



HAL
open science

Gene regulatory network.

Marc-Thorsten C Hütt, Annick Lesne

► **To cite this version:**

Marc-Thorsten C Hütt, Annick Lesne. Gene regulatory network.: Dissecting structure and dynamics. O. Wolkenhauer. Systems Medicine: Integrative, Qualitative and Computational Approaches, Elsevier, pp.77-85, 2020, Reference Modules in Biomedical Sciences (peer-reviewed, accepted 20/06/2019). hal-02313633

HAL Id: hal-02313633

<https://hal.science/hal-02313633>

Submitted on 11 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Gene regulatory network inference I

Dissecting structure and dynamics

Marc-Thorsten Hütt^{a,*}, Annick Lesne^{b,c}

^a*Department of Life Sciences and Chemistry, Jacobs University Bremen, D-28759 Bremen, Germany*

^b*Sorbonne Université, CNRS, Laboratoire de Physique Théorique de la Matière Condensée, LPTMC, F-75252, Paris, France*

^c*Institut de Génétique Moléculaire de Montpellier, University of Montpellier, CNRS, F-34293, Montpellier, France*

Abstract

Gene network inference is the task of reconstructing regulatory networks among genes from high-throughput (in particular transcriptomic) data. Here we introduce the main concepts of this rich and rapidly evolving field. In order to illustrate the basic principles of gene network inference we simulate gene expression patterns using two distinct computational models, Boolean dynamics and ordinary differential equations. These synthetic data are then analyzed with basic gene network inference methods based on information theory and correlation analysis. We emphasize that a careful distinction between the underlying network architecture and the effective network inferred from the dynamical patterns, similar to the interplay of structural and functional connectivity often discussed in Computational Neuroscience, may open up a new perspective on the data.

Keywords:

Boolean dynamics, ChIP–chip data, ENCODE project, Functional connectivity, Gene expression, GeneNetWeaver software, Information theory, Regulon DB, Simulation, Statistical inference, Structural connectivity, Transcriptional regulatory network

*Corresponding author

Email address: m.huett@jacobs-university.de (Marc-Thorsten Hütt)

1. Several notions of gene network

Networks are abstractions of real-life systems in terms of graphs consisting of nodes and links. They have a prominent role in Systems Biology and Systems Medicine, as they allow the integration of diverse data in a common framework. Furthermore, graph theory offers a rich toolbox for analyzing network structure and its implications for the dynamics, and hence the function, of such networks.

Network inference is the challenging task of estimating the underlying network from dynamical observations at each node of a system. It obviously depends on the meaning of a link between two nodes. This task has been explored with diverse methods across many disciplines, ranging from computer science [13] and bioinformatics [26] to physics [36], information theory [63] and statistics [58]. Machine learning has an increasingly important role in this issue [52, 14, 3].

Several notions of gene network are coexisting. The most prominent notion is the *transcriptional regulatory network*, where nodes are genes and a (directed) link indicates that the first gene encodes a transcription factor having a binding site in the regulatory region of the second gene. Evidence for the binding site of the transcription factor can either be provided by bioinformatic prediction (genomic motif analysis) or by experiments. The principal technique for experimentally assessing transcriptional regulatory interactions are Chromatin immunoprecipitation microarrays (ChIP-chip). Often, transcription regulation is represented as a bipartite graph comprising transcription factor nodes and gene nodes, with directed links from the encoding gene to the transcription factor and from the transcription factor to the regulated gene(s). Then the transcriptional regulatory network introduced above is the gene-centric projection of this bipartite graph.

Regulation via transcription factors is only one of many biological aspects of gene regulation. Other possible points of action of gene regulatory mechanisms include chromosomal organization and chromatin state, modification of the RNA transcript, mRNA transport or degradation and, when going further to the level of proteins, translation, protein degradation or activation and deacti-

vation. For instance, chromatin conformation is involved either through binding sites accessibility, or through mechanical constraints modifying binding affinity of transcription factors [46]. As such, a transcriptional regulatory network does not capture the whole complexity of gene regulation, but only the component based on transcription factor targeted action.

In contrast to a transcriptional regulatory network, a *genetic interaction network* rather addresses the functional relationships between genes. It is based on the empirical evaluation of phenotypes, comparing single-gene knockout mutants and two-gene knockout mutants (double mutants). A link in a genetic interaction network indicates that the phenotype of the double mutant differs from that of each individual mutant (e.g., the double mutation is lethal, while both single-gene mutants have a viable phenotype). This type of networks is not discussed here [see 11, for a detailed review].

Other types of gene networks frequently employed are *co-expression networks*, where a link between two genes means that they are jointly expressed, and *co-regulation networks*, where a link between two genes means that they are regulated by a common transcription factor. It is today acknowledged that in metazoan species, topological domains delineate sets of genes that are co-regulated, and separate genes and regulatory sequences that should not interfere with their regulation [50]. Gene co-regulation networks are thus essentially shaped by 3D genome organization. Even in prokaryotic gene regulation, the regulatory influence of chromosomal organization is well documented [74, 75] (see also Section 6). Another notion of gene network is derived from *protein-protein interaction networks*, by considering that a link between two genes means that the encoded proteins physically interact.

A shortcoming in these various definitions is their limitation to pairwise relationships, while the expression of more than two genes could intermingle in complex crosstalks. Hyper-networks could in principle be used to capture n -ary relationships [39, 61] but their mathematical intricacy makes them far less operational than plain networks.

Henceforth, we focus on transcriptional regulatory networks.

