INFERRING DISEASE-RELATED PATHWAYS USING A PROBABILISTIC EPISTASIS MODEL

P. N. KANABAR*[1], C. J. VASKE *[1], C. H. YEANG[2], F. H. YILDIZ[3], AND J. M. STUART[1]†

[1] *Department of Biomolecular Engineering, University of California Santa Cruz*
1156 High Street, Santa Cruz, CA 95062, USA
[2] *Institute for Advanced Study, Princeton, New Jersey 08540, USA*
[3] *Department of Microbiology and Environmental Toxicology, University of California Santa Cruz*

Motivation: We present a probabilistic model called a Joint Intervention Network (JIN) for inferring interactions among a chosen set of regulator genes. The input to the method are expression changes of downstream indicator genes observed under the knock-out of the regulators. JIN can use any number of perturbation combinations for model inference (e.g. single, double, and triple knock-outs). **Results/Conclusions**: We applied JIN to a *Vibrio cholerae* regulatory network to uncover mechanisms critical to its environmental persistence. *V. cholerae* is a facultative human pathogen that causes cholera in humans and responsible for seven pandemics. We analyzed the expression response of 17 *V. cholerae* biofilm indicator genes under various single and multiple knock-outs of three known biofilm regulators. Using the inferred network, we were able to identify new genes involved in biofilm formation more accurately than clustering expression profiles.

1. Introduction and Previous Work

DNA microarray expression data provides a rich source of phenotypes to infer gene regulatory networks. Recently, several methods have been developed to learn a cellular network from quantitative observations[9,2]. Notably, Bayesian Networks (BNs) have been used to infer networks from both observational[7,16] and interventional[11] gene expression microarray data. Interventional data can be used to identify causal networks[10]. In addition to gene expression data, BNs have been used to learn signaling networks from phosphorylation data collected from single-cells under perturbations[12]. Al-

[*] equal contributions.
[†] correspondence to jstuart@soe.ucsc.edu

gorithms for exact inference of BN structure with uncertain perturbations have been applied to similar data[5].

In contrast to the standard BN approach, the Nested Effects Model (NEM)[9] infers gene networks from expression changes of "secondary genes" observed under the knock-out of regulatory genes. NEMs search for a network of regulators that are consistent with a nested hierarchy among the expression changes. Like BNs, NEMs provide a generative model that can score a candidate network in proportion to the data likelihood.

Geneticists have used epistasis analysis of complex phenotypes to order genes into pathways for nearly a century[3]. The models considered here assume phenotypes (or expression states in our case) result from the loss of a signal propagating along a genetic network. Epistasis analysis uses the phenotypes of single- and multiple-gene deletions to place genes into a pathway order. In this work, we assume the genes of interest signal to one another according to a "switch regulatory," rather than a "substrate dependent," model as defined by Huang and Sternberg (2006)[8]. In switch regulatory epistasis, if the knock-out of gene A produces phenotype P, the knock-out of B produces phenotype Q, and the double knock-out of AB produces phenotype Q, B is said to be epistatic to, and is placed downstream of, A. In the context of using microarray expression data, the up- or down-regulation of a gene across a panel of mutant regulators does not directly convey the information about the network of regulators. Instead, a secondary gene's relative expression differences in single- versus double-mutants may be informative. Recently, Van Driessche *et al.*[14] used an *ad hoc* approach based on Euclidean distance to compare double and single knock-out expression profiles.

In this paper, we describe the Joint Intervention Network (JIN) model that implements a probabilistic epistasis analysis for quantitative, multivariate phenotypes. JIN differs from previous methods in the following aspects. First, it employs relative expression changes formed from the comparison of all possible pairs of perturbation datasets. Second, it extracts epistatic information from data on multiple perturbation combinations. Third, it is able to determine the order of genes as well as their functional dependencies (e.g., inhibition, activation, multiplicative, or additive) from the downstream effects, and can therefore find functional dependencies between regulators that may not be observed in direct transcript levels. Fourth, it identifies new targets of the regulators that have expression profiles consistent with the predicted regulatory network.

