with the FFL motif. The networks generated by FFLatt demonstrate structural properties that agree with biological GRNs, and have good robustness in stability analyses. Given their realistic properties, they are well suited for fair and robust evaluation of the performance of GRN inference algorithms.

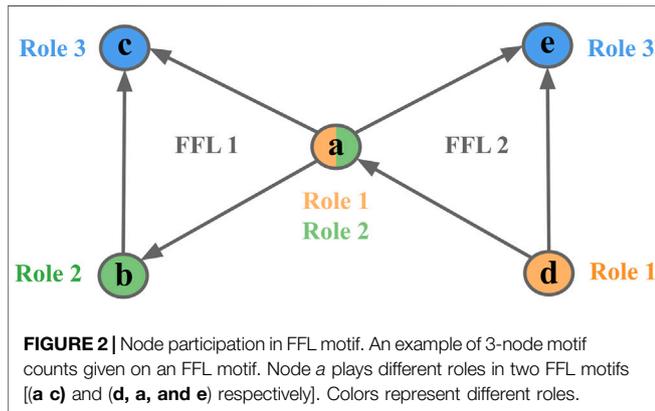
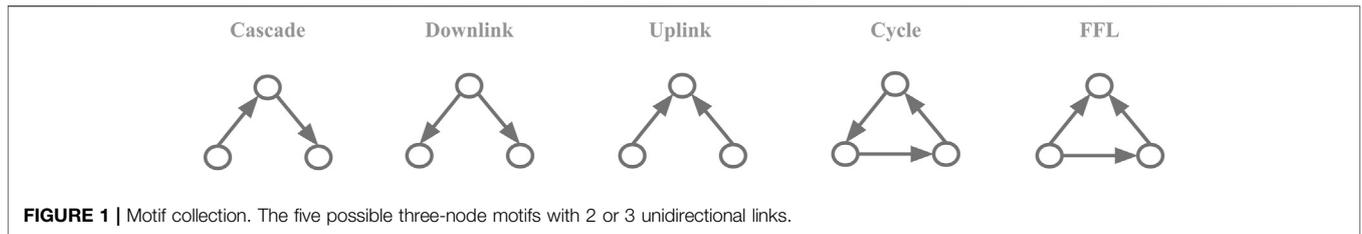
METHODS

Transcriptional Interaction Databases

Three biological databases that contain information of experimentally validated transcriptional regulation were chosen as ground-truth networks: RegulonDB (Santos-Zavaleta et al., 2019) for *E. coli* (Balaji et al., 2006), for *S. cerevisiae*, and TRRUST v2 (Han et al., 2018) for *M. musculus* and *H. sapiens* transcription factor—target regulatory relationships.

Motif-Node Participation and Motif Enrichment

We chose to test for node-motif participation for all possible connected three-node motifs with no reciprocal links between them (Figure 1). Reciprocal links were not considered as they are



very rare in the biological networks studied here. To calculate the motif-node counts, N_{real} for every node in the network we calculated the presence of a given node in all different roles of a given motif, $N(i)$. and so for a set of nodes $\{1 = 1, \dots, M\}$ in the network of size M it could be framed as:

$$N_{real} = \sum_{i=1}^M N_{role1}(i) + N_{role2}(i) + N_{role3}(i) \quad (1)$$

For example, node *a* could either participate in Role 1 (2 outgoing edges, 0 incoming), Role 2 (1 outgoing edge, 1 incoming), and Role 3 (0 outgoing edges, 2 incoming) of FFL motif 1 but at the same time participate in different role of other FFL motif 2 (Figure 2).

To test for motif enrichment, we calculated Z-score for every motif type:

$$\frac{N_{real} - \mu_{shuffled}}{\sigma_{shuffled}} \quad (2)$$

where N_{real} is the number of motif counts in the original network, $\mu_{shuffled}$ and $\sigma_{shuffled}$ are the mean and standard deviation of motif counts in the distribution of shuffled networks. Every network was shuffled with a preserved in/out-degree for all nodes until at least 80% of edges in the original network were swapped. To calculate the mean and standard deviation of motif counts in the shuffled networks every network was shuffled 10,000 times. To ensure that the same type of nodes stay connected after shuffling, we calculated the correlations between the degree of connected nodes as weighted average nearest-neighbors degrees (Barrat et al., 2004) in the original and shuffled networks.

Algorithm Description

The FFL-based generation algorithm starts with a nucleation step where an input network is used to find a subnetwork of predefined size (default 20 nodes) with all FFLs connected via shared nodes as in all analyzed networks, almost all FFL motifs share a common node with another FFL motif (Table 1). To avoid excessive parameters that could additionally control for in/out degree distribution, the *E. coli* GRN graph was used for the nucleation step. The degree distribution in the “FFL nucleus” sampled from a biological GRN was utilized by the preferential attachment rules as initial conditions to reconstruct a scale-free topology when attaching new edges and nodes to the growing network. The outline of the algorithm is presented graphically (Figure 3).

Once the substrate is selected the algorithm adds nodes and edges iteratively such that at every iteration, a candidate node is selected with a random uniform probability. Once selected, one of the four attachment rules (R1, R2, R3, and R4) is applied (Figure 4) based on four predetermined probabilities ($p_1, p_2, p_3,$ and p_4) that add up to 1. The iterations are repeated until the required number of nodes in the network is reached.

If the random float number $r1$ is less or equal to p_1 then R1 is picked. For the R1 rule we applied the modified preferential attachment algorithm from Abdelzaher et al. (2015a) with a power-law kernel:

$$P(g) = \frac{K_g^\gamma}{\sum_{i=1}^n K_i^\gamma} \quad (3)$$

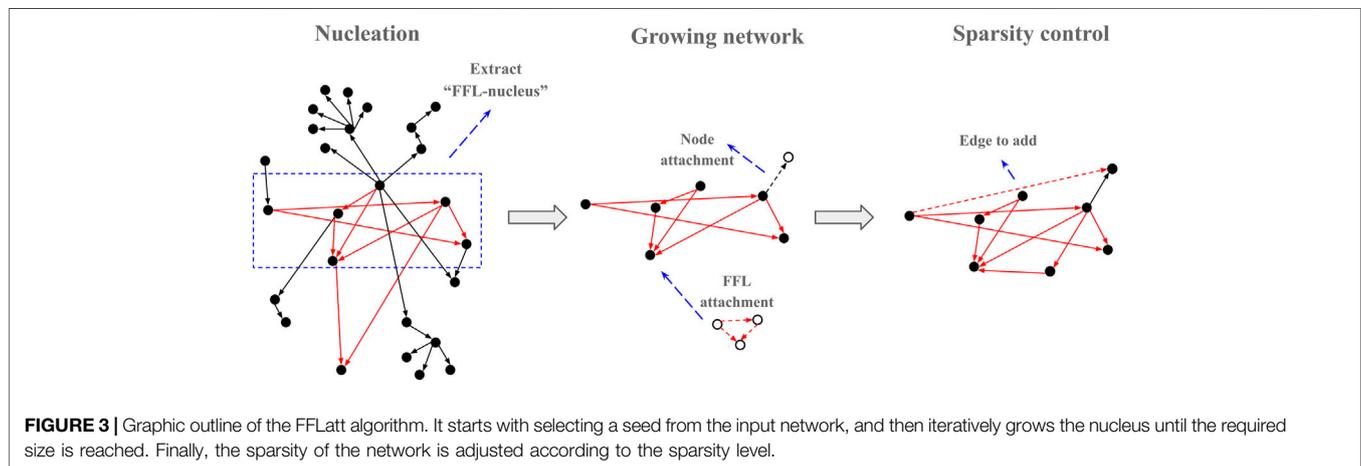
where K_i denotes node-degree connectivity, $P(g)$ is the probability that a new node will be connected to existing node g , and γ is a parameter that controls the shape of the out-degree distribution.