2. Some landmarks on gene regulatory networks

The concept of gene regulation and first drawings of gene regulatory networks date back to the work of Jacob and Monod in 1961 [38, 57]. They proposed a model in which a metabolite induces or represses the expression of a regulator gene, in turn controlling the expression of an array of co-expressed structural genes (forming an operon). Today, the notion of structural and regulatory genes is no longer in use, replaced by those of regulatory sequences and transcriptional regulatory networks.

Due to the effort over many years and the careful manual curation of the underlying database, the most comprehensive transcriptional regulatory network currently available in electronic form is the network stored in RegulonDB [30] for the gut bacterium *Escherichia coli*. Figure 1 shows a network representation of the main segment of RegulonDB. Here, various pieces of computational and empirical information are integrated, leading to a list of edges with controlled levels of reliability.

For the yeast *Saccharomyces cerevisiae*, as an example, early compilations of the transcriptional regulatory network, are provided in Guelzim et al. [34] and Lee et al. [45]. However more recent information is still accumulating, and needs to be compiled from individual publications [e.g., 33, 6]. A review of the current status of its transcriptional regulatory network is provided in Liu et al. [49]. A recently developed database for edges derived from ChIP–chip experiments is RegulatorDB [16].

With such a substantial volume of information available about gene regulatory networks, we can think of these networks as the 'hardware', on which the dynamical processes of gene activation and deactivation, as well as transcription and translation run. Gene expression profiles (or transcriptomic data) can be regarded as snapshots of these dynamics running on the underlying network. We will see in Sections 5 and 6 that it is indeed a productive approach to discriminate the structural information (i.e., the network architecture) and the dynamical information (i.e., the gene expression pattern).

When going to multicellular organisms, data availability for gene regulatory networks deteriorates in two ways: (1) Data (for example for gene regulation via transcription factors) becomes less and less complete. (2) Other processes of gene regulation (beyond transcription factors) become more and more important.

Insights on gene regulation during the development of multicellular organisms came from investigations taking as a model the *Drosophila* embryo [60]. Here a set of genes, the gap genes, control via the concentrations of their products the expression of developmental pair-rule genes [64]. The complexity of the developmental embryonic regulatory networks has been best demonstrated by the work of Davidson on sea urchin [22, 47].

In Human, the ENCODE (Encyclopedia of DNA Elements) project [19] undertook among other tasks that of mapping the gene regulatory network [31].

Another direction, far different from high-throughput experimental investigations, is to study the design of gene regulation through the construction of synthetic networks with a few genes [35] or the analysis of functional network motifs [68, 1]. These approaches provide insight in the logic and the main building blocks of gene regulatory networks.

As for the dynamics of gene regulation, beyond ordinary differential equations an acknowledged paradigm has been established with the development of random Boolean networks [42, 10]. In the following section we will make use of these two approaches to generate sample data for illustrating simple principles of network inference.

3. An illustrative example of gene regulatory network inference

In situations, where reliable and comprehensive data on the network structure are not available, the distinction between the network architecture derived from structural information and the dynamical patterns obtained from gene expression is not possible, and the task of *network inference* becomes an important approach.

Here we illustrate the key ideas of network inference in their simplest form using as a benchmark the small (fictitious) transcriptional regulatory network shown in Figure 2A and synthetic gene expression data.

For this purpose of illustrating some of the challenges associated with network inference, we simulated *stylized but realistic* gene expression profiles. We then inferred a gene regulatory network from these data, in order to subsequently compare the inferred with the original regulatory network. We used two fundamentally different models of regulatory dynamics: (1) Boolean dynamics, where nodes can be on or off, (2) stochastic ordinary differential equations, as represented by the GeneNetWeaver software [67], where node states take continuous values representing the amount of transcripts.

The network shown in Figure 2 consists of 20 nodes with 30 directed links, 15 activating and 15 inhibitory. First, we looked at this network from the perspective of Boolean dynamics. In this modeling framework time is discrete and genes can be on (1) or off (0). As an update rule, which computes the state of a gene (0 or 1) as a function of its inputs, we used the simple, but biologically plausible, majority rule employed for example in Li et al. [48]. A node i is switched on at time $t + 1$, $x_i(t + 1) = 1$, if the sum over all activating and inhibitory influences entering from active nodes at time t is larger than 0, $S_i(t) = \sum_j A_{ij}x_j(t) > 0$, where matrix elements A_{ij} are -1 ($+1$) for an inhibitory (activating) link from node j to node i and 0, if node j is not connected to node i . If the sum $S_i(t)$ is smaller than 0, the node i will be switched off, $x_i(t + 1) = 0$. If the activating and inhibitory influences balance, $S_i(t) = 0$, the state of the node does not change, $x_i(t + 1) = x_i(t)$. Simulating this system, starting from random initial conditions, leads to fixed-point attractors and cyclic attractors, into which the system settles after a short transient of a few time steps. We assumed that the measured gene expression patterns are sampled from these attractors, rather than from the transients. Furthermore, we did not distinguish between fixed-point attractors and cyclic attractors, each of which is allowed to contribute a single time point ('snapshot') to our data acquisition procedure. Figure 2B shows examples of such attractors.

Using attractors derived from 2000 random initial conditions, we employed a variant of the ESABO ('entropy shift under Boolean operations') method described in Claussen et al. [17], which yields a score for each pair of nodes in the network. This ESABO score takes the vectors formed by the on-off states of each of the two genes across all attractors and evaluates, how these vectors change under a Boolean AND. A z-score is then computed, comparing the composition of this resulting vector to a vector obtained via an AND operation now applied to pairs of randomly shuffled vectors, i.e. random vectors with the same overall composition in 0 and 1 as the original two gene vectors. This z-score is the ESABO score of the two genes. Employing upper and lower thresholds for the ESABO score allowed us to define positive and negative interactions among genes and thus infer a regulatory network with inhibitory and activating links. For the example at hand, the inferred network is depicted in Figure 2C. Note that this method determines only undirected links. In principle, it is possible to augment this method to infer directed links, for example by looking at the asymmetry between (0,1) and (1,0) in the pairs of gene vectors entering the ESABO analysis or other information about the state fluctuations at each end point of the undirected links in the reconstructed network. In this simple illustrative example we did not pursue this line of investigation.

