Network Benchmarking: A Happy Marriage between Systems and Synthetic Biology

Jeremy J. Minty,[1] S. Marjan Varedi K.,[1] and Xiaoxia Nina Lin[1,*]
[1]Department of Chemical Engineering, University of Michigan at Ann Arbor, 2300 Hayward, 3074 H.H. Dow, Ann Arbor, MI 48109, USA
*Correspondence: ninalin@umich.edu
DOI 10.1016/j.chembiol.2009.03.004

In their new *Cell* paper, Cantone et al. (2009) present exciting results on constructing and utilizing a small synthetic gene regulatory network in yeast that draws from two rapidly developing fields of systems and synthetic biology.

Systems biology and synthetic biology are two emerging research fields that have risen in recent years at the intersection of conventional biology and other more quantitative disciplines. The former advocates a systems perspective for the elucidation of complex biological networks, from which critical global properties emerge as a consequence of the synergistic interactions among a large number of components. It often features large-scale "omics" datasets generated by high-throughput technologies and sophisticated mathematical modeling tools drawn from other fields such as physics, computer science, and engineering. One representative example is the ultimate objective of recovering large gene regulatory networks based on gene expression data, which in the past decade has attracted numerous talented researchers for the development of intelligent "reverse-engineering" network-inferring algorithms. However, due to the inherent complexity of biology and the lack of "golden standards," it is usually difficult or even impossible to evaluate the performance and/or applicability of these diverse algorithms. On the other hand, the field of synthetic biology engages in the design and construction of artificial biological networks to help understand how natural systems function (e.g., build a synthetic oscillatory gene network to help understand how circadian rhythm is created) or to develop technologies for biomedical, environmental, or other applications.

Given the complementarity of these two fields, one can expect that concepts/tools from one field might be useful for the other. For example, knowledge and modeling formalisms from systems biology research could enable the rational design of new and improved biological functions for synthetic biology (Church, 2005). An alternative type of synergy between the two fields is also possible; namely, the utilization of synthetic biology tools in improving systems biology methods. In the forthcoming issue of *Cell*, Cantone and colleagues (Cantone et al., 2009) presented a pioneering work in this new direction, which beautifully united systems and synthetic biology by constructing a small synthetic gene regulatory network in yeast and utilizing it for in vivo benchmarking of several reverse engineering and modeling approaches.

In the framework of the in vivo reverse-engineering and modeling assessment (IRMA, see Figure 1), Cantone and colleagues (Cantone et al., 2009) started with the design and construction of a small synthetic gene regulatory network in *S. cerevisiae*, a model eukaryotic organism. They created a five-gene network, termed the IRMA network, with well-characterized, nonessential transcription factors (Swi5, Gal80, Ash1, Cbf1, and Gal4) and promoters. Comprehensive knowledge on the chosen components and their interactions were utilized to obtain a network with several distinct features. First, despite its small scale, the IRMA network includes several representative interactions in natural gene regulatory networks, such as transcriptional cascading, positive and negative feedback loops, as well as protein-protein interactions. In fact, the proposed network topology is potentially capable of generating oscillations, a popular theme both in systems-level elucidation of natural networks and in the investigation of artificial gene networks since the birth of synthetic biology (Elowitz and Leibler, 2000). Second, the network could be controlled with an administrable small molecule, galactose, because of the inclusion of a mutually repressive interaction between Gal80 and Gal4 proteins. Hence, the IRMA network could be either activated or inactivated, depending on the presence of galactose. Third, the IRMA network was designed to be robust against inputs from the cellular environment, except the regulation from the carbon source (i.e., galactose or glucose), while the network can affect other genes in the cell. Carefully planned experiments were carried out to rewire the connections among the chosen genes and delete the endogenous ones. This concept of "insulation" is a brilliant one and exemplifies the merit of employing synthetic gene networks for assessing modeling approaches (Stolovitzky et al., 2007). The previous approach was to utilize natural gene networks, which, even when extensively studied, are often connected to unknown (or neglected) cellular or environmental components not captured in the modeling framework and thus compromise, in an unpredictable manner, the reliability of the assessment results. However, it is worth noting that these authors did not experimentally validate the insulation of the constructed network, and some of the experimental complications they later encountered (i.e., transient increase of mRNA levels of two genes during switching of carbon sources) might be due to unknown regulations of the network by external factors (e.g., regulation of protein degradation by environmental signals).