If $r1$ is greater than p_1 then one of the motif-based preferential attachment rules (R2, R3 or R4) is applied, and so $1-p_1$ corresponds to the desired percentage of nodes that participate in FFL motifs. For R2-R4 rules, one of the already existing FFL motifs is picked based on it’s connectivity with the others.

Once the candidate motif and rule are chosen, a new random float number, $r2$, is generated. If $0 < r2 \leq p_2$, the R2 rule is applied. In that case, two new edges and one new node will be added to the existing node so the new FFL motif is formed. If $r2 > p_2$, one of the R3 or R4 rules is selected with equal probability. For the R3 rule, two edges are added to nodes in existing FFL motifs to create a new FFL motif. For the R4 rule, and one edge is added between nodes in two existing FFL motifs to create a new FFL motif. If R2 is applied, it creates an FFL motif where one node has

TABLE 1 | Biological GRNs' graph properties.

Organism	# Of nodes	% Of nodes that participate in FFL motifs	% Of FFL motifs sharing nodes with other FFLs	Sparsity	In-degree	Out-degree
<i>E. coli</i>	1,917	37.4	99.1	2.328	1.106	1.222
<i>S. cerevisiae</i>	4,441	27.0	100	2.899	1.421	1.477
<i>M. musculus</i>	2,862	31.5	99.7	2.643	1.274	1.369
<i>H. sapiens</i>	2,456	34.7	99.9	2.944	1.364	1.580



only incoming edges. If R2 or R3 is applied, it creates an FFL motif where all participating nodes have at least one incoming and one outgoing edge. See **Figure 4** for details.

All nodes have to have an out-degree smaller or equal to a threshold K_{max} after which no new outgoing edges are added. If the candidate motif doesn't satisfy the conditions for a chosen FFL attachment rule, another candidate motif is picked and this is repeated until a motif is found that meets the rule conditions. If a new motif is created, the library with FFL motifs is updated.

When the desired network size is reached, the algorithm adjusts the sparsity (average number of connections per gene) until it reaches the set sparsity level in terms of average links per node. If the network is too dense, edges are selected for removal based on out-degree node connectivity so that an edge is proportionally more likely to be removed if it is attached to a node with a high out-degree. If the network is too sparse, edges are added to nodes selected proportionally to their out-degree connectivity, connecting them to randomly selected nodes. When network generation is completed, the network is saved as an unweighted directed graph.

Network Generation

For network simulation comparison five algorithms were chosen: FFLatt (developed in present study), GeneNetWeaver (GNW; Schaffter et al., 2011), NetworkX directed scale-free graph algorithm (NetworkX; Hagberg et al., 2008), and sparse

uniformly distributed random matrix with and without allowing for feedback loops in the network (DAG and RandG; Guo and Amir, 2021). DAG and RandG matrices were binarized by setting all non-zero elements equal to 1. The NetworkX graph algorithm was modified to control for sparsity as the FFLatt algorithm does, i.e., edges are added to or removed from nodes proportionally to their out-degree node connectivity. For network generation of different sizes with FFLatt, the set of transcriptional interaction graph properties estimated from the *E. coli* transcriptional interaction network (**Table 1**) was used. For each organism, the number of nodes that participate in FFL motif was used to set $p1$, with $p2$ equal to $(1-p1)*0.9$, and $p3=p4=(1-p1)*0.05$ respectively. For network generation of different sizes with other algorithms (except GNW), only network size and sparsity parameters were taken into account as only controllable parameters available. For network generation/subselection with GNW the following (default) parameters were used: `-random-seed`, `--greedy-selection`, `--keep-self-interactions` as well as the size of the subtracted network.

When mimicking the *E. coli* transcription network model, all three-node cycles were disrupted, by removal of one edge, as they are absent in the target network. The removal was done by deleting the outgoing edge of the node with the highest out-degree and an edge was instead attached to a random node with a probability based on the connectivity of each node.

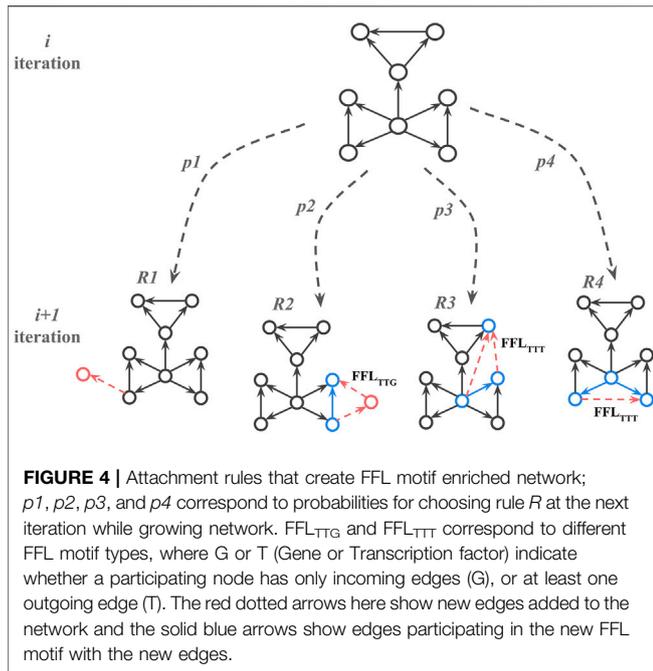


FIGURE 4 | Attachment rules that create FFL motif enriched network; $p1, p2, p3,$ and $p4$ correspond to probabilities for choosing rule R at the next iteration while growing network. FFL_{TTG} and FFL_{TTT} correspond to different FFL motif types, where G or T (Gene or Transcription factor) indicate whether a participating node has only incoming edges (G), or at least one outgoing edge (T). The red dotted arrows here show new edges added to the network and the solid blue arrows show edges participating in the new FFL motif with the new edges.