The survival of *Vibrio cholerae* both outside and inside a host, has

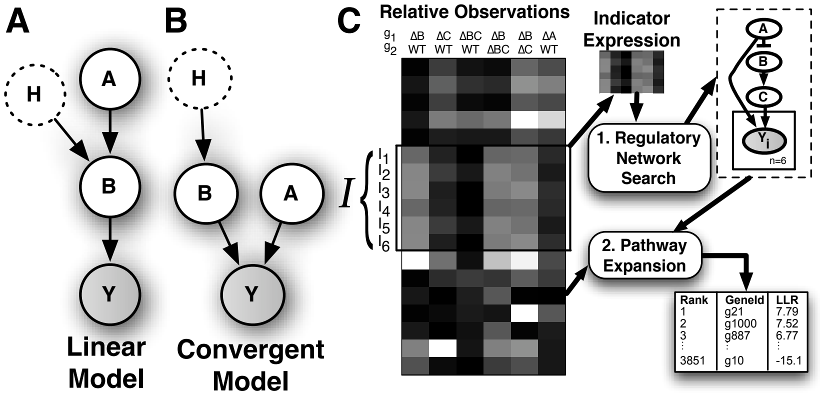

Figure 1. **A**–**B**. Signaling networks used in the toy example discussed in the text. A and B are known regulators, while H is a hidden regulatory influence, Y. **A**. Linear network. B directly controls Y's expression; A indirectly controls Y via B. **B**. Convergent network. A and B both directly control Y's expression. **C**. The input to the Joint Intervention Network model are relative gene expression changes of a set of indicator genes, I, measured under different comparisons of knock-out combinations of a set of regulator genes (A, B, and C), and the genotypes of the comparisons (g_1 and g_2). First, a high scoring network is found that connects regulators (nodes) to each other by predicted interactions (edges) and regulators to indicators (shown in plate notation using a solid box to indicate replicated indicator nodes and edges). Next, this regulatory model is used to find other genes in the genome that have consistent expression changes.

been linked to its ability to form biofilm. We applied JIN to the perturbation expression data of three *vps* biofilm regulators and automatically reconstructed the model consistent with the network proposed by the experimentalists that collected the data[4]. Furthermore, the model identified new target genes involved in biofilm formation.

2. Methods

We assume n signaling genes participate in the regulation of a pathway of interest and that the effect of their regulatory input is observable through indicator expression states, Y. Under different combinations of knock-outs to the signaling genes, indicator expression is used as quantitative phenotypes to order the signaling genes into a network.

2.1. *Theoretical Motivation*

We motivate our approach with an example to illustrate the use of using (1) knock-out combinations and (2) relative expression changes observed under knock-outs. Multiple knock-outs provide information that cannot be revealed by single knock-outs. For example, the networks in Figures 1A and 1B are indistinguishable using single knock-outs since both networks will exhibit down-regulation of indicator gene Y in both single knock-outs ΔA and ΔB. However, the results of the double knock-out ΔAB can help distinguish between the two models. If the model is linear, ΔB and ΔAB have the same effects on Y, but ΔA and ΔAB have distinct effects. ΔA has weaker effect on Y because (hidden) regulators of B, in addition to A, may be present. If the model is convergent and additive, both ΔA and ΔB have weaker effects on Y compared to ΔAB. If the model is convergent and multiplicative, both ΔA and ΔB have the same effect on Y compared to ΔAB.

Adopting relative changes alleviates the quantization problem inherent in discrete models that use direct expression levels. Let the relative expression change (REC) for an indicator gene i, $x_i(g_1, g_2)$, be the $\log_2$-ratio of i's expression level in genotype g_1 over its level in genotype g_2. Consider the linear model in Figure 1A. H represents a hidden factor such as an unknown regulator or an environmental variable not controlled in the experiment. Both single knock-outs ΔA and ΔB and double knock-out ΔAB down-regulate Y relative to wild-type. In addition, Y is lower in ΔB or ΔAB compared to ΔA, as H may provide a constituent activation of Y in ΔA. It is problematic to apply a binary quantization on the direct expression level of Y under ΔA, since Y is between the level of the double knock-out and wild type. More refined quantization also suffers from the arbitrary quantization levels.

