DEFINING THE PLAYERS IN HIGHER-ORDER NETWORKS: PREDICTIVE MODELING FOR REVERSE ENGINEERING FUNCTIONAL INFLUENCE NETWORKS

JASON E. MCDERMOTT[1], MICHELLE ARCHULETA[1], SUSAN L. STEVENS[2], MARY P. STENZEL-POORE[2], AND ANTONIO SANFILIPPO[1]

[1]*Pacific Northwest National Laboratory, Richland, WA, USA*

[2]*Oregon Health & Science University, Portland, OR, USA*

Determining biological network dependencies that can help predict the behavior of a system given prior observations from high-throughput data is a very valuable but difficult task, especially in the light of the ever-increasing volume of experimental data. Such an endeavor can be greatly enhanced by considering regulatory influences on co-expressed groups of genes representing functional modules, thus constraining the number of parameters in the system. This allows development of network models that are predictive of system dynamics. We first develop a predictive network model of the transcriptomics of whole blood from a mouse model of neuroprotection in ischemic stroke, and show that it can accurately predict system behavior under novel conditions. We then use a network topology approach to expand the set of regulators considered and show that addition of topological bottlenecks improves the performance of the predictive model. Finally, we explore how improvements in definition of functional modules may be achieved through an integration of inferred network relationships and functional relationships defined using Gene Ontology similarity. We show that appropriate integration of these two types of relationships can result in models with improved performance.

1. Introduction

Stroke is currently the second leading cause of death in the Western world [1] and is estimated to cause 10% of deaths worldwide. Patients who do not die from a stroke suffer from neurological impairment that is significantly disabling in a large percentage of survivors. Preconditioning by induction of a small stroke or treatment with Toll-like receptor (TLR) agonists prior to induction of a large stroke provides a significant degree of neuroprotection in animal models [2]. To provide molecular level understanding of the dynamics of stroke processes we have previously used high-throughput transcriptomic profiling using microarrays to follow the dynamics of stroke and neuroprotection in a mouse model [2, 3]. Predictive models of regulatory and functional processes occurring during neuroprotection and stroke would offer a very powerful tool to investigate novel methods for prevention and treatment of this important disease.

Models that can predict aspects of system behavior from the observation of a small number of system inputs or components have been largely limited to very general models [4], focused models that can be fully parameterized, or models for which there is a large body of existing data about the molecular interactions between components [5]. Inference of specific interactions between large numbers of system components is limited by the number of observations of the

system being examined. Specific interactions of interest include protein-protein interactions, interactions between signal transduction pathway members (e.g. phosphorylation events), and transcription factor mediated regulatory events (activation or repression of a gene or set of genes). Even with high-throughput experimental techniques most experimental designs are limited in their ability to produce detailed molecular networks of regulatory influences on a system-wide level. An alternative to determination of mechanistic networks between individual components is to constrain the parameter space by considering networks that describe the most important regulatory influences between groups of genes that represent important functions [6], here called functional influence networks. Functional influence networks involve regulatory processes that govern a specific set of system responses. The networks can be represented as causal influences between regulators, which mediate transitions between system states, and functional modules that provide the mechanism of action for the system [7]. For example, immune cells such as macrophages respond to certain kinds of stimuli (e.g. pathogen detection) by activating an inflammatory program that includes the transcriptional activation and subsequent release of inflammatory cytokines, pro-inflammatory effectors, and other components of the inflammatory program [8, 9]. These responses are regulated by a set of transcription factors (e.g. AP1, NFκB, and IRFs) that respond to pathogen associated molecular patterns (PAMPs) detected by TLRs. In a functional influence network the inflammatory response genes responding with similar dynamics would be considered to be functional modules and the genes that regulate their activation would be considered their regulatory influences. In this way the dynamic behavior of the network is simplified to facilitate modeling and represented only as expression patterns that represent collections of similarly behaving genes.

