Automated Inference of Gene Regulatory Networks Using Explicit Regulatory Modules

Clémence Réda and Bartek Wilczyński

école normale supérieure paris-saclay

Wydział Matematyki, Informatyki i Mechaniki
Uniwersytet Warszawski

Problem

Efficient identification of the causal connections between genes, their products and transcription factors is key to understanding how defects in their function may translate to developmental defects or other diseases.

Recent gene regulatory network models use binding measurements, but usually only to define gene-to-gene interactions, ignoring cis-regulatory structure.

Contributions

We suggest a modelling framework that explicitly encompasses cis-regulatory action on genes, that is flexible and can be used for existing Boolean Networks[3], using an "expansion" procedure. Expanded "Cis-Regulatory" models ensure that some useful, biologically meaningful properties hold for associated regulatory functions:

1. **Monotonicity** with respect to the regulators.
2. **Decomposability** into TF binding clusters.

We include this in an automated pipeline, similar to [1], to infer all gene regulatory networks that can account for input experimental gene expression patterns.

Method

Network Inference Given Experiments And Putative Model

Putative Model[2] ⟶ Expanded Putative Model ⟶ SMT Solving[1]

Cis-Regulatory Model M

- G set of biological entities of interest.
- $C = \{C_g | g \in G\}$ set of cis-regulatory modules (CRMs) by regulated gene.
- T set of TF bindings to each CRM.

$M = (N, I)$, N set of nodes, I set of regulatory interactions:

$$\begin{aligned} N =& G \cup C \cup T \\ I \subseteq& \cup_{g \in G} ((G - T) \times \{g\} \times \{+, -\}) \text{: post-transcriptional actions} \\ & \cup \cup_{g \in G} \cup_{c_g \in C_g} (T_{c_g} \times \{c_g\} \times \{+, -\}) \text{: TF bindings} \\ & \cup \cup_{g \in G} (C_g \times \{g\} \times \{+, -\}) \text{: cis-regulatory interactions} \end{aligned}$$

Decomposability into regulatory modules

$g \in G$, associated regulatory function f_g is the composition of every cis-regulatory module function f_c, associated to module $c \in C_g$, to which TFs from T_c bind, iff.

$$\exists r_g \in (\mathbb{B}^{|G|} \to \mathbb{B}), \forall \text{ system state } q \in \mathbb{B}^{|G|}, f_g(q) = r_g(\prod_{c \in C_g} f_c(q_{|T_c}), q_{|G-T}).$$

Monotonicity with respect to the regulators

Regulatory function f_g associated with $g \in G$ is monotonic with respect to regulators r of g

$$\text{iff. } \epsilon_g(r) \cdot f_g(G - \{r\} = q_{|G-\{r\}}, r = 1) \leq \epsilon_g(r) \cdot f_g(G - \{r\} = q_{|G-\{r\}}, r = 0)$$

where $\epsilon_g(r) = 1$ if r is a repressor (resp. -1 if it is an activator) of g.

Conclusion

This modelling framework allows:

1. **To observe directly redundancy in TF bindings**.