In contrast, it is straightforward to determine whether Y under one condition is higher, lower or equal to another condition. In this example, suppose $B = H + A + \epsilon$ and $Y = B + \epsilon$. Each regulator contributes a mean influence of μ to Y's activity and $\epsilon \ll \mu$ is the background noise level. Then $\frac{y(\Delta A)}{y(WT)} \approx \frac{1}{2}$, and $\frac{y(\Delta AB)}{y(WT)} = \frac{y(\Delta B)}{y(WT)} \approx \frac{\epsilon}{2\mu}$, where $y(g)$ represents the continuous expression of Y under genotype g. Clearly, $x_Y(\Delta AB, \Delta A) \approx \log(\frac{\epsilon}{\mu}) \ll 0$ and $x_Y(\Delta AB, \Delta B) \approx 0$. Epistatic reasoning could now place A above B since B's knock-out phenotype matches that of the double mutant.

2.2. *The Joint Intervention Network Model*

The input to the method consists of two parts: an $n \times p$ matrix x_{ij} of RECs and a $2 \times p$ matrix g of the genotypes of each comparison (Fig. 1C, "Indicator Expression"). A row in X corresponds to one indicator gene. Columns in

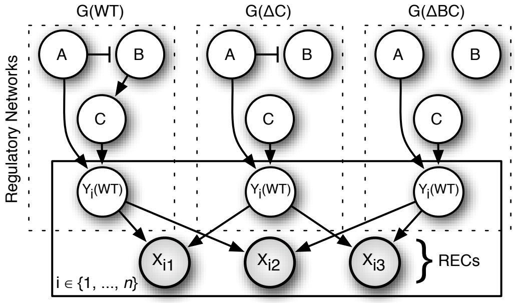

Figure 2. A JIN structure, $J(G)$, for the candidate network G from Fig. 1A, consists of a set of regulatory network instances $G(g)$ connected to the expression state of the indicator genes, Y, and connections from hidden genotype-specific expression $Y(g)$ to X observations representing RECs. A, B, and C are the perturbed signaling genes and the set of $Y_{n=17}$ represents the expression of the indicator genes. The JIN represents three different genotypes, WT (wild-type strain), ΔC (single knock-out strain), and ΔBC (double knock-out strain), and models observations of all possible comparisons. The solid box indicates plate notation to indicate replication of variables and dependencies.

both X and g correspond to a comparison of two of the p phenotypes. Thus, for each indicator i, $x_{ij} = \log(y_i(g_{1j})/y_i(g_{2j}))$, where $y_i(g)$ represents the expression of indicator gene i under genotype g. To form as many RECs as possible, X is augmented with virtual ratios. If a pair of genotypes does not have an observed REC, but there are comparisons to a common reference genotype (i.e. columns j and k where $g_{2j} = g_{2k}$), a new column h was constructed so that $x_{ih} = x_{ij} - x_{ik}$ for all such j and k. The new column(s) represent the pairwise comparison between genotypes g_{1j} and g_{1k} that may be important for detecting epistasis.

We first search for a high-scoring regulatory model by scoring network candidates and then use the highest-scoring candidate to find other genes that are under the same regulation model (Fig. 1C, "Regulatory Network Search" and "Pathway Expansion" respectively). We use a score that reflects how well G explains the RECs recorded in X (either observed or virtual). To score G, a Bayesian Network (BN), $J(G)$, is first constructed, which we refer to as the Joint Intervention Network of G (Figure 2). The log-posterior of the JIN is used as the score for G (see Supp. Methods "Regulatory Network Search"[17]). $J(G)$ is constructed from a candidate network by creating instances of G for each genotype g, referred to as a component network, $G(g)$. For each genotype g, the component network $G(g)$ encodes how the underlying network G is modified under perturbation, i.e. disconnecting perturbed nodes from their parents. Nodes associated with each

$G(g)$ are shown in dashed boxes in Figure 2. Conditional probability tables are shared between component networks. G contains a hidden binary variable for every regulator as well as every indicator gene, representing either *active* or *inactive* gene activity, and connections representing regulatory influences from regulator to regulator states and from regulator to indicator gene states (Y nodes).