To mimic the complete three-node motif profile in biological GNRs in which non-FFL motifs are depleted, an optional motif depletion step can be executed. Here all three-node cycles are converted to FFL motifs by swapping the direction of one of the edges. In addition, up to one tenth of the cascades that do not share edges with FFL motifs were used to create new FFLs by adding an edge. The total number of edges that was used for motif conversion was taken into account when adjusting the network sparsity.

For stability analysis, self-loops (if any) were removed from network graphs generated with above mentioned algorithms before applying the stability analysis model.

Stability Analysis Model

To measure the stability of a network, i.e., how a network graph structure affects the dynamical stability of a gene regulatory interaction model, we utilized the model developed by (Guo and Amir, 2021) that explores how the dynamics of protein and mRNA concentrations control the transcriptional regulation. The model allows for multiple proteins acting on the same gene, and is defined by the authors as:

$$g_i(\vec{c}) = g_{i0} + \prod_j (1 + \gamma_{ij} f_{ij}(c_j)) \tag{4}$$

where g_i and g_{i0} is the effective gene copy number of gene i with and without input of other genes respectively, c_j is the concentration of transcription factor j , and γ_{ij} relates to the strength of the regulation of gene i by c_j . The functional relationship between the transcription factor and target gene, f_{ij} , is modelled as a sigmoid Hill function:

$$f_{ij}(c_j) = \frac{c_j^h}{K_{ij}^h + c_j^h} \tag{5}$$

where h is the saturation binding coefficient, i.e. the number of proteins required for saturation of binding to DNA, and K is the protein concentration threshold needed to produce a significant increase in mRNA.

The process of gene expression could be described as coupled dynamics of protein and mRNA concentrations. It was shown that in yeast (Zhurinsky et al., 2010) and mammalian cells (Schmidt and Schibler, 1995), the RNA polymerase concentration limits the transcription of mRNA, and the number of ribosomes limits the process of translation. The general transcription model (4) that connects transcription rate of gene i and the number of RNA polymerases can then be described as:

$$\frac{dC_{mi}}{dt} = k_m \phi_i(\vec{c}) n - C_{mi} k_p c_r - \frac{C_{mi}}{\tau} \tag{6}$$

$$\frac{dc_i}{dt} = k_p c_r \left(\frac{C_{mi}}{C_{mT}} - c_i \right) \tag{7}$$

where n is the total number of RNA polymerases, C_{mi} is the mRNA concentration of gene i , C_{mT} is the concentration of all mRNAs, ϕ is the gene allocation fraction of $g_i(\vec{c})$ controlled by RNA polymerases active on gene i , k_m is the transcription rate of RNA polymerase, k_p is the translation rate of the ribosome, c_r is the ribosomal concentration, and τ is the degradation rate difference between proteins and mRNA.

We assume that mRNAs degrade much faster than proteins, and as suggested by (Guo and Amir, 2021) we can set $\frac{dC_{mi}}{dt} \approx 0$ to neglect fast dynamics aiming to simplify the model. By substituting C_{mi} from 6 into 7, the dynamics of transcription factors concentrations can be simplified as:

$$\frac{dc_i}{dt} \approx k_p c_r \left(\phi_i(\vec{c}) - c_i \right) \tag{8}$$

In such case, the stability of a steady-state in the dynamical model is dependent on the Jacobian matrix A of size $N \times N$:

$$A = k_p c_r^{ss} (M - I) \tag{9}$$

where c_r^{ss} is the steady-state ribosomal concentration, M is the gene-gene interaction matrix that consists of γ_{ij} weights of the regulation, I is the identity matrix, and N is the number of genes in the system. The system is stable if the maximal real part of all eigenvalues of M , λ_M , is smaller than 1, i.e., the real part of all eigenvalues of A are negative. As the imaginary part of the eigenvalues is ignored, both oscillatory systems and systems without oscillations around the steady state are considered to be stable.

In contrast to random matrix theory (May 1972) or the generalized models (Gross and Feudel, 2006; Gross et al., 2010), the Jacobian matrix here is not a random matrix nor approximated through studying system bifurcations. In the Guo and Amir model it is derived by applying a knowledge-driven modelling approach which we find convenient for such a well-studied biological process like transcription. We applied this model to all network graphs simulated with different algorithms. Each graph, in a form of adjacency matrix, was supplied as a binary interaction matrix. For each replicate of a

different size generated with a given algorithm, we repeated assigning the network graph with link strengths 10 times. To focus on the effect of the GRN structure and FFL content on stability, we forced the distribution of link strengths of all GRNs to be similar. This was done by randomly setting half of the links in the binary interaction matrix to be upregulated and the other half downregulated (setting $\max(\gamma_{ij})$ and $\min(\gamma_{ij})$ to 1.5 and -1.5 respectively as boundaries of a normal distribution). In every trial, we first numerically solved for the ribosomal concentration c_r^{ss} with which the system reaches its non-zero steady state with Eq. 8. Given c_r^{ss} , A was found such that it only has negative real part eigenvalues using Eq. 9 by optimizing M , and the highest eigenvalue in λ_M from this solution was compared across networks of different sizes.

RESULTS

Feed-Forward Loop is the Only Enriched Three-Node Motif in Biological Gene Regulatory Networks

Of all possible 3-gene network motifs with 2 or 3 unidirectional links, we found a strong enrichment relative to shuffled networks of the FFL motif in the networks studied here, which are networks that mainly capture transcription factor to target interactions (Supplementary Table S1). This was previously shown for *E. coli* (Milo et al., 2002) and *S. cerevisiae* (Lee et al., 2002). We also found that the cascade, uplink, and downlink motifs were consistently and significantly (p -value < 0.05) depleted in all four target networks. To ensure that the shuffling procedure produced topologically similar networks, we verified that the distribution of correlations between the degree of connected nodes was similar for the original and shuffled networks (Supplementary Figure S1).

All depleted motifs are 3-node motifs with two edges (Figure 1), and these have previously been shown to be significantly depleted in other biological networks, for instance in a protein structure network and a human brain functional network (Mirzasoileiman and Jalili, 2011). However, how the depletion of these motifs contributes to the function of the gene circuitry, and how it relates to the evolution of gene regulatory networks, remains to be answered.

We found that FFL is the only enriched motif, and this was observed in all analyzed networks (Supplementary Table S1). Almost all FFL motifs share a common node with another FFL motif, as this fraction ranges from 99.1% in the *E. coli* GRN to 100% in *S. cerevisiae* (Table 1). The fraction of nodes that participate in FFL motifs ranges from 27 to 37.4%. This inspired us to develop a GRN generation algorithm that attaches nodes to form connected FFL motifs at a high rate. For each GRN we also calculated the average number of edges per node, here referred to as sparsity, and average in- and out-degrees, and these properties were also used as targets for the algorithm.

