

Testing Hypotheses on Genetic Network Structures

Minimal Requirements for an Experimental Design

Dirk Repsilber
Evolutionary Biology Centre
Uppsala University
SE-75236 Uppsala, Sweden
dirk.repsilber@ebc.uu.se

Hans Liljenström
Biometry and Informatics
SLU, Box 7013
SE-750 07 Uppsala, Sweden
hans.liljenstrom@bi.slu.se

ABSTRACT

Reverse-Engineering-algorithms aim at using gene expression data to reconstruct interactions in small regulatory genetic networks [1,2,3]. This may help to understand the basis of gene regulation, the core task of functional genomics [4]. A number of "independent" samples are necessary to reengineer even the smallest regulatory networks with reasonable confidence [5].

We systematically tested the requirements for the experimental design necessary for ranking alternative hypotheses about the structure of a given regulatory network. A Genetic Algorithm [6] was used to explore the parameter space of a multistage discrete genetic network model with fixed connectivity and number of states per node.

We have used the genetic algorithm approach to identify the most probable structure of small genetic networks, for which a set of possible hypotheses exists. Our methods have been applied to simulated network data, but they are aimed for real gene expression data from microarrays. We found that 1) the ranking of hypotheses is easier the more experimental treatments (environments) used for the data set, 2) the number of "fixed" parameters increases with the number of treatments, 3) some errors in the hypothetical network structure may pass undetected, due to a maintained dynamical behavior, and 4) it is not necessary to determine all parameters in order to rank the hypotheses. The results clearly show that, for most cases, our methods are adequate in ranking the hypotheses for small genetic networks, pointing to the most probable network structure (interdependence of gene products).

In a previous study, Akatsu *et al.* [5] estimated the number of independently measured expression patterns necessary for the full reverse engineering of a Boolean model (two states; $S = 1$) of a genetic regulatory network. They found that at least 30 treatments are needed for the full reverse engineering of a Boolean network model with five nodes (even for non-noisy data). As we are employing regulatory network nodes with $S > 1$, we expect an even higher number of necessary treatments, but we have not performed any systematic investigations to obtain exact values. In our case study, three treatments were below the minimal requirements for most of the network types investigated. This means that results from expression level measurements based on less than three treatments are not sufficient to make any statement about the structure of a 5-node regulatory network (even given our most simple model). Here, the only alternative is

adding even more prior knowledge to further reduce the parameter space for the unknown parameters (for example for the interaction rules or the transition function), or increasing the number of treatments.

As one of our main results, we found that it is not necessary to reverse engineer all structural parameters, if the task is only to obtain a ranking of the hypotheses. In addition, since a number of parameters in the hypotheses (in our case the link-list) were fixed, the parameter space was narrowed down considerably. Consequently, for our ranking task we were able to considerably reduce the number of necessary treatments, as compared to the full reverse engineering task.

Based on our results, we emphasize the need for synchronized modeling and experimental design. Computationally automated algorithmic solutions that build upon previous biological knowledge are here indispensable.

REFERENCES

- [1] Somogyi, R. and Sniegoski, C.A. (1996) Modelling the complexity of genetic networks: Understanding multigenic and pleiotropic regulation, *Complexity*, 1, 45-63.
- [2] Smolen, P. Baxter, D.A. and Byrne, J.H. (2000) Modelling transcriptional control in gene networks - methods, recent results, and future directions, *Bulletin of Mathematical Biology* 62, 247-292.
- [3] D'haeseleer, P., Liang, S., and Somogyi, R. (2000) Genetic network inference: From co-expression clustering to reverse engineering, *Bioinformatics* 16, 707-726.
- [4] Wahde, M. and Hertz, J. (2000) Coarse-grained reverse engineering of genetic regulatory networks, *BioSystems* 55, 129-136.
- [5] Akutsu, T. Miyano, S., and Kuhara, S. (2000) Inferring qualitative relations in genetic networks and metabolic pathways, *Bioinformatics* 16, 727-734.
- [6] Corcoran, A.L. and Wainwright, R.L. (1993) Libga: A user-friendly workbench for order-based genetic algorithm research, In: *Proceedings of the 1993 ACM/SIGAPP Symposium on Applied Computing (SAC 93)*, Indianapolis, Indiana.