In our setting, the state of the indicator genes, $Y = Y_1, Y_2, \dots Y_n$, are forced to be descendants of the signaling genes since these genes serve as the phenotypic output of signaling regulation. A node, X_{ij}, represents the *discretized* REC of the i^{th} indicator gene under the j^{th} genotype comparison of g_{1j} and g_{2j}. The node X_{ij} is connected to two parents in the BN: $Y_i(g_{1j})$ and $Y_i(g_{2j})$, where $Y_i(g_{1j})$ and $Y_i(g_{2j})$ are binary random variables representing the unobserved expression state of gene i (either 0 "inactive" or 1 "active") under perturbations g_{1j} and g_{2j} respectively. A REC node X_{ij} is a ternary random variable and was set to +1 ("higher") if $x_{ij} > \tau$, -1 ("lower") if $x_{ij} < -\tau$, or 0 ("equal") otherwise. We used $\tau = 0.3$ to provide a conservative estimate of equal expression between two genotypes[4]. With the discrete values from the nodes in the JIN, we use a deterministic conditional probability table for X_{ij}: $\mathrm{P}(X_{ij}|Y_i(g_{1j}), Y_i(g_{2j})) = 1$ iff $X_{ij} = Y_i(g_{1j}) - Y_i(g_{2j})$.

The parameters of $J(G)$, θ, that define how regulators influence gene activity are restricted to encode biological intuitions about influences on gene expression. Each edge in the regulatory network can be either repressing or activating, with activation broken into two classes: *necessary* activators, and a pool of *alternative* activators. (see Supp. Methods "Regulatory Model"[17]).

3. Results

The biofilm regulatory pathway involves a complex interplay of transcription factors HapR, VpsT, and VpsR (see Fig. 3A). These regulators control the expression of *Vibrio* polysaccharide (VPS) genes (*vps*-I and *vps*-II gene clusters) that allow a community of *Vibrio* cells to change its surface properties by modulating cell-matrix and cell-surface contacts. Dual channel microarray gene expression data were obtained from Beyhan *et al.* 2007[4]. Briefly, the dataset consisted of a combination of single, double and triple knock-outs of three biofilm regulatory genes: *hapR*, *vpsR* and *vpsT*, representing eight different genotypes including wild type. A competitive microarray hybridization was performed using RNA isolated from a dele-

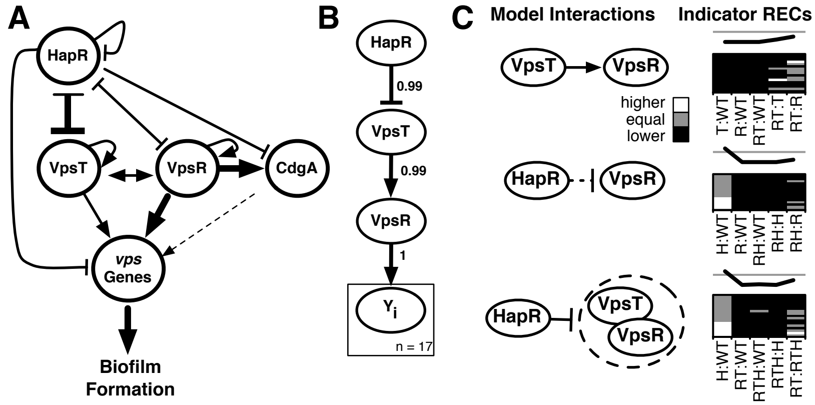

Figure 3. **A. Known biofilm regulation of *V. cholerae.*** HapR, VpsR, and VpsT influence *vps* gene expression. VpsT and VpsR activate the *vps* genes, while HapR represses the expression of *vps* genes. Line thickness indicates magnitude or strength of regulation. **B. Biofilm pathway predicted by JIN.** The regulatory structure learned by JIN. Links are annotated by their feature score (see Results). The solid box indicates plate notation. **C. Predicted interactions and supporting RECs** For three interactions in the predicted regulatory model, we show related discretized RECs. The dashed line from HapR to VpsR indicates an indirect interaction. The dashed circle around VpsT and VpsR indicates treatment as a single genetic unit. The deletions for each REC are coded with H=Δ*hapR*, R=Δ*vpsR*, and T=Δ *vpsT*.

tion mutant and that of wild-type. Six replicates were performed for each of seven deletion strains. Virtual hybridizations (see Methods) were added between compatible deletion strains, resulting in all 28 possible comparisons of strains.