Modeling the dynamic behavior of functional influence networks makes it possible to chart the development of a biological network through time, with reference to experimental evidence from gene expression data. For example, Tegner et al. (2003) have created a method that models the change in each gene's expression as a linear process [10]. Another algorithm created recently for such dynamic modeling uses an ODE model for regulatory dynamics and L1 shrinkage as a means of selecting parsimonious models [11, 12]. The result is a coupled set of ODEs, each ODE describing the expression of a set of co-regulated genes as a function of the expression of genes identified as being regulators. A separate model is learned for each functional module, with each model defining the network edge connections between that cluster and its regulators and assigning strengths (coefficients) to each such regulatory interaction. Thus, the approach infers the regulatory network structure as it builds individual dynamic models for each regulated functional module.

Despite the significance of dynamic regulatory models, the performance of many inference methods is highly dependent on the initial clustering techniques. Inference methods require determination of subtle differences in patterns of gene expression profiles to best identify co-regulated functional modules. Unfortunately gene expression data has inherent noise and standard clustering techniques applied to limited sets of observations will inappropriately identify clusters. Existing knowledge, for example functional information about genes represented in the Gene Ontology (GO), can be used to augment clustering approaches. Methods to incorporate knowledge-driven techniques into predictive models of pathways have been recently proposed in which the GO is used to filter [13], enrich [14] or restructure [15] gene associations inferred from gene expression data through reverse engineering methods. These approaches have been shown to improve the biological plausibility of the network inferences drawn and the accuracy of the predictive models built. However, they still treat data- and knowledge-based inferences as incommensurable inputs, and the impact of each approach on the inferred network is factored in separately.

The goal of the current study is to show how dynamic modeling using functional influence networks can be used to infer the important regulatory influences that drive neuroprotection or injury during stroke in a mammalian model system and how incorporation of data from other sources can be used to improve model performance. We accomplish this using clustering, network topology and functional associations to refine components of functional influence networks (regulators and functional modules). We then use a machine-learning approach to learn relationships between components that can be used to robustly predict system dynamics. Our results show that predictive modeling in complex eukaryotic systems can be a useful way of generating hypotheses about the high-level functional regulation of the system, even with relatively few observations of the system. This approach provides valuable information about the processes of neuroprotection and injury during stroke in a whole animal model system, and generates a number of interesting hypotheses that are being experimentally validated.

2. Methods

2.1. Data sets

Briefly, we used a dataset of microarray results from blood of mice in a neuroprotection study, and data processing was performed as previously described [3]. The dataset comprises five treatments; ischemic preconditioning, lipopolysaccharide (LPS) or CpG injection, or control treatments, saline injection and sham surgery. The samples were taken 3, 24 and 72 hours post-preconditioning treatment, a stroke was induced at 72 hours then two more samples, 3 and 24 hours post-stroke induction, were taken for each preconditioning treatment.

2.2. Co-expression networks

We filtered this data to exclude probes that do not change significantly (p value > 0.05, fold-change expression < 2.0) resulting in 7352 transcripts. The expression levels of these transcripts (fold change relative to control untreated animals) were used as input to the CLR method [16] and the resulting relationships were filtered to a Z score of 5.0, yielding a network with 1880 nodes and 14205 relationship edges. We inverted the adjacency matrix for this network and treated it as a distance matrix for hierarchical clustering using complete linkage agglomeration and cut the dendrogram to generate 46 clusters to serve as initial targets for modeling.

The igraph library in the R statistical language was used to calculate the topology of the inferred networks. Bottlenecks are considered to be those genes with high betweenness centrality measures in the network [17, 18]; the highest 2.5% in this study.

2.3. Predictive modeling cross-validation approach

To infer a predictive regulatory model of neuroprotection during stroke we expanded on an algorithm that was previously applied to transcriptomics from prokaryotes and yeast [11, 12, 19]. We first applied the multivariate regression method, the Inferelator, to the targets defined from network analysis (above) using sets of potential regulators as described in the text. This method infers parsimonious sets of regulatory influences between regulators and targets (functional modules). In the learned model the relation between the expression of a target (y) and the expression levels of regulators with non-null influences on y (X) is expressed as:

$$\tau \frac{dy}{dt} = -y + \sum \beta_j X_j \tag{1}$$

Here, τ is the time step used in model construction and β is the weight for relationship X on y as determined by L_1 shrinkage using least angle regression [20]. To make predictions using a learned model eq. 1 can be solved for y, the expression of the target cluster. Assuming equilibrium conditions the derivative dy/dt is 0 and so equation (1) can be represented simply as a linear weighted sum:

$$y = \sum \beta_j X_j \tag{2}$$

and the dynamic version (for time series) is expressed for each time point (m) as:

$$y_m = \frac{-y_{m-1} + \sum \beta_j X_{m-1\,j}}{\tau} - y_{m-1} \tag{3}$$

In our modeling we used a τ of 30 minutes, which is appropriate for mRNA dynamics in a eukaryote [21].