As a second approach for simulating gene expression patterns from the network given in Figure 2A, we employed the GeneNetWeaver software described in Schaffter et al. [67], which has also been used to generate synthetic data for the DREAM (Dialogue for Reverse Engineering Assessments and Methods) network inference challenges [52]. We generated 10 time series of 1000 (dimensionless) time steps. Computing the Pearson correlation between the resulting simulated expression vectors for pairs of genes and imposing an upper (lower) threshold on the correlation coefficients to define activating (inhibitory) links, we inferred the network shown in Figure 2D. Again, this simplistic method yields only undirected links. Analyzing the change of the correlation coefficient under time delays between the two gene expression arrays may be a plausible strategy for inferring the direction of a link. Also, many of the methods described in the

next section (e.g., partial correlations, versions of the mutual information and dynamic Bayesian networks) make possible such directionality inference.

For the inferred network shown in Figure 2C for example, the Jaccard index for the positive links comparing the reconstruction with the original from Figure 2A (intersection of the two link sets divided by the union) is 0.52, compared to 0.05 ± 0.04 for a random selection of links. For the negative links the Jaccard index is 0.46, compared to 0.048 ± 0.034 for random data. With our choice of ± 5 as the threshold for the ESABO scores, the inferred network is denser than the original network. In Figure 3 the number of links is shown as a function of the threshold (Figure 3A), together with the corresponding Jaccard indices (Figure 3C). The same analysis for the network inferred from the GeneNetWeaver simulation is shown in Figures 3B and 3D.

In addition to illustrating important aspects of network inference, this little example introduces several fundamental principles of modeling and analyzing gene regulation, and the range of mathematical models potentially employed for this type of modeling [details see 41, 44], in particular the notion and usefulness of Boolean attractors ([details see 48, 21, 9]. Many of these points have been excellently summarized in Bornholdt [8]. An early review of mathematical models of gene expression patterns, as well as data analysis approaches employing clustering of expression data to identify genes under common regulation is found in D’Haeseleer et al. [26].

4. Several methods of gene network inference

Network inference is less demanding than the broader topic of identifying from data the (often nonlinear) dynamical rules underlying the system at hand [see, e.g., 77, for a detailed discussion of this more general field].

The first approach described in the previous section, for the data derived from Boolean dynamics, is an information-theoretical method. The second approach, for continuous data, is the simplest example of a correlation-based method. Beyond these approaches, there are causality-based methods, meth-

ods based on Bayesian networks [62], regression-based methods and various approaches relying on machine learning [see, e.g., 77, 44, 3]. A combination of Bayesian networks and Markov chain Monte Carlo (MCMC) has been proposed in Mukherjee and Speed [58] to refine the graphical model approaches to network inference. A useful classification of inference methods is given in De Smet and Marchal [24]. A successful strategy, proposed in [52] in the context of the DREAM challenges, is to use an aggregation of different methods.

A major issue is to distinguish between direct and indirect interactions, that is, to avoid the spurious identification as a link of what is merely a network path between two nodes. This issue has been addressed with partial correlations [23], conditional mutual information [79], network deconvolution [29], by a silencing method of indirect effects based on network structure [5], as well as by refinements of the mutual information approach (e.g., 'part mutual information' [80] or 'conditional mutual inclusive information' [78]) and many more [see 53, for additional examples].

The challenges of network inference, as well as the current panel of methods, have recently been summarized in Brugere et al. [13]. One challenge is to integrate data from multiple sources (e.g., gene expression profiles from different data sets). In Castro et al. [15] this challenge is addressed via a multitask learning approach, in contrast to inferring a single network per data source and then combining these individual networks into a consensus network (as frequently done in typical network inference methods; see Marbach et al. [52]). An alternative is to include prior information (e.g., regulatory interactions already known or inferred from other data sets) in the network inference process [70]. Inferring networks from noisy data [76] and assessing the robustness of network reconstructions are also prominent topics in the recent literature.

Important algorithmic approaches and methodological innovations are excellently summarized in a range of review articles and book chapters [e.g., 24, 44, 27, 13, 66, 40]. Delgado and Gómez-Vela [25] provide a clear and detailed survey of the underlying assumptions and parameters of network inference methods, as well as a summary of frequently used measures of reconstruction

quality of the resulting methods.

5. Structure and dynamics of gene regulation

In the spirit of the distinction done in neural networks between structural and functional connectivity, it is here essential to distinguish the hardware of gene regulation and the dynamics producing gene expression patterns.

The hardware refers to potential links, corresponding to the ability of a transcription factor (or a small RNA) to bind some regulatory domain on the genome, generally the promoter or enhancers of a gene. Each link thus represents only a potentiality. Actually many hardware architectures are compatible with a given set of expression data. Henceforth this hardware cannot be inferred from gene expression measurements but rather established link-wise from ChIP-chip data or other dedicated experiments. The dynamics refers to the actual activity of these potential physical links, namely which links are actually at work, with which strength. The experimental data underlying the dynamical information (microarray or RNA-Seq data) is thus distinct from structural information.