Using known *vps* indicator genes, we used JIN scoring to infer a signaling network among the three regulators. The likelihood of the *vps* data under the learned regulatory network was found to be significant compared to the likelihood of arbitrary, non-*vps* genes. Bayesian feature scoring[11] was used to rank predicted interactions by a confidence measure (see Supp. Methods “Bayesian Feature Scoring”[17]). In addition, the inferred regulatory network was used to implicate new members of the *V. cholerae* biofilm pathway. Several of these genes have been tested and validated for involvement in biofilm formation.

3.1. *Inference of the V. cholerae Biofilm Regulatory Network*

A high-scoring biofilm regulatory network of VpsR, VpsT and HapR was found using known downstream gene clusters *vps*-I (VC0917–VC0927) and *vps*-II (VC0934–VC0939). Their relative expression under each deletion strain compared to wild-type is shown in Figure 3C. Each gene's expression was treated as an independent observation of the biofilm phenotype. A greedy hill-climb with random restarts was used to sample biofilm regulatory networks. We ran 1000 independent hill climbs, each starting from a different G and/or with different parameter values of $J(G)$ (see Supp. Methods "Regulatory Network Search"[17]).

Figure 3B shows the inferred *vps* regulatory network with edges annotated with feature scores. We calculated a feature score[11] for each edge as the proportion of final networks found by hill climbing that contained the edge, weighted by their likelihoods. Three edges in the network were associated with high feature scores.

Consistent with JIN's predictions, two of the three edges are well documented. HapR was found to inhibit VpsT, and VpsR was found to have the most direct influence on the *vps* genes. Several of the relations predicted by the JIN method are consistent with the observed RECs (Figure 3C). The RECs contrasting Δ*vpsT,vpsR* to Δ*vpsR* reveal more "equal" observations than the RECs contrasting Δ*vpsT,vpsR* to Δ*vpsT*, indicating that the regulatory influence from VpsT to VpsR is stronger than the direction from VpsR to VpsT. The model predicts an indirect influence of HapR on VpsR (via VpsT). This prediction is supported by the presence of more "equal" Δ*hapR,vpsR* to Δ*vpsR* RECs compared to the Δ*hapR,vpsR* to Δ*hapR* RECs. If VpsT and VpsR are considered as a single genetic unit, their epistatic relationship to HapR is evident from the triple knock-outs. The Δ*vpsR,vpsT,hapR* to Δ*vpsR,vpsT* RECs have more "equal" observations than the Δ*vpsR,vpsT,hapR* to Δ*hapR* RECs.

3.2. *Significance of the learned network*

To assess the significance of the *vps* network inferred by JIN, we asked how well the expression of *vps* genes fit the model compared to "unrelated" genes. A log-likelihood ratio (LLR) for each gene i was calculated from its vector of data, X_i as $log(P(X_i|J(G),\theta)/P(X_i|Null))$, where the null model was formed by disconnecting all Y_i's from the regulator genes. An LLR for all of the *vps* genes and for genes whose expression was not correlated with