Each regulatory interaction in the FFL motif can be either positive or negative, i.e., activating or inhibiting, resulting in

eight different types that can act as e.g. accelerators, delay-generators or pulsers (Mangan and Alon, 2003), resulting in different dynamics of gene circuits. Given the wide variety of FFL types and their importance to GRN dynamics, an unsigned *in silico* GRN graph needs a large number of FFLs to accommodate these. A combination of the eight signed types of FFL motifs will in turn reflect a realistic flow of GRN circuits.

We generated a set of GRNs of different sizes from 500 to 1,500 nodes, 10 replicates for each size, using five different algorithms: FFLatt, GNW, NetworkX graph, RandG, and DAG. For each algorithm we analyzed four properties of their GRNs: the number of nodes that participate in FFL motifs, network sparsity, average in- and out-degree within the network. We repeated these simulations for all four organisms, as they have different graph properties. The results for *E. coli* are shown in Figure 5, and for the other organisms in Supplementary Figures S2, S3, and S4. Each organism-related GRN was used to set the topological parameters in the GRN simulated by FFLatt as described in Methods.

To assess the accuracy of GRN inference algorithms, the topological parameters such as in- and out-degree distribution and sparsity should be controlled when simulating data for benchmark analysis. We found that sparsity as well as out-degree of artificial networks generated with the subnetwork selection based GNW algorithm deviates considerably from the target networks for *E. coli* in sizes 500 and 750 (Figures 5B,D), for *S. cerevisiae* in size 500 (Supplementary Figures S2B, S2D), and in all sizes for *M. musculus* and *H. sapiens* (Supplementary Figures S3B, S3D, S4B, and S4D). While this alone does not indicate a poor performance of the GNW algorithm, it does advocate for the necessity of network generation algorithms to control topological parameters.

More importantly, when subsetting networks from biological GRNs with the GNW algorithm, we obtained a significant underrepresentation of FFL motifs in sizes 500, 750, and 1,000 for *E. coli* (Figure 5A) in comparison with FFLatt networks. Similar results were obtained for GRNs of other organisms (Supplementary Figures S2A, S3A, and S4A). To confirm and extend these findings, we performed motif enrichment analysis on the simulated networks as well as on biological GRNs (Figure 6; Supplementary Table S1). This showed that FFL motifs are not significantly overrepresented in GNW networks, but they are highly significantly enriched in the *E. coli* GRN (Z-score 7.4). In networks generated with other algorithms, the FFL motif was also not significantly overrepresented, with the exception of FFLatt whose networks were significantly enriched with Z-scores between 2.95 and 4.98. By default, FFLatt does not deplete other 3-node motifs, and but this is possible with an optional motif depletion step. We explored how this step in combination with various parameter values can mimic the complete 3-node motif distribution profile with the FFL motif enriched, and all other motifs depleted (Supplementary Table S2).

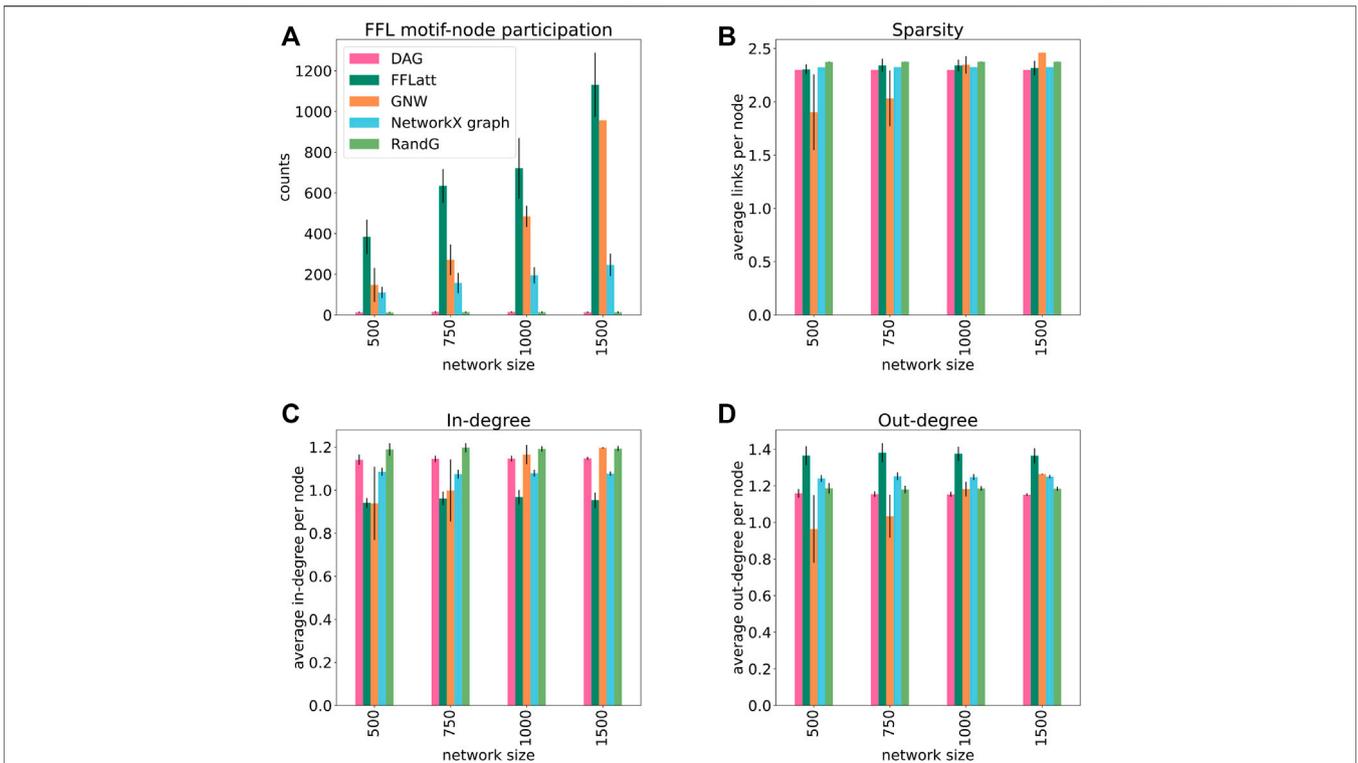


FIGURE 5 | Topological properties of simulated networks (*E. coli*). FFL motif node participation, average sparsity, in- and out-degree distribution in simulated networks. For FFL-motif node participation counts, up to three participations for each node were allowed (in different roles). Each data point was calculated as the average of ten different replicates of each network size. Error bars represent standard deviation.