Given the limited amount of transcriptomic data available for training we wanted to ensure that the models being inferred were robust, that is, that they were predictive of target expression under novel conditions not included in the training data. To accomplish this we employed a cross-validation approach to evaluate the performance of inferred models using different starting components (sets of regulators or target clusters, as described in the text). In the cross-validation the transcriptomic data is divided into five sets based on the treatment (i.e., LPS, CpG or ischemic preconditioning pretreatment, or saline or sham control treatments; see Figure 2B), five models are trained on the data excluding each treatment set in turn, and the performance of each model is evaluated on the left out set. Performance is evaluated as the average correlation of observed versus predicted expression values for each target weighted by the number of genes in each target, to produce a weighted gene-normalized overall performance score for the model, as:

$$P = \frac{\sum_{i=1}^{T} corr(pred_i, obs_i) n_i}{\sum_{i=1}^{T} n_i} \qquad (4)$$

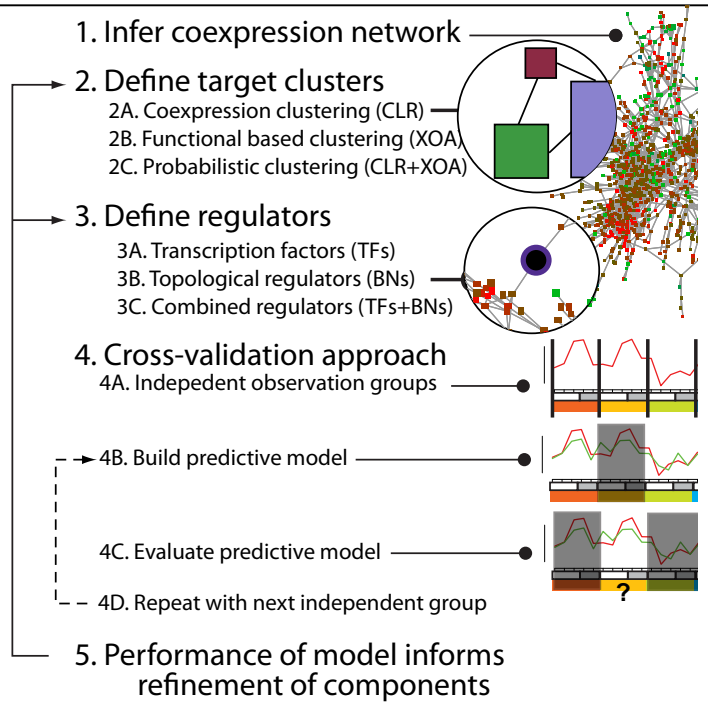

Figure 1. Overview of iterative cross-validation predictive modeling approach. 1). Network inference from transcriptomic data using CLR. **2)** Definition of target clusters for modeling using several partitioning methods. **3)** Definition of potential regulators from existing knowledge or topological analysis. **4)** Cross-validation of predictive model: **A**. Divide expression data into related independent groups of observations (i.e. different treatments); **B**. Build a predictive model using all but one group with the Inferelator; **C.** Evaluate the performance of the model using the left out group; **D**. Repeat with next independent group. **5)** Use the overall predictive performance to evaluate and refine methods used to determine the network components (targets [2] and regulators [3]).

where P is the overall performance score, T is the number of targets in the model, *pred* and *obs* are the predicted and observed expression patterns, respectively, and n is the number genes in the target i. This cross-validation approach allows relatively unbiased assessment of model performance because the data used to evaluate the model is not included in the training data.