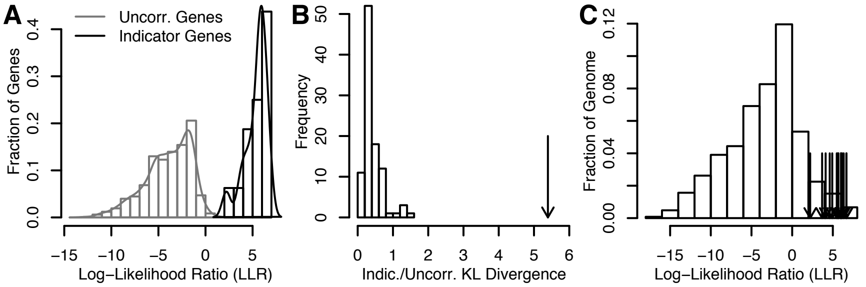

Figure 4. **A. Separation of LLR.** Histograms of the LLRs of indicator genes and of genes uncorrelated to the indicator genes. **B. Significance of learned regulatory network.** Histogram of KL-divergences from sets of randomly drawn indicator genes. The arrow indicates the separation shown in **A**. **C. LLR of entire genome.** Histogram of LLRs from entire genome. Arrows indicated the LLR of the indicator genes.

the *vps* genes was calculated. Uncorrelated genes were selected as those with absolute Pearson correlation to the *vps* genes' mean centroid less than 0.2. The LLRs of the *vps* genes were all above zero, while nearly all of the uncorrelated genes had an LLR less than zero, showing nearly no overlap with the LLRs of uncorrelated genes. The distribution of the *vps* LLRs and the uncorrelated LLRs is shown in Figure 4A.

To quantify the difference between these distributions, we calculated the Kullback-Leibler divergence. We found that the distributions had a KL divergence of 5.4 bits and next asked if this divergence was significant. KL divergences of randomly selected indicators were much lower than the *vps* indicators (see Figure 4B) yielding an empirical P value of 3×10^{-12}. [a]

3.3. *Expansion of the vps Pathway*

We ranked all of the genes by the LLR scores assigned by the JIN *vps* model and chose the top 15 for further analysis. The majority of these top-scoring genes have evidence supporting their association with the biofilm pathway (gray rows in Table 1).

Three genes, *exeA* (VC2445), VC0483, and VC0930, contain a VpsR binding site in their promoter[15], indicating that they are under direct regulation of the furthest downstream biofilm regulator. Additionally, VC0930, along with chromosomal neighbors VC0931 and VC0933 which lie between the *vps*-I and *vps*-II clusters, were recently shown to modulate biofilm

[a] P value computed from a Gamma fit to the empirical distribution.

Table 1. Top predictions for new *vps* pathway members.

LLR	Corr. Rank	Locus	Name	Description
7.14	8	VC2445	*exeA*	general secretion pathway protein A
7.03	34	VC1888	*bap1*	biofilm-associated protein
6.83	116	VC2732	*epsE*	general secretion pathway protein E
6.77	99	VC2730	*epsG*	general secretion pathway protein G
6.77	270	VCA0570		Sui1 family protein
6.74	69	VC0483		hypothetical protein
6.73	287	VC1064		lipoprotein-related protein
6.69	114	VCA0612	*mscL*	large-conductance mechanosensitive channel
6.67	86	VC0930	*rbmC*	rugosity-biofilm modulator
6.67	30	VC0931	*rbmD*	rugosity-biofilm modulator
6.67	12	VC1701		hypothetical protein
6.62	67	VC1320		DNA-binding response regulator
6.51	62	VC1935		CDP-diacylglycerol–glycerol-3-phosphate 3-phosphatidyltransferase-related protein
6.48	133	VC1195		lipoprotein, putative
6.48	9	VC0933	*rbmF*	rugosity-biofilm modulator

formation[6]. VC0930, along with its paralog VC1888, also a top JIN biofilm candidate, are secreted proteins critical to pellicle and biofilm formation.

Several other genes are involved in secretion and are predicted to be *vps* pathway members. VC2730–VC2733, of which JIN predicts *epsE* (VC2732) and *epsG* (VC2730) as top biofilm candidates, are involved in secretion of VPS[1]. *prtV* (rank 18, not shown in Table 1) is a protease that has recently been found in the extra-cellular matrix (Fong and Yildiz, in preparation).