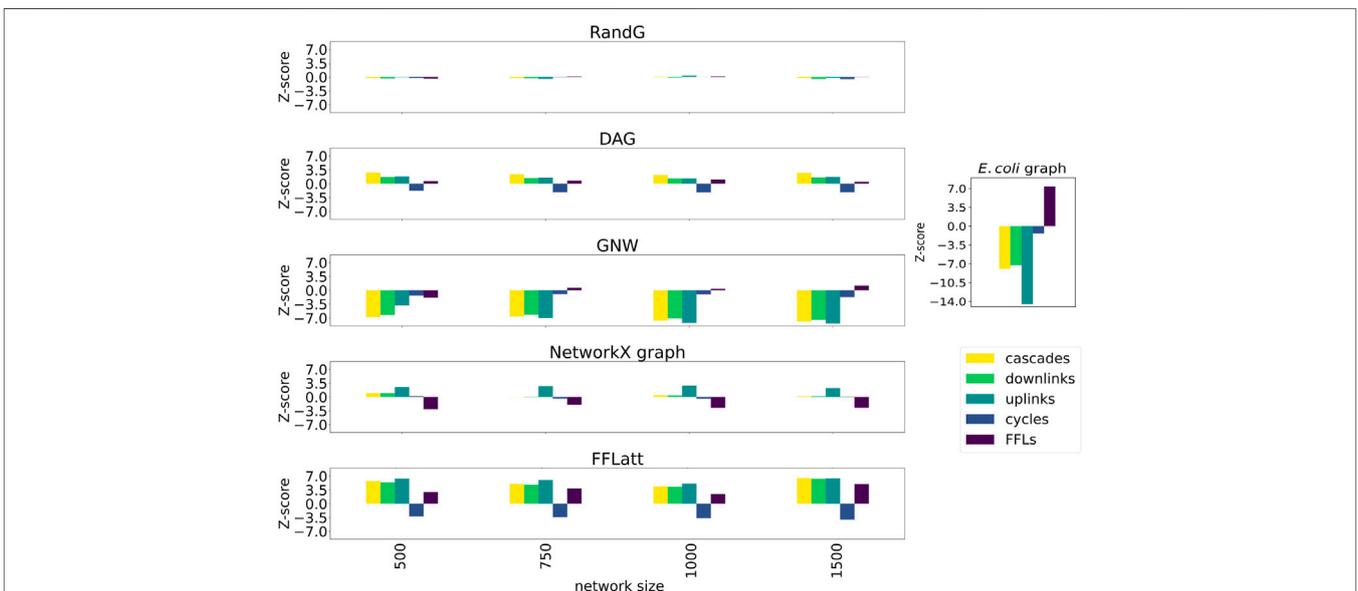
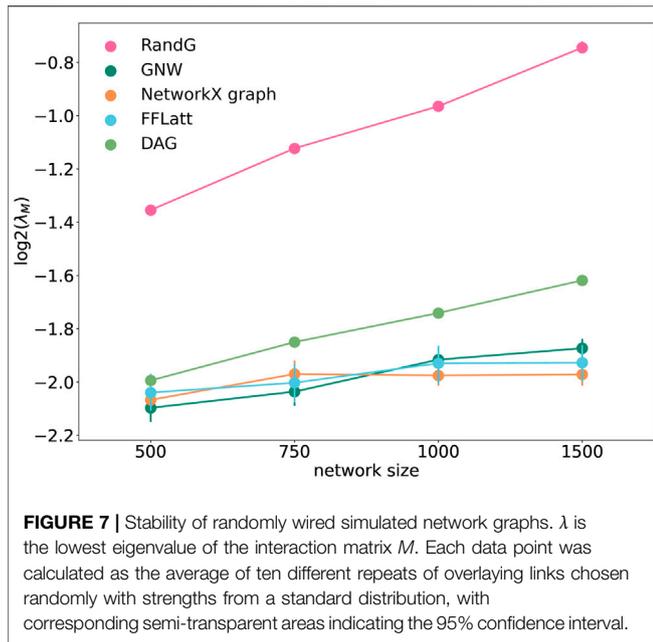


FIGURE 6 | Motif enrichment analysis of 3-node network motifs in simulated networks (*E. coli*). For networks generated with GNW, the *E. coli* RegulonDB (Santos-Zavaleta et al., 2019) database was used. For networks generated with FFLatt, we used the graph properties for *E. coli* specified in **Table 1**. RandG is a random assignment of links and DAG is the same with cycles removed. NetworkX graph GRNs are scale-free. For RandG, DAG, and NetworkX graph GRNs we used the *E. coli* network sparsity.



Topology, Motif Composition, and Network Stability

In biology, random matrix theory, that seeks to understand the properties of matrices with randomly drawn elements, is known from R. May’s research on the stability of large biological systems (May 1972). He demonstrated that the stability of a large ecological system depends on satisfying the following inequality:

$$1 > \alpha \sqrt{nC} \tag{10}$$

where α is the average interaction strength, n is the number of species, and C is the density of interactions between them. Therefore, the larger a system gets the more unstable it becomes unless the sparsity and/or interaction strengths are scaled down accordingly. May’s approach has been proven to be highly valuable to other biological networks (Aljadeff et al., 2015), including those that aim to describe gene regulations (Prill et al., 2005; Stone, 2018).

It was earlier suggested that motif composition contributes to fault-tolerance in transcriptional networks (Roy et al., 2020). To test if the structural composition is important for stability in artificially generated networks, we analysed the stability of the five network models using the method by Guo and Amir (2021). As expected, all GRNs with fixed sparsity and interaction strengths became more fragile when increasing in size. We found that GRNs with different motif profiles demonstrated different levels of network stability (Figure 7). The RandG GRNs that were neither enriched nor depleted with any 3-node motifs (Figure 6) were far less stable than the other ones. The DAG GRNs which are generated like RandG GRNs but without cyclic motifs were more stable but still considerably less stable than NetworkX, GNW, and FFLatt GRNs. We note that NetworkX, GNW, and FFLatt GRNs have different network motif abundances, such as either depleted or enriched FFL motifs, and yet they show similar

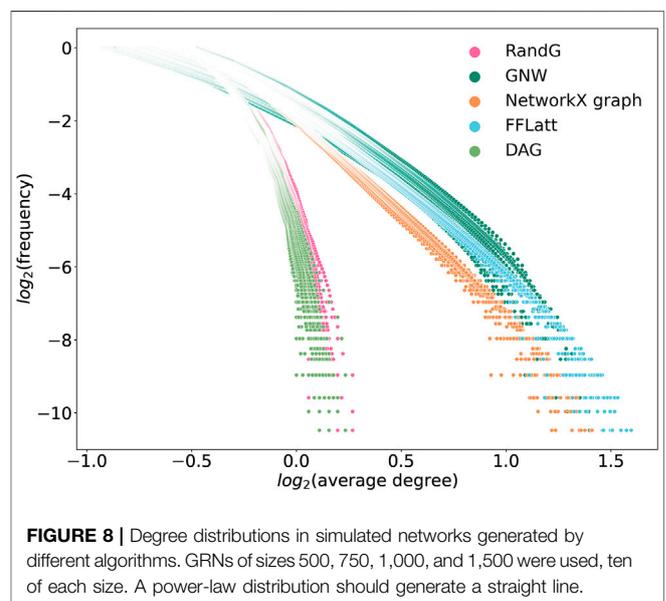
stability. The abundance of the FFL motif alone therefore does not seem to be a major factor for network stability, which is congruent with previous findings about non-importance of the FFL motif to system robustness under random node failure test (Abdelzaher et al., 2015b).