The current view is that the dynamic network can be inferred from measurements of gene expression and correlation analysis, with possibly taking into account prior structural knowledge in the form of constraints. We suggest another way of addressing the question of gene regulation inference, which can be formulated as: given the hardware, how constrained are expression data. A related question is, whether the observed gene expression patterns are compatible with the network. Approaches for measuring the agreement of a gene expression profile with a given regulatory network have, for example, been discussed in Cline et al. [18], Marr et al. [56, 55], Cowen et al. [20]. This approach of quantifying the agreement between a gene expression pattern and a network can also be addressed from a slightly different perspective: Not only can we ask for the evidence that a gene regulatory network has *produced* this expression pattern. We can also ask, whether this pattern is meaningful to *emulate* the other networks in a cell e.g., the metabolic network or the protein interaction

network. In the case of metabolic networks such an approach has been pursued in Sonnenschein et al. [73, 72], Knecht et al. [43]. Another question is whether the observed gene expression patterns can even be predicted from the information contained in the network. Simulation here provides part of the answer, as detailed in the next section.

6. The promises of gene regulatory network simulation

To exploit the dynamical systems viewpoint on gene regulatory networks, a requirement is to solve the challenge of simulating expression data. Benchmark studies proceed by first, generating a realistic/plausible gene network, then generating realistic/plausible gene expression patterns (i.e. expression matrix of g genes by c chip measurements). In our above example, gene expression patterns for the *in silico* network were derived from GeneNetWeaver [54].

However, it should be noted that even in the case of *E. coli*, gene expression patterns can by no means be fully explained by the available transcriptional regulatory network. A range of other biological processes and factors exert a systematic influence on the gene expression pattern. It is, for example, by now widely accepted that even in the case of bacteria, chromosomal structure and the positioning of genes in the circular chromosome exerts an important regulatory influence [74, 75, 65, 55, 12].

7. Perspectives in Systems Medicine

Since long ago, genetics, based for instance on family studies, has determined the impact of gene sequence variation in diseases. More recently, genome-wide association studies performed on large cohorts of patients and controls provided more quantitative and wider-scope results in the form of statistical association between single-nucleotide polymorphisms and diseases [37]. In a second step, these pointwise variants are mapped to genes (currently the closest genes along the genome) [2]. These results are available in a public database, the GWAS catalog, and offer an unprecedented ground for developing personalized medicine.

In the present context, we suggest that functional insights could be gained by investigating the location of these at-risk genes on regulatory networks.

However, several difficulties have to be circumvented in the use of gene regulatory networks for personalized medicine. That external factors, e.g. a metabolite, may influence gene expression has been discovered long ago [38, 57]; they could have either an action of induction or repression of the gene transcription. As a consequence, the context-dependence of gene networks should be taken into account for any efficient and reliable medical application. Also, gene expression levels involve not only transcriptional regulation but also mRNA and protein degradation or sequestration (RNA granules). Additional levels of regulation are thus present. The notion of multiplexes, that is, networks sharing their set of nodes but composed of different kinds of links, is promising to account for several types of relationships [7, 59]. However, much work remains to be done before turning this notion in an operational tool.

8. Conclusion

In this introductory chapter, we provide a simple illustrative example of gene network inference. We emphasized the distinction between network structure and dynamical processes 'running on' this structure, i.e. the distinction between structural connectivity and functional connectivity, as an underrated conceptual foundation of the field. We finally embedded the topic in a broader interdisciplinary context with some guidelines to explore the existing literature.

The paradigm of 'network medicine' is based upon the notion that diseases, as well as treatments, can be understood as alterations of the 'cellular-molecular network' [71]. This view is driven by the representation of disease-gene associations as a bipartite network, termed the 'diseasome' [32, 4], and the idea that functional biological networks (protein interaction networks, gene regulatory networks, metabolic networks, etc.) can be related to the diseasome via the identification in these networks of dysregulated 'disease modules', embedding the associated genes on the diseasome [4, 51].

The current trend in medical applications of network methods is rather data integration than network inference. Data integration here can mean

- (1) merging multi-'omics' data in a common, network-based framework (e.g., as represented by the 'Mergeomics' approach in Shu et al. [69]),
- (2) using a given biological network to interpret 'omics' data (as the network coherences from Sonnenschein et al. [72], Knecht et al. [43]),
- (3) enhancing data using a given network (e.g., via network propagation, as summarized in Cowen et al. [20]).

Several promising findings suggest that the detailed reconstruction and analysis of regulatory networks, together with the integration of ever more and diverse data, can shed light on some basic principles underlying disease onset and progression. In Emilsson et al. [28], the co-variation of extracellular proteins circulating in the blood (the serum proteome) among elderly individuals revealed strong associations between features of the serum protein network and disease states. In a case study focusing on psoriasis, Zhao et al. [81] show that a regulatory network derived from multi-'omics' data is informative about potential key drivers of the disease.

Machine learning will undoubtedly play a prominent role in the applications of network inference and network-based data analysis techniques to medicine. Here it could be helpful to first create a machine learning model of the healthy situation (e.g., represented by the corresponding regular network) and then use this model to predict differences between the disease state and the healthy state, as suggested in Camacho et al. [14].

9. Further Reading

Bornholdt, S., 2005. Less is more in modeling large genetic networks. *Science* 310, 449–451.

Brugere, I., Gallagher, B., Berger-Wolf, T.Y., 2018. Network structure inference, a survey: Motivations, methods, and applications. *ACM Computing Surveys (CSUR)* 51, 24.

Jurman, G., Filosi, M., Visintainer, R., Riccadonna, S., Furlanello, C., 2019. Stability in grn inference, in: *Gene Regulatory Networks*. Springer, pp. 323–346.

Karlebach, G., Shamir, R., 2008. Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology* 9, 770.

Le Novère, N., 2015. Quantitative and logic modelling of molecular and gene networks. *Nature Reviews Genetics* 16, 146.