2.4. Probabilistic integration of relationships

Our previous results showed that partitioning co-regulated clusters of genes using either CLR or XOA associations could improve the performance

of our cross-validated model. We were interested in combining the networks generated by both methods. Our approach was to treat the score for association between two genes as a p value for each method (see below), then partition the parent target cluster into subclusters using hierarchical clustering. We then used the predictive model generated for the genes in the subclustered target to assess which approach provides the best performance. We tested several approaches for integrating p values: maximum p value, minimum p value, mean p value, and the product of p values. Associations unique to either approach were transferred into the final similarity matrix directly, thus creating a union set of associations. Though the product of p values is the appropriate probabilistic combination of p values, the other methods were used because they may be more appropriate for specific instances. Additionally, the p values from each method do not have exactly the same meaning due to the differences in assumptions used in generating them. For CLR p values we converted the output of CLR (Z scores for the edge relative to the all other edges for each interaction partner) to p values using the normal distribution in R.

The p value for an XOA relationship is obtained by comparing the observed XOA score against the distribution of (a sample of) all possible scores obtained by computing the XOA similarity between all pairs of GO terms from the three subontologies. For example, the p value 0.14 associated with the XOA score of 3.76 assessing the similarity of GO:0007179 (BP: TGF-beta receptor signaling pathway) and GO:0016301 (MF: kinase activity) indicates that fewer than 14% of all XOA scores have higher semantic similarity than 3.76. Higher XOA scores are regularly found in association with lower p values. For example, statistically relevant values (< 0.05) typically correspond to XOA scores above 4.73. The p value across gene expresses the same idea, since the semantic similarity between two genes is the highest XOA score found pairing GO categories across the two genes:

$$XOA(GP1, GP2) = max\ XOA(c1_i, c2_j) \quad (5)$$

where i=1,…,n and j=1, …,m, $GP1$ and $GP2$ are genes, $c1_i$ is one of the GO categories associated with $GP1$, and $c2_j$ one of the GO categories associated with $GP2$.

3. Results and Discussion

3.1. Reverse-engineering by predictive modeling of transcriptomic data

We are interested in developing a predictive model of neuroprotection in stroke at a systems level. There are significant gaps in knowledge about the regulation, functional mechanisms, and components that are involved in neuroprotection and stroke. These gaps prevent the development of molecular-level representations of the stroke process. We therefore have chosen to use a reverse-engineering approach that considers the regulatory influences and functional processes

that these influences induce at a more abstract level. The resulting models will still provide useful and interpretable predictions that can be used for further experimental or computational investigation.

Our approach is to develop a predictive model of transcriptomic data using a machine-learning approach and cross-validation, and use the ability of this model to predict behavior under novel conditions as a way to refine the reverse engineering process (Figure 1). The reverse-engineering algorithm [11, 12] uses multivariate regression to learn ordinary differential equations (ODEs) that describe the relationship between the expression levels of a parsimonious set of regulators and the target functional module. Here, we apply this approach to a higher eukaryotic system with observations that are focused specifically on stroke response and neuroprotection.