We note that the two lines that represent size-dependent stability of DAG and RandG GRNs have a steeper slope than the other three. This means that as the GRN increases in size, DAG and RandG GRNs become less stable faster than the other three. To find a reason for this, we analyzed the degree distribution of the GRNs. Since RandG and DAG networks are sparse uniformly distributed random binary matrices, their degree distributions do not follow the power-law and therefore they are not scale-free (Figure 8). This suggests that a scale-free topology which has been previously found to be central for creating a robust system, protecting the GRN from random mutations (Greenbury et al., 2010), can in fact help gene regulatory systems to reach a stable state after perturbation.

DISCUSSION

Here we present a new algorithm, FFLatt, for generating realistic directed GRN graphs to enable more accurate and authentic performance evaluation of GRN inference methods. The novelty of the presented algorithm is that it generates networks with boosted FFL motifs, which are known to be important for network dynamics. Besides being enriched with the FFL motif, the resulting GRN graphs generated with FFLatt exhibit topological properties similar to experimentally validated biological GRNs.

We show that the motif profile and topological properties of FFLatt network graphs demonstrate a biological stability comparable with other models, such as the NetworkX and GNW algorithms. It is particularly important for network



inference methods working with steady-state gene expression data as many of them, for instance Least-Squares with Cut-Off (LSCO; (Tjärnberg et al., 2013), LASSO (Tibshirani, 1996; Friedman et al., 2010), LASSO-VAR (Larvie et al., 2016), and GENIE3 (Huynh-Thu et al., 2010) aim to infer a stable static network from steady-state data. To summarize, the FFLatt graph generation algorithm provides an opportunity to simulate biologically meaningful network graphs that can be wired with realistic biological dynamics.

We also noted that the FFLatt networks were enriched with three other motifs: uplinks, downlinks and cascades whereas in GNW networks and biological GRNs these motifs are usually depleted. Sorrells and Jonhson (2015) suggested that in biological GRNs, FFL formation proceeds through a non-adaptive rewiring of gene regulatory regulation which could explain how the abundance of FFLs and the depletion of uplinks, downlinks, and cascades is coupled. The algorithm can be run to allow for depletion of other 3-node motifs while growing the network. However a reason that such depletions are important for network dynamics is yet to be found. A thorough search of the relevant literature did not yield in related articles. We also could not find evidence that different three-node motif profiles affect network stability. NetworkX, GNW, and FFLatt motif profiles are fairly different yet they demonstrated comparable stability across different sizes. While being out of scope for this study, it remains an interesting question how the composition of more complex and higher-order structures known to be present in GRNs (Benson et al., 2016; Goroehowski et al., 2018) could contribute to stability of the system.

In this article we focus on the proof of concept of the FFL attachment algorithm to demonstrate its necessity and feasibility. However, to increase model performance, it could be extended with other parameters. For example, to better capture “small world” (Watts and Strogatz, 1998) structural properties that are known to be present in biological networks, one parameter could be a desired number of biological modules so that within each module the connectivity is higher than in between them. The clustering algorithm should however be biologically motivated so that the connection between modular graph structure and expression dynamics is clear.

Despite a continued uncertainty of how structural properties and functional modularity of GRNs relate to each other, some patterns such as FFLs are known to be key signatures of transcriptional regulation networks. Here we developed a novel algorithm that generates biologically realistic structures of large artificial gene regulatory networks with controlled size, sparsity, topology, and number of FFLs. The implementation executes with reasonable runtimes (**Supplementary Figure S5**). FFLatt graphs are binary and can thus assume a wide range of dynamical structures with signed strengths. They could be used as

input to already established tools based on Hill function kinetics such as GNW, which allows for knock-out and knock-down perturbation designs when generating expression data, and some control of the number of nodes, including the number of transcription factors, based on a user-defined input network. To generate expression data it utilizes a non-linear ordinary differential equations (ODE) model for gene expression, and stochastic differential equations (SDEs) for molecular noise generation. Potentially, they could also become a part of future deep learning frameworks that aim to model gene expression from DNA sequence (Zrimec et al., 2020; Avsec et al., 2021). In such frameworks, FFLatt networks could be used as a deep learning model constraint to incorporate prior knowledge of each node participation in FFL motifs. As a result, we believe that it will contribute to future development of benchmarking tools that could fairly and accurately evaluate the performance of GRN inference methods.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. The source code of the algorithm is available at <https://bitbucket.org/sonnhammergrni/flatt>.

AUTHOR CONTRIBUTIONS

EZ and OV devised and implemented the algorithm. EZ and TH performed the calculations, analyzed the results, contributed to the discussion, designed the figures, and wrote the manuscript. ES participated in the design and coordination of the study, contributed to the discussion and design of figures, supervised and reviewed the writing of the manuscript. All authors read and approved the final version of the manuscript.

FUNDING

This work was supported by the Swedish Research Council Project 2019-04095.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.815692/full#supplementary-material>

REFERENCES

- Abdelzaher, A. F., Al-Musawi, A. F., Ghosh, P., Mayo, M. L., and Perkins, E. J. (2015b). Transcriptional Network Growing Models Using Motif-Based Preferential Attachment. *Front. Bioeng. Biotechnol.* 3, 157. doi:10.3389/fbioe.2015.00157
- Abdelzaher, A. F., Mayo, M. L., Perkins, E. J., and Ghosh, P. (2015a). Contribution of Canonical Feed-Forward Loop Motifs on the Fault-Tolerance and Information Transport Efficiency of Transcriptional Regulatory Networks. *Nano Commun. Networks* 6, 133–144. doi:10.1016/j.nancom.2015.04.002
- Ahnert, S. E., and Fink, T. M. A. (2016). Form and Function in Gene Regulatory Networks: The Structure of Network Motifs Determines Fundamental