Marbach, D., Costello, J.C., Küffner, R., Vega, N.M., Prill, R.J., Camacho, D.M., Allison, K.R., Aderhold, A., Bonneau, R., Chen, Y., et al., 2012. Wisdom of crowds for robust gene network inference. *Nature methods* 9, 796.

Sanguinetti, G., Huynh-Thu, V.A., 2019. Gene regulatory network inference: an introductory survey, in: *Gene Regulatory Networks*. Springer, pp. 1–23.

References

- [1] Alon, U., 2007. Network motifs: theory and experimental approaches. *Nature Reviews Genetics* 8, 450.
- [2] Altshuler, D., Daly, M.J., Lander, E.S., 2008. Genetic mapping in human disease. *science* 322, 881–888.
- [3] Babbie, A.C., Chan, T.E., Stumpf, M.P., 2017. Learning regulatory models for cell development from single cell transcriptomic data. *Current Opinion in Systems Biology* 5, 72–81.
- [4] Barabási, A.L., Gulbahce, N., Loscalzo, J., 2011. Network medicine: a network-based approach to human disease. *Nature reviews genetics* 12, 56.
- [5] Barzel, B., Barabási, A.L., 2013. Network link prediction by global silencing of indirect correlations. *Nature biotechnology* 31, 720.
- [6] Bergenholm, D., Liu, G., Holland, P., Nielsen, J., 2018. Reconstruction of a global transcriptional regulatory network for control of lipid metabolism

in yeast by using chromatin immunoprecipitation with lambda exonuclease digestion. *Msystems* 3, e00215–17.

- [7] Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C.I., Gómez-Gardenes, J., Romance, M., Sendina-Nadal, I., Wang, Z., Zanin, M., 2014. The structure and dynamics of multilayer networks. *Physics Reports* 544, 1–122.
- [8] Bornholdt, S., 2005. Less is more in modeling large genetic networks. *Science* 310, 449–451.
- [9] Bornholdt, S., 2008. Boolean network models of cellular regulation: prospects and limitations. *Journal of the Royal Society Interface* 5, S85–S94.
- [10] Bornholdt, S., Kauffman, S., 2019. Ensembles, dynamics, and cell types: Revisiting the Statistical Mechanics perspective on cellular regulation. *Journal of theoretical biology* .
- [11] Boucher, B., Jenna, S., 2013. Genetic interaction networks: better understand to better predict. *Frontiers in genetics* 4, 290.
- [12] Browning, D.F., Grainger, D.C., Busby, S.J., 2010. Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Current opinion in microbiology* 13, 773–780.
- [13] Brugere, I., Gallagher, B., Berger-Wolf, T.Y., 2018. Network structure inference, a survey: Motivations, methods, and applications. *ACM Computing Surveys (CSUR)* 51, 24.
- [14] Camacho, D.M., Collins, K.M., Powers, R.K., Costello, J.C., Collins, J.J., 2018. Next-generation machine learning for biological networks. *Cell* .
- [15] Castro, D.M., De Veaux, N.R., Miraldi, E.R., Bonneau, R., 2019. Multi-study inference of regulatory networks for more accurate models of gene regulation. *PLoS computational biology* 15, e1006591.

- [16] Choi, J.A., Wyrick, J.J., 2017. RegulatorDB: a resource for the analysis of yeast transcriptional regulation. Database 2017.
- [17] Claussen, J.C., Skiecevičienė, J., Wang, J., Rausch, P., Karlsen, T.H., Lieb, W., Baines, J.F., Franke, A., Hütt, M.T., 2017. Boolean analysis reveals systematic interactions among low-abundance species in the human gut microbiome. PLoS computational biology 13, e1005361.
- [18] Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B., et al., 2007. Integration of biological networks and gene expression data using Cytoscape. Nature protocols 2, 2366.
- [19] Consortium, E.P., et al., 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57.
- [20] Cowen, L., Ideker, T., Raphael, B.J., Sharan, R., 2017. Network propagation: a universal amplifier of genetic associations. Nature Reviews Genetics 18, 551.
- [21] Davidich, M.I., Bornholdt, S., 2008. Boolean network model predicts cell cycle sequence of fission yeast. PloS one 3, e1672.
- [22] Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., et al., 2002. A genomic regulatory network for development. science 295, 1669–1678.
- [23] De La Fuente, A., Bing, N., Hoeschele, I., Mendes, P., 2004. Discovery of meaningful associations in genomic data using partial correlation coefficients. Bioinformatics 20, 3565–3574.
- [24] De Smet, R., Marchal, K., 2010. Advantages and limitations of current network inference methods. Nature Reviews Microbiology 8, 717.

- [25] Delgado, F.M., Gómez-Vela, F., 2019. Computational methods for gene regulatory networks reconstruction and analysis: A review. *Artificial Intelligence in Medicine* 95, 133 – 145. URL: <http://www.sciencedirect.com/science/article/pii/S09333365718303865>, doi:<https://doi.org/10.1016/j.artmed.2018.10.006>.
- [26] D’Haeseleer, P., Liang, S., Somogyi, R., 2000. Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics* 16, 707–726.
- [27] Dong, X., Yambartsev, A., Ramsey, S.A., Thomas, L.D., Shulzhenko, N., Morgun, A., 2015. Reverse enGENEering of regulatory networks from big data: a roadmap for biologists. *Bioinformatics and biology insights* 9, BBI–S12467.
- [28] Emilsson, V., Ilkov, M., Lamb, J.R., Finkel, N., Gudmundsson, E.F., Pitts, R., Hoover, H., Gudmundsdottir, V., Horman, S.R., Aspelund, T., et al., 2018. Co-regulatory networks of human serum proteins link genetics to disease. *Science* 361, 769–773.
- [29] Feizi, S., Marbach, D., Médard, M., Kellis, M., 2013. Network deconvolution as a general method to distinguish direct dependencies in networks. *Nature biotechnology* 31, 726.
- [30] Gama-Castro, S., Salgado, H., Santos-Zavaleta, A., Ledezma-Tejeida, D., Muniz-Rascado, L., García-Sotelo, J.S., Alquicira-Hernández, K., Martínez-Flores, I., Pannier, L., Castro-Mondragón, J.A., et al., 2015. RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. *Nucleic acids research* 44, D133–D143.
- [31] Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R., et al., 2012. Architecture of the human regulatory network derived from ENCODE data. *Nature* 489, 91.