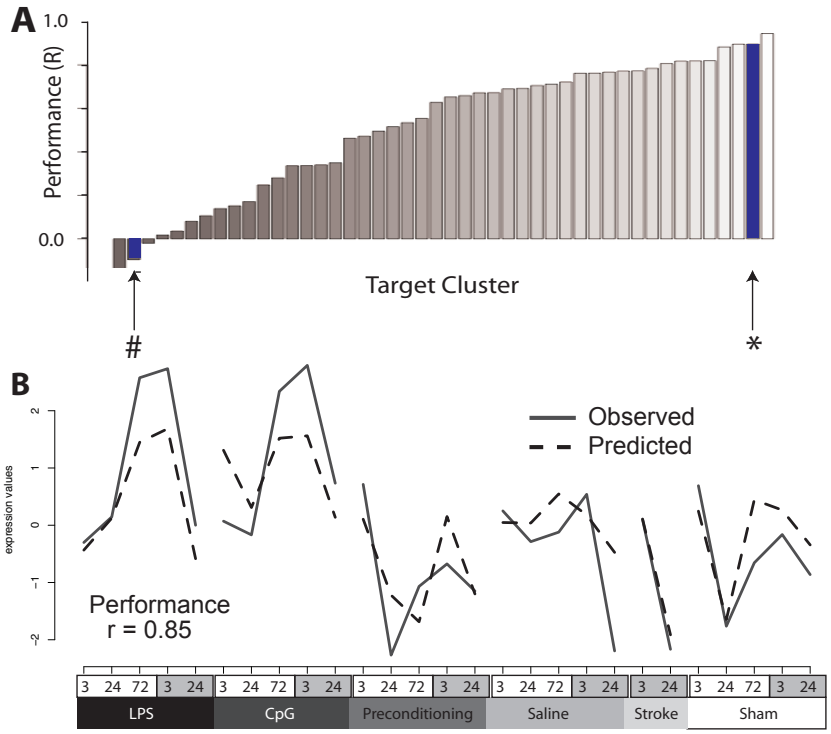

Figure 2. Performance of a predictive model of neuroprotection and injury during stroke in a mouse model system. A. Target cluster performance. The coexpressed clusters used as targets for modeling are shown (X axis) with bar height (Y axis) indicating the performance (correlation of predicted versus observed expression) for that target in the cross-validation approach. # indicates the poorly performing cluster used in further partitioning and * indicates the accurately predicted cluster shown in panel B. **B. Expression of an accurately predicted target.** The observed (red line) versus predicted (green line) expression levels (Y axis) for one cluster representing 180 genes is shown over the treatments/time points (X axis). The independent groups used in the cross-validation are indicated in colored boxes, and time points post-treatment (white boxes) and post-stroke induction (grey boxes) are also shown.

To define functional modules that are the targets in the model we used a transcriptomic data set from a mouse stroke model to infer functional relationships between genes using the context likelihood of relatedness (CLR) method [16] and used hierarchical clustering to define targets (see Methods). We initially treated all genes annotated as transcription factors (85 genes in the network) as potential regulators for reverse engineering.

To evaluate the performance of the model in a relatively unbiased manner we used a cross-validation approach (see Figure 1) that allows all the observations of the system to be treated as

'independent' data sets. We obtained an overall model performance of 0.52 (mean correlation per gene) observed versus predicted expression. In Figure 2A we show the performance (Y axis) of each cluster in the model (bars) ordered by performance. We mark the performance bar corresponding to the poorly performing cluster used for further analysis (see below) with a number sign and mark the bar corresponding to a well-predicted cluster with an asterisk. In Figure 2B we show the predicted (green line) and observed (red line) expression of the well-predicted cluster marked in panel A, over all the conditions examined (Y axis). The shaded bars below the X axis in Fig. 2B show the independent groups used for cross-validation. This correlation between observed and predicted expression shows that the model is robustly predictive of the behavior of the majority of the genes considered. This is an important result as it shows that regulatory influences that act as predictors can be learned from a relatively limited set of expression data. We note that the model itself provides a large number of interesting predictions about regulatory influences and expression of particular functional groups that are the focus of future studies. In this study we use this output of the model (predicted target behavior) to refine the components and relationships that are used for model generation.

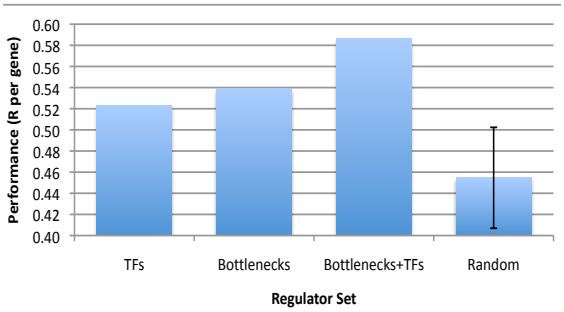

Figure 3. Bottlenecks are complementary to transcription factors as candidate regulatory influences. Predictive models were constructed using annotated transcription factors (TFs), topological bottlenecks, or a combination of the two groups (X axis). The mean and standard deviation (error bar) of ten randomly selected sets of genes is shown as a control. Performance (Y axis) using our cross-validation approach indicates that bottlenecks are robustly predictive of system behavior.

3.2. Network topology identifies important points of regulatory control

Many approaches for reverse-engineering regulatory networks preselect regulators based on sequence-based annotation, and then attempt to identify regulatory relationships between these sets of transcriptional regulators. Functional influence networks may be driven by mediators that are not transcriptional regulators, but could include effectors (e.g. immune effectors), signaling pathway components, metabolic enzymes, or any other component whose change mediates or reflects major changes in the state of the system. Previously our research has suggested the hypothesis that topological bottlenecks identified from transcriptional coexpression networks represent mediators of state transitions in systems [18, 22]. We thus tested the ability of topological bottlenecks to predict system behavior reasoning that true mediators of system transitions should be more predictive of system behavior than randomly chosen differentially regulated genes.