- Properties of Their Dynamical State Space. *J. R. Soc. Interf.* 13, 20160179. doi:10.1098/rsif.2016.0179
- Aljadeff, J., Stern, M., and Sharpee, T. (2015). Transition to Chaos in Random Networks with Cell-type-Specific Connectivity. *Phys. Rev. Lett.* 114, 088101. doi:10.1103/PhysRevLett.114.088101
- Avsec, Ž., Agarwal, V., Visentin, D., Ledsam, J. R., Grabska-Barwinska, A., Taylor, K. R., et al. (2021). Effective Gene Expression Prediction from Sequence by Integrating Long-Range Interactions. *Nat. Methods* 18, 1196–1203. doi:10.1038/s41592-021-01252-x
- Balaji, S., Babu, M. M., Iyer, L. M., Luscombe, N. M., and Aravind, L. (2006). Comprehensive Analysis of Combinatorial Regulation Using the Transcriptional Regulatory Network of Yeast. *J. Mol. Biol.* 360, 213–227. doi:10.1016/j.jmb.2006.04.029
- Barabási, A., and Albert, R. (1999). Emergence of Scaling in Random Networks. *Science* 286, 509–512. doi:10.1126/science.286.5439.509
- Barrat, A., Barthelemy, M., Pastor-Satorras, R., and Vespignani, A. (2004). The Architecture of Complex Weighted Networks. *Proc. Natl. Acad. Sci.* 101, 3747–3752. doi:10.1073/pnas.0400087101
- Benson, A. R., Gleich, D. F., and Leskovec, J. (2016). Higher-Order Organization of Complex Networks. *Science* 353, 163–166. doi:10.1126/science.aad9029
- Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., et al. (2005). Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. *Cell* 122, 947–956. doi:10.1016/j.cell.2005.08.020
- Chen, S., and Mar, J. C. (2018). Evaluating Methods of Inferring Gene Regulatory Networks Highlights Their Lack of Performance for Single Cell Gene Expression Data. *BMC Bioinformatics* 19, 232. doi:10.1186/s12859-018-2217-z
- Chouvardas, P., Kollias, G., and Nikolaou, C. (2016). Inferring Active Regulatory Networks from Gene Expression Data Using a Combination of Prior Knowledge and Enrichment Analysis. *BMC Bioinformatics* 17, 181. doi:10.1186/s12859-016-1040-7
- Davidson, E. H. (2010). Emerging Properties of Animal Gene Regulatory Networks. *Nature* 468, 911–920. doi:10.1038/nature09645
- Emmert-Streib, F., and Dehmer, M. (2018). Inference of Genome-Scale Gene Regulatory Networks: Are There Differences in Biological and Clinical Validations? *Make* 1, 138–148. doi:10.3390/make1010008
- Faith, J. J., Hayete, B., Thaden, J. T., Mogno, I., Wierzbowski, J., Cottarel, G., et al. (2007). Large-Scale Mapping and Validation of *Escherichia coli* Transcriptional Regulation from a Compendium of Expression Profiles. *Plos Biol.* 5, e8. doi:10.1371/journal.pbio.0050008
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Soft.* 33, 1–22. doi:10.18637/jss.v033.i01
- Gardner, T., and Faith, J. (2005). Reverse-engineering Transcription Control Networks. *Phys. Life Rev.* 2, 65–88. doi:10.1016/j.plrev.2005.01.001
- Gorochowski, T. E., Grierson, C. S., and di Bernardo, M. (2018). Organization of Feed-Forward Loop Motifs Reveals Architectural Principles in Natural and Engineered Networks. *Sci. Adv.* 4, 12. doi:10.1126/sciadv.aap9751
- Greenbury, S. F., Johnston, I. G., Smith, M. A., Doye, J. P. K., and Louis, A. A. (2010). The Effect of Scale-Free Topology on the Robustness and Evolvability of Genetic Regulatory Networks. *J. Theor. Biol.* 267, 48–61. doi:10.1016/j.jtbi.2010.08.006
- Gross, T., and Feudel, U. (2006). Generalized Models as a Universal Approach to the Analysis of Nonlinear Dynamical Systems. *Phys. Rev. E Stat. Nonlin Soft Matter Phys.* 73, 016205. doi:10.1103/PhysRevE.73.016205
- Gross, T., Stiefs, D., Rudolf, L., and Zumsande, M. (2010). Generalized Modeling of Heterogeneous Nonlinear Networks. *IEICE Proc. Ser.* 44, A2L–A1. doi:10.3485/proc.44.A2L-A1
- Guo, Y., and Amir, A. (2021). Exploring the Effect of Network Topology, Mrna and Protein Dynamics on Gene Regulatory Network Stability. *Nat. Commun.* 12, 130. doi:10.1038/s41467-020-20472-x
- Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). “Exploring Network Structure, Dynamics, and Function Using NetworkX,” in *Proceedings of 7th Python in Science Conference (SciPy2008)*. Editors G. Varoquaux, T. Vaught, and J. Millman, 11–15.
- Han, H., Cho, J.-W., Lee, S., Yun, A., Kim, H., Bae, D., et al. (2018). TRRUST V2: An Expanded Reference Database of Human and Mouse Transcriptional Regulatory Interactions. *Nucleic Acids Res.* 46, D380–D386. doi:10.1093/nar/gkx1013
- Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. *PLoS ONE* 5, e12776. doi:10.1371/journal.pone.0012776
- Iglesias-Martinez, L. F., De Kegel, B., and Kolch, W. (2021). KBoost: A New Method to Infer Gene Regulatory Networks from Gene Expression Data. *Sci. Rep.* 11, 15461. doi:10.1038/s41598-021-94919-6
- Kang, Y., Liow, H.-H., Maier, E. J., and Brent, M. R. (2018). NetProphet 2.0: Mapping Transcription Factor Networks by Exploiting Scalable Data Resources. *Bioinformatics* 34, 249–257. doi:10.1093/bioinformatics/btx563
- Kaplan, S., Bren, A., Dekel, E., and Alon, U. (2008). The Incoherent Feed-forward Loop Can Generate Non-monotonic Input Functions for Genes. *Mol. Syst. Biol.* 4, 203. doi:10.1038/msb.2008.43
- Krek, A., Grün, D., Poy, M. N., Wolf, R., Rosenberg, L., Epstein, E. J., et al. (2005). Combinatorial microRNA Target Predictions. *Nat. Genet.* 37, 495–500. doi:10.1038/ng1536
- Larvie, J., Sefidmazgi, M., Homaifar, A., Harrison, S., Karimodini, A., and Guiseppi-Elie, A. (2016). Stable Gene Regulatory Network Modeling from Steady-State Data. *Bioengineering* 3, 12. doi:10.3390/bioengineering3020012
- Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., et al. (2002). Transcriptional Regulatory Networks in *Saccharomyces cerevisiae*. *Science* 298, 799–804. doi:10.1126/science.1075090
- Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes Are microRNA Targets. *Cell* 120, 15–20. doi:10.1016/j.cell.2004.12.035
- Mangan, S., and Alon, U. (2003). Structure and Function of the Feed-Forward Loop Network Motif. *Proc. Natl. Acad. Sci.* 100, 11980–11985. doi:10.1073/pnas.2133841100
- Mangan, S., Zaslavler, A., and Alon, U. (2003). The Coherent Feedforward Loop Serves as a Sign-Sensitive Delay Element in Transcription Networks. *J. Mol. Biol.* 334, 197–204. doi:10.1016/j.jmb.2003.09.049
- Marbach, D., Costello, J. C., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J., et al. (2012). Wisdom of Crowds for Robust Gene Network Inference. *Nat. Methods* 9, 796–804. doi:10.1038/nmeth.2016
- Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Favera, R. D., et al. (2006). ARACHNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics* 7, S7. doi:10.1186/1471-2105-7-s1-s7
- May, R. M. (1972). Will a Large Complex System Be Stable? *Nature* 238, 413–414. doi:10.1038/238413a0
- Mendes, P., Sha, W., and Ye, K. (2003). Artificial Gene Networks for Objective Comparison of Analysis Algorithms. *Bioinformatics* 19, ii122–ii129. doi:10.1093/bioinformatics/btg1069
- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network Motifs: Simple Building Blocks of Complex Networks. *Science* 298, 824–827. doi:10.1126/science.298.5594.824
- Mirzasoileman, B., and Jalili, M. (2011). Failure Tolerance of Motif Structure in Biological Networks. *PLoS ONE* 6, e20512. doi:10.1371/journal.pone.0020512
- Pratap, A., Jaliha, A. P., Law, J. N., Bharadwaj, A., and Murali, T. M. (2020). Benchmarking Algorithms for Gene Regulatory Network Inference from Single-Cell Transcriptomic Data. *Nat. Methods* 17, 147–154. doi:10.1038/s41592-019-0690-6
- Prill, R. J., Iglesias, P. A., and Levchenko, A. (2005). Dynamic Properties of Network Motifs Contribute to Biological Network Organization. *Plos Biol.* 3, e343. doi:10.1371/journal.pbio.0030343
- Roy, S., Ghosh, P., Barua, D., and Das, S. K. (2020). Motifs Enable Communication Efficiency and Fault-Tolerance in Transcriptional Networks. *Sci. Rep.* 10, 9628. doi:10.1038/s41598-020-66573-x
- Santos-Zavaleta, A., Salgado, H., Gama-Castro, S., Sánchez-Pérez, M., Gómez-Romero, L., Ledezma-Tejeida, D., et al. (2019). RegulonDB V 10.5: Tackling Challenges to Unify Classic and High Throughput Knowledge of Gene Regulation in *E. coli* K-12. *Nucleic Acids Res.* 47, D212–D220. doi:10.1093/nar/gky1077
- Schaffter, T., Marbach, D., and Floreano, D. (2011). GeneNetWeaver: in silico Benchmark Generation and Performance Profiling of Network Inference Methods. *Bioinformatics* 27, 2263–2270. doi:10.1093/bioinformatics/btr373
- Schmidt, E. E., and Schibler, U. (1995). Cell Size Regulation, a Mechanism that Controls Cellular RNA Accumulation: Consequences on Regulation of the