Using the IRMA network, Cantone and colleagues (Cantone et al., 2009) then

 ©2009 Elsevier Ltd All rights reserved

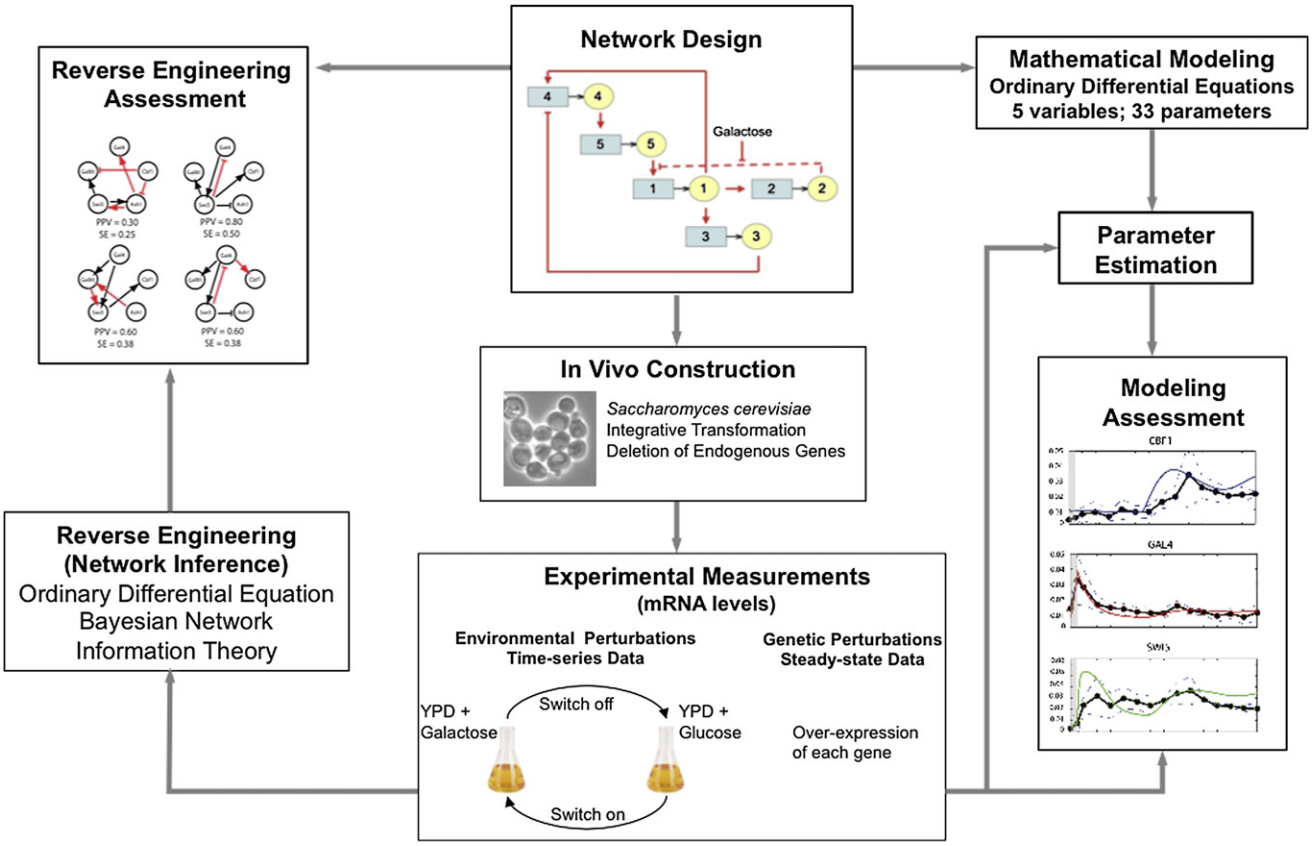

Figure 1. Schematic Diagram of the Overall Framework of IRMA

Schematic diagram of the overall framework of the in vivo reverse-engineering and modeling assessment (IRMA) (Cantone et al., 2009). Starting from an artificial network design involving 5 yeast transcription factors (center), the authors combined reverse engineering assessment, mathematical modeling, and in vivo measurements into an itergrated process for synthetic network based benchmarking of modeling methods. This figure is adapted from Cantone et al., 2009.

performed two types of perturbation experiments. In environmental perturbations, they switched on or off the network by culturing cells in galactose or in glucose, respectively, and collected samples in a time series. In genetic perturbations, they overexpressed the five network genes one-by-one by utilizing a strong constitutive promoter and focused on samples at the steady state. In both cases, mRNA levels (i.e., the transcriptional response of the network) were measured through quantitative real-time PCR.

With a "true" gene network and a good amount of experimental data (both time series and steady state) in hand, Cantone and colleagues (Cantone et al., 2009) were well positioned to evaluate a range of network modeling approaches. They first developed a dynamic model of the IRMA network based on the ordinary differential equation (ODE) formalism. The ODE model included only mRNA concentrations, considering transcription and translation as a single step, and eventually involved 5 variables and 33 parameters, which are related to promoter response, degradation, and specific time delay. The authors then estimated the parameters from the switch-on time series and additional promoter strength experimental data, using a stochastic optimization framework. Next, the ODE model was validated by comparing model predictions with experimental network perturbation data. There was a semiquantitative agreement between the experimental data and the model predictions, and important dynamic behaviors were reasonably captured by the model. Some discrepancies between the model and the in vivo data were noted, however. Besides possible biological complications that the authors commented on, other more fundamental reasons might include the incomplete "insulation" of the network and/or the oversimplification of the model. Including protein dynamics in the model could lead to a more accurate representation of the network, although at the expense of introducing additional parameters, a recurring trade-off between accuracy and knowledge/data requirement in modeling of complex biological networks.