We examined the ability of bottlenecks to serve as regulators in our cross-validated modeling. As a comparison we randomly selected ten sets of differentially expressed genes in the network and evaluated their ability to predict the behavior of the targets in the model. Our results (Figure 3) show the performance of models that include transcription factors only, bottlenecks only, a combination of bottlenecks and transcription factors, or the mean of ten randomly chosen sets of genes. Bottlenecks provide modestly better performance than either the transcription factors set used initially or randomly selected genes. Furthermore, combining the transcription factors with the list of bottlenecks further improved the ability of the resulting model to predict expression behavior under novel conditions. This shows that the expression of bottlenecks is somewhat predictive of system behavior.

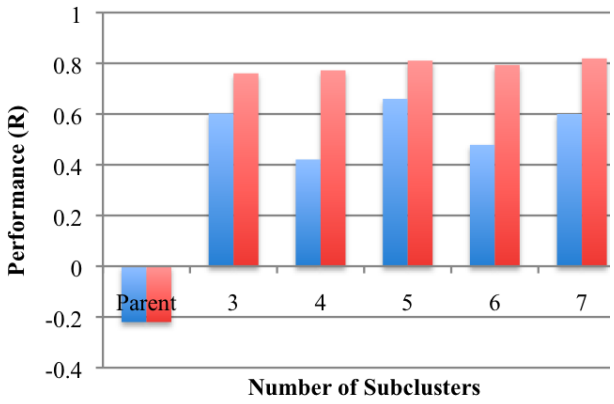

Figure 4. Performance of CLR and XOA defined subclusters for prediction. The parent cluster was subclustered using either the CLR (red)- or XOA (blue)-derived associations between genes into the indicated number of subclusters. Performance (mean correlation of observed versus predicted expression levels) is shown on the Y axis. These results support our previous observations that both methods can improve performance.

A surprising result of this analysis was that the randomly selected gene sets performed significantly worse than any of the selected regulator groups but the performance was still high (R = 0.45). This is likely to be due to the limited number of observations of the system that we are using for this work. Essentially the model is able to identify randomly selected genes which are somewhat predictive of the behavior of the targets because the dynamics of expression over the limited observations are relatively simple. Adding additional observations and/or data gathered for other purposes (TLR agonist treatment of mice, e.g.) should improve performance of our model. Further study is required to determine whether bottlenecks are indeed robustly predictive of system behavior.

3.3. Probabilistic integration of relationships improves delineation of functional modules

We next wanted to examine how the model could be further improved by better determination of target clusters. We examined how best to partition target gene clusters by combining results from the CLR and XOA algorithms to delineate subclusters. As a test case we focused on a problematic cluster with very poor performance (Figure 2A) identified in our previous study [15]. This cluster is made up of 335 genes and has a performance of -0.22 (correlation of predicted versus observed behavior) in the original model.

In Figure 4 we show the cross-validation performance on this cluster subdivided the cluster into 3-7 subclusters using either CLR (red bars) or XOA (blue bars) associations. These results show that using both expression-driven (CLR) or function-driven clustering can improve performance of the predictive model dramatically over that of the parent cluster.

We next examined how combining the two sets of associations could improve results. We chose to consider the strength of the associations as p values in order to directly compare the scores from different algorithms. We used four simple methods for combining p values for XOA and CLR scores when there were overlapping associations within a cluster: the minimum XOA or CLR p value, the maximum XOA or CLR p value, the mean of the XOA and CLR p values, and the product of the XOA and CLR p values. As shown in Figure 5, either the mean p value or maximum p value strategy provides the best performing solution for most cluster sizes, showing significant, but modest, improvement for a model composed of four subclusters. These findings indicate that an appropriate combination of approaches can improve the performance of predictive transcriptomic models

4. Conclusions

We have presented an approach to reverse-engineering from limited, but focused, transcriptional datasets and used it to infer functional influence networks of mouse blood during stroke. This approach uses a machine-learning method to iteratively define and refine the components of the network, both potential regulatory influences and coexpressed functional modules that are the targets of prediction (Figure 1). The approach is applicable to problems in which there are not well-established regulatory pathways already understood, where there are a limited number of observations of the system available, and where there may be complex and multiscale effects that need to be captured by the model, but not necessarily explicitly modeled. Our results demonstrate that the approach can be applied to provide biological insight into a complex and poorly understood pathology, such as neuroprotection and injury during ischemic stroke.