- [32] Goh, K.I., Cusick, M.E., Valle, D., Childs, B., Vidal, M., Barabási, A.L., 2007. The human disease network. *Proceedings of the National Academy of Sciences* 104, 8685–8690.
- [33] Gonçalves, E., Nakic, Z.R., Zampieri, M., Wagih, O., Ochoa, D., Sauer, U., Beltrao, P., Saez-Rodriguez, J., 2017. Systematic analysis of transcriptional and post-transcriptional regulation of metabolism in yeast. *PLoS computational biology* 13, e1005297.
- [34] Guelzim, N., Bottani, S., Bourguin, P., Képès, F., 2002. Topological and causal structure of the yeast transcriptional regulatory network. *Nature genetics* 31, 60.
- [35] Guet, C.C., Elowitz, M.B., Hsing, W., Leibler, S., 2002. Combinatorial synthesis of genetic networks. *Science* 296, 1466–1470.
- [36] Han, X., Shen, Z., Wang, W.X., Di, Z., 2015. Robust reconstruction of complex networks from sparse data. *Physical review letters* 114, 028701.
- [37] Hardy, J., Singleton, A., 2009. Genomewide association studies and human disease. *New England Journal of Medicine* 360, 1759–1768.
- [38] Jacob, F., Monod, J., 1961. On the regulation of gene activity, in: *Cold Spring Harbor Symposia on Quantitative Biology*, Cold Spring Harbor Laboratory Press. pp. 193–211.
- [39] Jeffrey, J., 2013. *Hypernetworks in the science of complex systems*. volume 3. World Scientific.
- [40] Jurman, G., Filosi, M., Visintainer, R., Riccadonna, S., Furlanello, C., 2019. Stability in GRN inference, in: *Gene Regulatory Networks*. Springer, pp. 323–346.
- [41] Karlebach, G., Shamir, R., 2008. Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology* 9, 770.

- [42] Kauffman, S.A., 1969. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of theoretical biology* 22, 437–467.
- [43] Knecht, C., Fretter, C., Rosenstiel, P., Krawczak, M., Hütt, M.T., 2016. Distinct metabolic network states manifest in the gene expression profiles of pediatric inflammatory bowel disease patients and controls. *Scientific reports* 6, 32584.
- [44] Le Novère, N., 2015. Quantitative and logic modelling of molecular and gene networks. *Nature Reviews Genetics* 16, 146.
- [45] Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., et al., 2002. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298, 799–804.
- [46] Lesne, A., Bécavin, C., Victor, J.M., 2012. The condensed chromatin fiber: an allosteric chemo-mechanical machine for signal transduction and genome processing. *Physical biology* 9, 013001.
- [47] Levine, M., Davidson, E.H., 2005. Gene regulatory networks for development. *Proceedings of the National Academy of Sciences* 102, 4936–4942.
- [48] Li, F., Long, T., Lu, Y., Ouyang, Q., Tang, C., 2004. The yeast cell-cycle network is robustly designed. *Proceedings of the National Academy of Sciences* 101, 4781–4786.
- [49] Liu, G., Marras, A., Nielsen, J., 2014. The future of genome-scale modeling of yeast through integration of a transcriptional regulatory network. *Quantitative Biology* 2, 30–46.
- [50] Lonfat, N., Duboule, D., 2015. Structure, function and evolution of topologically associating domains (TADs) at HOX loci. *FEBS letters* 589, 2869–2876.

- [51] Loscalzo, J., Barabasi, A.L., 2011. Systems biology and the future of medicine. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 3, 619–627.
- [52] Marbach, D., Costello, J.C., Küffner, R., Vega, N.M., Prill, R.J., Camacho, D.M., Allison, K.R., Aderhold, A., Bonneau, R., Chen, Y., et al., 2012. Wisdom of crowds for robust gene network inference. *Nature methods* 9, 796.
- [53] Marbach, D., Prill, R.J., Schaffter, T., Mattiussi, C., Floreano, D., Stolovitzky, G., 2010. Revealing strengths and weaknesses of methods for gene network inference. *Proceedings of the national academy of sciences* 107, 6286–6291.
- [54] Marbach, D., Schaffter, T., Mattiussi, C., Floreano, D., 2009. Generating realistic in silico gene networks for performance assessment of reverse engineering methods. *Journal of Computational Biology* 16, 229–239.
- [55] Marr, C., Geertz, M., Hütt, M.T., Muskhelishvili, G., 2008. Dissecting the logical types of network control in gene expression profiles. *BMC systems biology* 2, 18.
- [56] Marr, C., Theis, F.J., Liebovitch, L.S., Hütt, M.T., 2010. Patterns of subnet usage reveal distinct scales of regulation in the transcriptional regulatory network of *Escherichia coli*. *PLoS computational biology* 6, e1000836.
- [57] Monod, J., Jacob, F., 1961. General conclusions: teleonomic mechanisms in cellular metabolism, growth, and differentiation, in: *Cold Spring Harbor symposia on quantitative biology*, Cold Spring Harbor Laboratory Press. pp. 389–401.
- [58] Mukherjee, S., Speed, T.P., 2008. Network inference using informative priors. *Proceedings of the National Academy of Sciences* 105, 14313–14318.