- Ubiquitous Transcription Factors Oct1 and NF-Y and the Liver-Enriched Transcription Factor DBP. *J. Cel Biol.* 128, 467–483. doi:10.1083/jcb.128.4.467
- Shalgi, R., Lieber, D., Oren, M., and Pilpel, Y. (2007). Global and Local Architecture of the Mammalian microRNA-Transcription Factor Regulatory Network. *Plos Comput. Biol.* 3, e131. doi:10.1371/journal.pcbi.0030131
- Shen-Orr, S. S., Milo, R., Mangan, S., and Alon, U. (2002). Network Motifs in the Transcriptional Regulation Network of *Escherichia coli*. *Nat. Genet.* 31, 64–68. doi:10.1038/ng881
- Siahpirani, A. F., and Roy, S. (2017). A Prior-Based Integrative Framework for Functional Transcriptional Regulatory Network Inference. *Nucleic Acids Res.* 45, gkw963. doi:10.1093/nar/gkw963
- Sohka, T., Heins, R. A., Phelan, R. M., Greisler, J. M., Townsend, C. A., and Ostermeier, M. (2009). An Externally Tunable Bacterial Band-Pass Filter. *Proc. Natl. Acad. Sci.* 106, 10135–10140. doi:10.1073/pnas.0901246106
- Sorrells, T. R., and Johnson, A. D. (2015). Making Sense of Transcription Networks. *Cell* 161, 714–723. doi:10.1016/j.cell.2015.04.014
- Stone, L. (2018). The Feasibility and Stability of Large Complex Biological Networks: A Random Matrix Approach. *Sci. Rep.* 8, 8246. doi:10.1038/s41598-018-26486-2
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B (Methodological)* 58, 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x
- Tjärnberg, A., Nordling, T. E. M., Studham, M., and Sonnhammer, E. L. L. (2013). Optimal Sparsity Criteria for Network Inference. *J. Comput. Biol.* 20, 398–408. doi:10.1089/cmb.2012.0268
- Tsang, J., Zhu, J., and van Oudenaarden, A. (2007). MicroRNA-Mediated Feedback and Feedforward Loops Are Recurrent Network Motifs in Mammals. *Mol. Cel* 26, 753–767. doi:10.1016/j.molcel.2007.05.018
- Van den Bulcke, T., Van Leemput, K., Naudts, B., van Remortel, P., Ma, H., Verschoren, A., et al. (2006). Syntren: A Generator of Synthetic Gene Expression Data for Design and Analysis of Structure Learning Algorithms. *BMC Bioinformatics* 7, 43. doi:10.1186/1471-2105-7-43
- Watts, D. J., and Strogatz, S. H. (1998). Collective Dynamics of 'small-world' Networks. *Nature* 393, 440–442. doi:10.1038/30918
- Zavlanos, M. M., Julius, A. A., Boyd, S. P., and Pappas, G. J. (2011). Inferring Stable Genetic Networks from Steady-State Data. *Automatica* 47, 1113–1122. doi:10.1016/j.automatica.2011.02.006
- Zhang, C., Tsoi, R., Wu, F., and You, L. (2016). Processing Oscillatory Signals by Incoherent Feedforward Loops. *Plos Comput. Biol.* 12, e1005101. doi:10.1371/journal.pcbi.1005101
- Zhurinsky, J., Leonhard, K., Watt, S., Marguerat, S., Bähler, J., and Nurse, P. (2010). A Coordinated Global Control over Cellular Transcription. *Curr. Biol.* 20, 2010–2015. doi:10.1016/j.cub.2010.10.002
- Zrimec, J., Börlin, C. S., Buric, F., Muhammad, A. S., Chen, R., Siewers, V., et al. (2020). Deep Learning Suggests that Gene Expression Is Encoded in All Parts of a Co-Evolving Interacting Gene Regulatory Structure. *Nat. Commun.* 11, 6141. doi:10.1038/s41467-020-19921-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhivkopoulos, Vavulov, Hillerton and Sonnhammer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.