We show that a machine-learning method that employs multivariate regression techniques to learn ODEs describing relationships between regulators and target clusters can be applied to model transcriptomic dynamics from multicellular eukaryotic time course samples (Figure 2). This is an advance in modeling such systems that have traditionally been underrepresented in reverse-engineering applications due to their complexity and lack of ‘gold standard’ networks for validation. The results from cross-validation show that the models we produce can predict transcriptomic behavior of the majority of the genes considered under conditions not used to train the models. This approach is limited by the requirement that the gene-expression level changes of the regulatory influences must be indicative of their activity, an assumption that is clearly not true for many regulators. Additionally, regulatory influences inferred from such a

limited set of observations, though predictive of system behavior to a significant degree, are unlikely to be highly accurate. However, this approach provides the foundation for more detailed investigation, both computationally and experimentally. These results represent an important first step toward more detailed and nuanced models of complex systems.

Using network topology we show that highly central bottlenecks are more predictive of system behavior than a similarly sized group of transcription factors (Figure 3). This result is consistent with the notion that bottlenecks from inferred networks represent mediators of transitions between system states [18, 22]. We further show that combining transcription factors and bottlenecks provides even better predictive performance. These gains are modest but statistically significant and we foresee that including more varied observations of the system will improve the results of the modeling, and should improve the definition of important mediators that we identify through network topology. However, the integration of such data will have to be undertaken carefully [23].

In our approach the performance of the predictive models is dependent on definition of the underlying functional modules used as targets for prediction. We initially define functional modules using hierarchical clustering based on expression profiles of genes. This approach gives good performance for a number of resulting clusters (Figure 2A) but does not provide accurate predictions for a number of significantly sized clusters. We show that further subclustering of a poorly performing cluster using either co-expression relationships from CLR or functional relationships from XOA [24] can dramatically improve the gene-wise performance of the parental cluster. Further, we use a probabilistic integration method and show that the combination of the two relationships can provide better performance than either individual method. This relatively simple approach has the advantage of being able to integrate arbitrary kinds of relationships between genes, so long as they can be associated with p values. We are currently examining what other kinds of relationships between genes will improve performance of the predictive models (e.g. protein-protein interactions, phylogenetic relationships).

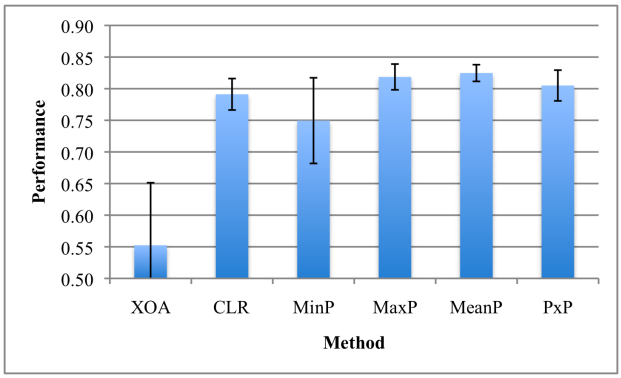

Figure 5. Comparison of subclustering methods. The mean performance of the methods examined (X axis) across different subclustering levels (3-7 clusters, as in Figure 1) is shown (Y axis). The error bars represent one standard deviation. The methods used are XOA and CLR alone, minimum p value (MinP), maximum p value (MaxP), mean of p values (MeanP) and product of p values (PxP). These results show that combining the CLR and XOA associations using probabilities can improve performance over the individual methods alone, but that only when non-standard methods (maximum p value or mean of p values) are employed to do so.

5. Acknowledgements

This work was supported under NIH/NINDS grant R01NS057484-03.