- [59] Nissim, L., Perli, S.D., Fridkin, A., Perez-Pinera, P., Lu, T.K., 2014. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Molecular cell* 54, 698–710.
- [60] Nüsslein-Volhard, C., Wieschaus, E., 1980. Mutations affecting segment number and polarity in *Drosophila*. *Nature* 287, 795–801.
- [61] Pearcy, N., Chuzhanova, N., Crofts, J.J., 2016. Complexity and robustness in hypernetwork models of metabolism. *Journal of theoretical biology* 406, 99–104.
- [62] Perrin, B.E., Ralaivola, L., Mazurie, A., Bottani, S., Mallet, J., d’Alche Buc, F., 2003. Gene networks inference using dynamic Bayesian networks. *Bioinformatics* 19, ii138–ii148.
- [63] Porfiri, M., Ruiz Marín, M., 2018. Inference of time-varying networks through transfer entropy, the case of a Boolean network model. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28, 103123.
- [64] Rickels, R., Shilatifard, A., 2018. Enhancer logic and mechanics in development and disease. *Trends in Cell Biology* .
- [65] Rocha, E.P., 2008. The organization of the bacterial genome. *Annual review of genetics* 42, 211–233.
- [66] Sanguinetti, G., Huynh-Thu, V.A., 2019. Gene regulatory network inference: an introductory survey, in: *Gene Regulatory Networks*. Springer, pp. 1–23.
- [67] Schaffter, T., Marbach, D., Floreano, D., 2011. GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods. *Bioinformatics* 27, 2263–2270.
- [68] Shen-Orr, S.S., Milo, R., Mangan, S., Alon, U., 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature genetics* 31, 64.

- [69] Shu, L., Zhao, Y., Kurt, Z., Byars, S.G., Tukiainen, T., Kettunen, J., Orozco, L.D., Pellegrini, M., Lusi, A.J., Ripatti, S., et al., 2016. Mergeomics: multidimensional data integration to identify pathogenic perturbations to biological systems. *BMC genomics* 17, 874.
- [70] Siahpirani, A.F., Roy, S., 2016. A prior-based integrative framework for functional transcriptional regulatory network inference. *Nucleic acids research* 45, e21–e21.
- [71] Silverman, E.K., Loscalzo, J., 2013. Developing new drug treatments in the era of network medicine. *Clinical Pharmacology & Therapeutics* 93, 26–28.
- [72] Sonnenschein, N., Dzib, J.F.G., Lesne, A., Eilebrecht, S., Boulkroun, S., Zennaro, M.C., Benecke, A., Hütt, M.T., 2012. A network perspective on metabolic inconsistency. *BMC systems biology* 6, 41.
- [73] Sonnenschein, N., Geertz, M., Muskhelishvili, G., Hütt, M.T., 2011. Analog regulation of metabolic demand. *BMC systems biology* 5, 40.
- [74] Travers, A., Muskhelishvili, G., 2005a. Bacterial chromatin. *Current opinion in genetics & development* 15, 507–514.
- [75] Travers, A., Muskhelishvili, G., 2005b. Dna supercoiling—a global transcriptional regulator for enterobacterial growth? *Nature Reviews Microbiology* 3, 157.
- [76] Wang, B., Pourshafeie, A., Zitnik, M., Zhu, J., Bustamante, C.D., Batzoglou, S., Leskovec, J., 2018. Network enhancement as a general method to denoise weighted biological networks. *Nature communications* 9, 3108.
- [77] Wang, W.X., Lai, Y.C., Grebogi, C., 2016. Data based identification and prediction of nonlinear and complex dynamical systems. *Physics Reports* 644, 1–76.

- [78] Zhang, X., Zhao, J., Hao, J.K., Zhao, X.M., Chen, L., 2015. Conditional mutual inclusive information enables accurate quantification of associations in gene regulatory networks. *Nucleic acids research* 43, e31–e31.
- [79] Zhang, X., Zhao, X.M., He, K., Lu, L., Cao, Y., Liu, J., Hao, J.K., Liu, Z.P., Chen, L., 2011. Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information. *Bioinformatics* 28, 98–104.
- [80] Zhao, J., Zhou, Y., Zhang, X., Chen, L., 2016. Part mutual information for quantifying direct associations in networks. *Proceedings of the National Academy of Sciences* 113, 5130–5135.
- [81] Zhao, Y., Jhamb, D., Shu, L., Arneson, D., Rajpal, D.K., Yang, X., 2019. Multi-omics integration reveals molecular networks and regulators of psoriasis. *BMC systems biology* 13, 8.