Finally, Cantone and colleagues (Cantone et al., 2009) also utilized all the available gene expression data of the IRMA network to assess systematically three types of reverse-engineering methods for inferring gene regulatory networks. The algorithms evaluated are representative of the ODE-based framework (Gardner et al., 2003), Bayesian network (Friedman et al., 2000), and information theory (IT)-based approaches (Basso et al., 2005). Quantitative metrics, incorporating true positives, false positives, and false negatives, were calculated to evaluate the algorithm performance. The results of this benchmarking exercise yielded new insights into the robustness and the

 ©2009 Elsevier Ltd All rights reserved

strengths/weaknesses of different network inference algorithms. The authors concluded that the ODE approaches performed the best with data sets that captured strong responses (e.g., the "switch-on" dataset, which is rich in dynamic behavior). The tested dynamic Bayesian network algorithm tended to perform better with larger data sets (e.g., the "switch-off" series), because larger datasets facilitate the estimation of probability density distributions utilized in inferring Bayesian networks. The IT-based approach did not perform well for any of the data sets; this algorithm was not expected to work well on small networks. It should be pointed out that comparison of different network inference algorithms must be done with prudence. Most importantly, it remains to be determined whether assessment results from small gene networks, such as the IRMA network, can be reliably extended to large-scale networks with a much larger number of components as well as more diverse and complex interactions. In addition, other factors such as computational requirements and data requirements must also be considered.

Combining all the above modules, the IRMA framework, a first-of-its-kind work, represents a laudable contribution to the development of a synthetic model gene network for use in benchmarking network inference and modeling algorithms. It pushes the envelope of standardization and benchmarking, a trend that is increasingly gaining momentum in synthetic biology (Canton et al., 2008). We envision that more advanced approaches, inspired by and extended from the IRMA framework, will be developed in the coming years to realize the full benefit of utilizing synthetic gene networks for evaluating and improving reverse-engineering and other modeling approaches. We further expect that two types of developments could potentially move forward this new line of research substantially. First, a rich reservoir of gene components for engineering synthetic networks exists; for instance, the yeast genome was found to contain 106 transcription factors that form a larger range of interactions (e.g., auto-regulation and feedforward loops) (Lee et al., 2002). Computational frameworks could be developed to support more systematic and comprehensive design of larger and more complex synthetic gene networks. Second, it will be highly desirable to include proteins in future IRMA-like frameworks. Intriguingly, although protein concentrations were not reported in their current work, Cantone and colleagues (Cantone et al., 2009) labeled all proteins in the IRMA network with different markers, including a green fluorescence tag which can be used to monitor the network response directly at the single-cell resolution. Therefore, the full value of the constructed IRMA network could be further explored by incorporating protein levels. It has been well accepted that the correlation between mRNA and protein abundance is often quite weak in yeast cells (Gygi et al., 1999) due to post-transcriptional regulations (most importantly, translation and degradation). To upgrade the performance of reverse-engineering and other modeling approaches, protein dynamics must be included, both in experimental measurements and in modeling. The ultimate outcome we look forward to seeing from this happy marriage between systems and synthetic biology is the advance of systems biology approaches to the next level of quantitativeness through the support of synthetic biology tools, which in turn will enable the rational design and realization of new in vivo biological functions, a main goal of synthetic biology.

REFERENCES

Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Nat. Genet. *37*, 382–390.

Canton, B., Labno, A., and Endy, D. (2008). Nat. Biotechnol. *26*, 787–793.

Cantone, I., Marucci, L., Iorio, F., Ricci, M.A., Belcastro, V., Bansal, M., di Bernardo, M., Santini, S., di Bernardo, D., and Cosma, M.P. (2009). Cell, in press. Published online March 26, 2009. 10.1016/j.cell.2009.01.055.

Church, G.M. (2005). Mol. Syst. Biol. *1*, 32.

Elowitz, M.B., and Leibler, S. (2000). Nature *403*, 335–338.

Friedman, N., Linial, M., Nachman, I., and Pe'er, D. (2000). J. Comput. Biol. *7*, 601–620.

Gardner, T.S., di Bernardo, D., Lorenz, D., and Collins, J.J. (2003). Science *301*, 102–105.

Gygi, S.P., Rochon, Y., Franza, B.R., and Aebersold, R. (1999). Mol. Cell. Biol. *19*, 1720–1730.

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). Science *298*, 799–804.

Stolovitzky, G., Monroe, D., and Califano, A. (2007). Ann. N Y Acad. Sci. *1115*, 1–22.

 ©2009 Elsevier Ltd All rights reserved