6. References

1. Donnan, G.A., et al., *Stroke.* Lancet, 2008. **371**(9624): p. 1612-23.
2. Stenzel-Poore, M.P., et al., *Effect of ischaemic preconditioning on genomic response to cerebral ischaemia: similarity to neuroprotective strategies in hibernation and hypoxia-tolerant states.* Lancet, 2003. **362**(9389): p. 1028-37.
3. Marsh, B., et al., *Systemic lipopolysaccharide protects the brain from ischemic injury by reprogramming the response of the brain to stroke: a critical role for IRF3.* J Neurosci, 2009. **29**(31): p. 9839-49.
4. Thakar, J., et al., *Modeling systems-level regulation of host immune responses.* PLoS Comput Biol, 2007. **3**(6): p. e109.
5. Oberhardt, M.A., B.O. Palsson, and J.A. Papin, *Applications of genome-scale metabolic reconstructions.* Mol Syst Biol, 2009. **5**: p. 320.
6. De Smet, R. and K. Marchal, *Advantages and limitations of current network inference methods.* Nat Rev Microbiol, 2010.
7. McDermott, J.E., et al., *Separating the drivers from the driven: Integrative network and pathway approaches aid identification of disease biomarkers from high-throughput data.* Dis Markers, 2010. **28**(4): p. 253-66.
8. Glass, C.K. and K. Saijo, *Nuclear receptor transrepression pathways that regulate inflammation in macrophages and T cells.* Nat Rev Immunol, 2010. **10**(5): p. 365-76.
9. Jenner, R.G. and R.A. Young, *Insights into host responses against pathogens from transcriptional profiling.* Nat Rev Microbiol, 2005. **3**(4): p. 281-94.
10. Tegner, J., et al., *Reverse engineering gene networks: integrating genetic perturbations with dynamical modeling.* Proc Natl Acad Sci U S A, 2003. **100**(10): p. 5944-9.
11. Bonneau, R., et al., *A predictive model for transcriptional control of physiology in a free living cell.* Cell, 2007. **131**(7): p. 1354-65.
12. Bonneau, R., et al., *The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo.* Genome Biol, 2006. **7**(5): p. R36.
13. Gamalielsson, J., P. Nilsson, and B. Olsson, *A GO-Based Method for Assessing the Biological Plausibility of Regulatory Hypotheses*, in *ICCS 2006, Part II, LNCS 3992*, V.N.A.e. al., Editor. 2006. p. 879–886.
14. Sanfilippo, A., et al., *Using the gene ontology to enrich biological pathways.* Int J Comput Biol Drug Des, 2009. **2**(3): p. 221-35.
15. McDermott, J., et al., *An Integrated Approach to Predictive Genomic Analytics.*, in *ACM International Conference on Bioinformatics and Computational Biology,*. 2010: Niagra Falls, New York, USA.
16. Faith, J.J., et al., *Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles.* PLoS Biol, 2007. **5**(1): p. e8.
17. Yu, H., et al., *The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics.* PLoS Comput Biol, 2007. **3**(4): p. e59.
18. McDermott, J.E., et al., *Bottlenecks and hubs in inferred networks are important for virulence in Salmonella typhimurium.* J Comput Biol, 2009. **16**(2): p. 169-80.
19. Madar, A., et al., *DREAM3: network inference using dynamic context likelihood of relatedness and the inferelator.* PLoS ONE. **5**(3): p. e9803.
20. Efron, B., et al., *Least angle regression.* Annals of Statistics, 2003. **32**: p. 407-499.
21. Ross, J., *mRNA stability in mammalian cells.* Microbiol Rev, 1995. **59**(3): p. 423-50.
22. Diamond, D.L., et al., *Temporal proteome and lipidome profiles reveal hepatitis C virus-associated reprogramming of hepatocellular metabolism and bioenergetics.* PLoS Pathog, 2010. **6**(1): p. e1000719.
23. Cosgrove, E.J., T.S. Gardner, and E.D. Kolaczyk, *On the Choice and Number of Microarrays for Transcriptional Regulatory Network Inference.* BMC Bioinformatics, 2010. **11**(1): p. 454.
24. Posse, C., et al., *Cross-ontological analytics: Combining associative and hierarchical relations in the gene ontologies to assess gene product similarity.* Lecture Notes in Computer Science, 2006. **3992**: p. 871-878.