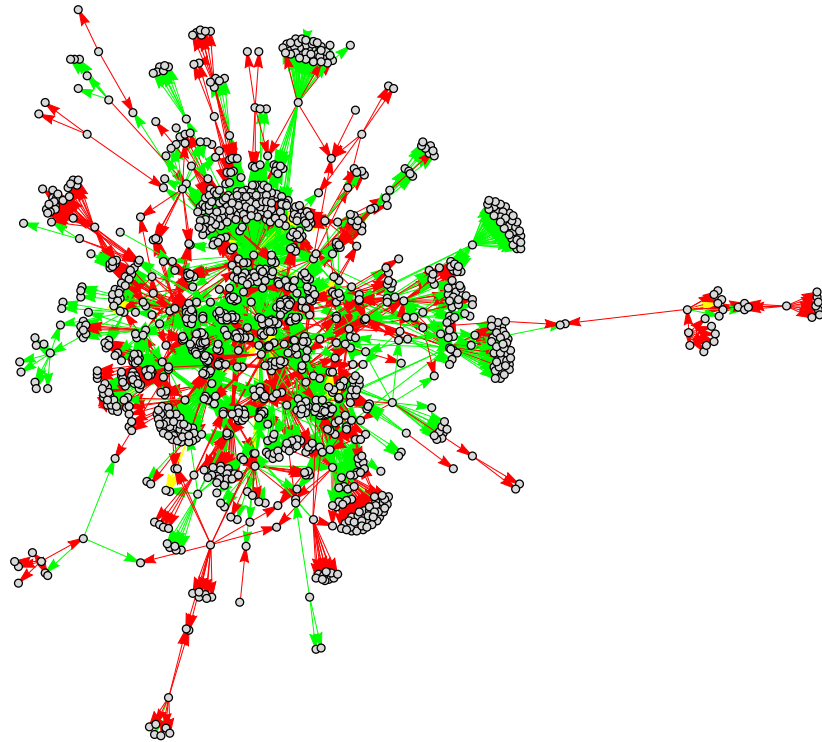


Figure 1: Transcriptional regulatory network of *E. coli* derived from RegulonDB (version 10.5; only data supported by literature with experimental evidence has been used). Nodes are genes, directed links indicate the regulatory action of the transcription factor encoded by the start-point gene onto the end-point gene. Only links with level of evidence labeled as 'strong' in RegulonDB are shown. Activating links are shown in green, inhibitory links are shown in red. Dual links (where the sign of the interaction depends on the conditions) are displayed in yellow.

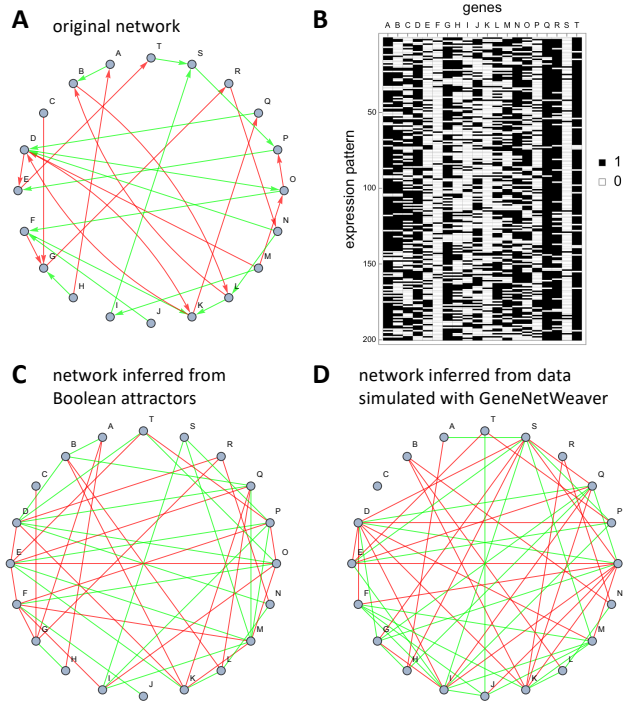


Figure 2: Illustrative example of gene network inference. (A) A random graph consisting of 20 nodes (labeled A to T) with 30 directed links, 15 activating (green) and 15 inhibitory (red). (B) 200 examples of fixed-point attractors obtained by simulating Boolean dynamics on the network shown in (A). (C) Network inferred from a set of 893 distinct Boolean attractors obtained from 2000 random initial conditions using ± 5 as the threshold for the ESABO score between two nodes (see text). (D) Network inferred from continuous data simulated with the GeneNetWeaver tool [67] using a threshold of ± 0.3 for the Pearson correlation coefficient between two nodes (see text).

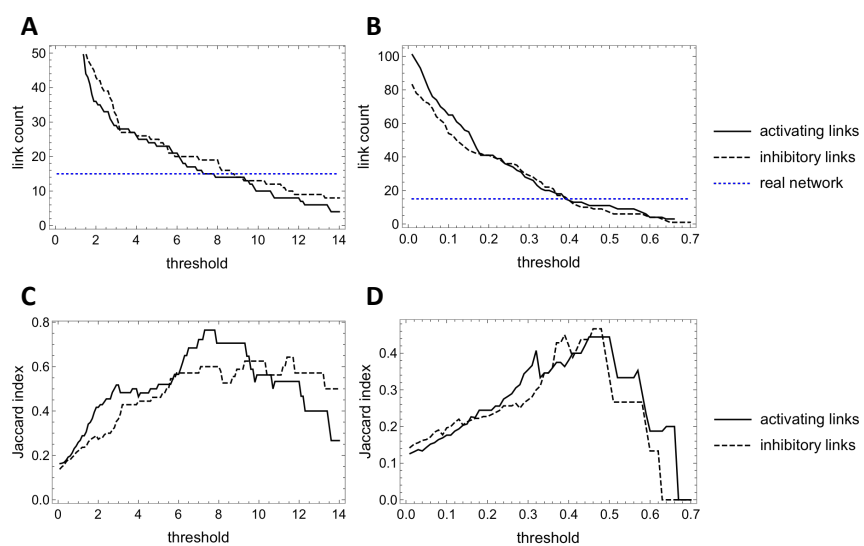


Figure 3: (A) Number of activating (full curve) and inhibitory (dashed curve) links in the network inferred from Boolean dynamics (as in Figure 2C). The blue dotted line indicates the number of activating and inhibitory links in the original network. (C) Jaccard index for the comparison of link sets between the inferred network and the original network. Full curve: activating links; dashed curve: inhibitory links. (B)+(D): Same for the network inferred from the GeneNetWeaver simulation.