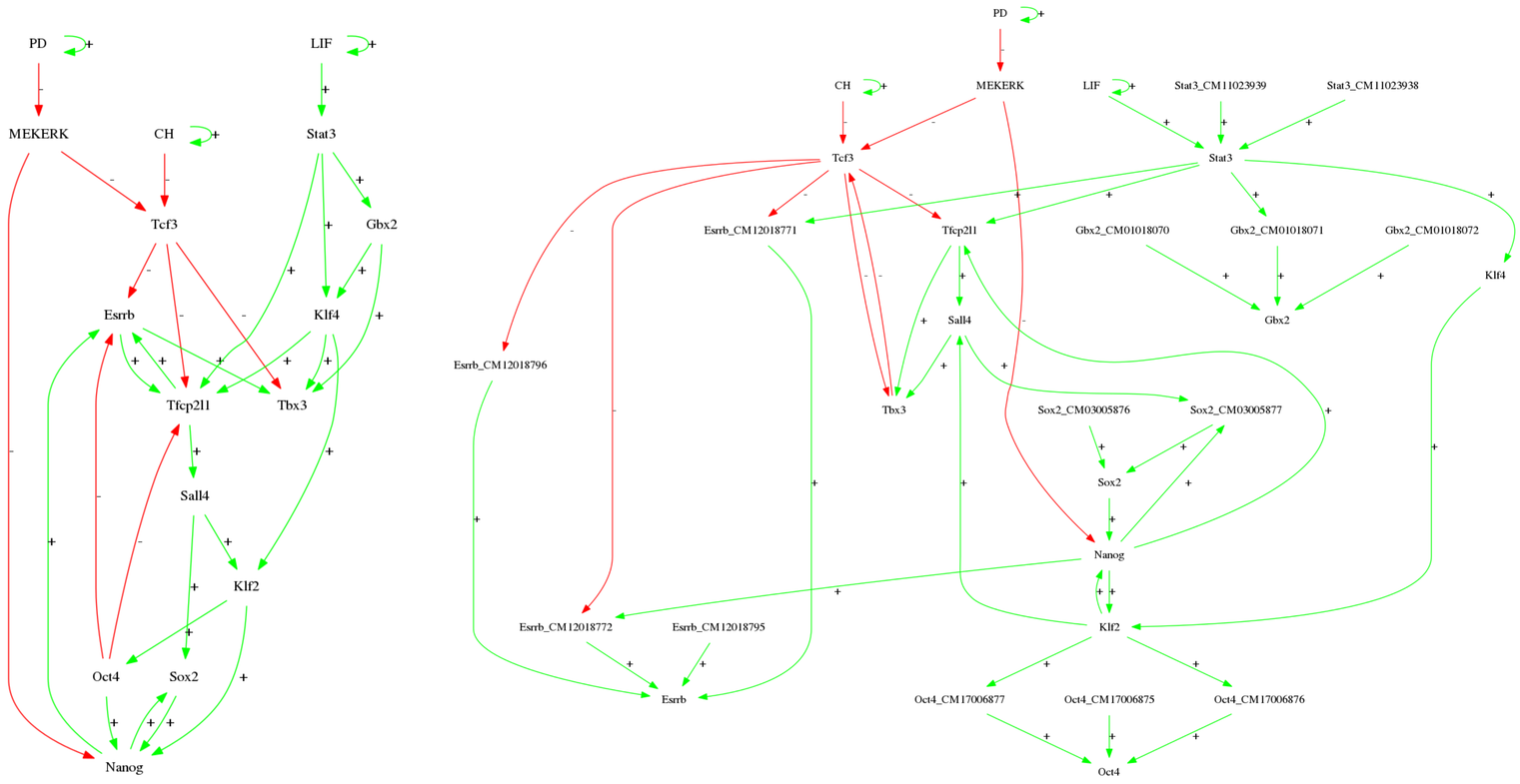
   This might be helping the system to maintain its behavior even in case of mutations. This redundancy between regulatory elements matches redundancy of gene expression patterns.

2. **To give a possible interpretation of the different solutions found by the automated inference method in terms of *binding site availability***.

   This, in turn, can lead to speculation on different possible molecular scenarii behind the identified regulatory functions.

Given the fast progress both in identification of CRMs and quicker sampling of expression, there is an increasing need for extension of regulatory network models to include actual connectivity based on molecular measurements. Our work shows, as a proof-of-concept, that this supplementary information may be useful to better understand regulatory models.

Non-expanded and expanded Dunn[2] model



The decomposition of the gene regulatory functions with respect to their CRMs is clearly shown in the model, and helps interpreting the influence of regulators on gene expression.

Source Code

Automated method for network inference and expansion algorithm have been implemented in Python and R:

`github.com/regulomics/expansion-network`

References

[1] B. Yordanov et al. A method to identify and analyze biological programs through automated reasoning In NPJ systems biology and applications (2016)

[2] S.-J. Dunn et al. Defining an essential transcription factor program for naive pluripotency In Science (2014)

[3] S. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets In Journal of theoretical biology (1969)