Several regulators were also found. VC1320 negatively regulates biofilm formation in the smooth variant of *V. cholerae* (submitted, Bilecen and Yildiz). *rpoN*, ranked at 25 (not shown in Table 1), has been shown to positively regulate biofilm formation[15]. Other genes predicted are candidates for cell surface sensing and/or modification. *mscL*, predicted to be a high-conductance mechanosensitive channel, may sense surface contact, which is one of the first steps in biofilm formation. Finally, VC1701 and VC1935 have been shown to be differentially regulated between the $\Delta vpsR$ and $\Delta vpsT$ genotypes, indicating that they may play a role in modulating biofilm formation.

We compared the learned regulatory network to that predicted by ModuleNetworks[13] from the same data (see Supplemental File 1[17]). ModuleNetworks only found HapR and VpsR to be significant regulators, with P values of 4.10E-03 and 1.12E-10, respectively. Additionally, robustness analysis revealed that ModuleNetworks used HapR and VpsT as regulators in less than 25% of the trials.

JIN had a higher precision for identifying new candidates than ranking by correlation, which simulates a cluster-based approach. The top 25 candidates were selected from both methods, and genes common to both lists were removed, resulting in 17 genes in each list. Seven genes unique to JIN's list were known to be related to biofilm formation while only three genes unique to the correlation-based list were related.

4. Conclusion and Discussion

The method presented uses a novel probabilistic structure, the Joint Intervention Network, for explicitly linking different genotypes to their phenotypic consequences. It models epistasis analysis by using relative expression changes, rather than direct expression levels, as the quantitative phenotype. The pathway reconstruction method differs from previous approaches, such as NEMs defined by Markowetz *et al.*, in that it uses the downstream effects of knock-out combinations of two or more genes and expression changes of indicator genes already known to be relevant to the pathway rather than the expression of genes selected through data preprocessing.

JIN successfully predicted several genes involved in *V. cholerae* biofilm formation, one of the pathogen's survival mechanisms. With respect to VpsT and VpsR, our model predicts a linear, rather than convergent regulation of the *vps* genes, and the validity of this needs to be tested. Prior studies have found a VpsR binding site in the promoter of *vpsT*, which would suggest VpsR acts upstream of VpsT and not downstream as predicted by our model. The JIN model therefore predicts VpsT regulates VpsR by a different, currently unknown mechanism, and that this regulation has greater strength than VpsR's regulation on VpsT. No transcription factor binding sites have been identified for VpsT. Therefore, it would be interesting to identify targets of VpsT (e.g. through chromatin-immunoprecipitation) and search for the presence of sequence motifs. Binding sites found in the promoter region of *vpsR* but not in *vps* genes would support JIN's prediction that VpsT directly regulates VpsR.

In this work, we assumed the observed epistatic interactions result from the propagation of a signal according to the state of activity of the gene products. However, different signal propagation mechanisms, such as interactions that depend on substrates or other intermediates, are known to reveal different epistatic relationships[8]. Also, feedback present in the network may significantly influence our ability to recover its structure, as may be the case for the predicted interaction between VpsT and VpsR. Thus, further development of computational approaches for ordering genes into pathways using high-dimensional phenotypes, which can incorporate more general models of epistasis than presented here, may hold promise for application to a wide variety of genotype-phenotype investigations to accelerate our understanding of various disease-related processes.

5. Acknowledgments

JMS was supported by NSF grant DBI-0543197 and a fellowship from the Alfred P. Sloan foundation. CHY was supported by the Leon Levy Foundation and the Simons Foundation at the IAS, Princeton, NJ.

References

1. A. Ali, J. Johnson, A. Franco, D. Metzger, T. Connell, J. Morris. JR., S. Sozhamannan, *Infect Immun.* **68:6** 3792 (2000).
2. M. Bansal, V. Belcastro, A. Ambesi-Impiombato, and D. di Bernard, *Molecular Systems Biology* **3:78** (2007).
3. W. Bateson, *Mendel's Principles of Heredity*, Cambridge University Press, Cambridge (1909).
4. S. Beyhan, K. Bilecen, S. Salama, C. Casper-Lindley, and F. Yildiz, *J Bacteriol.*, **189** 388 (2007).
5. D. Eaton, K. Murphy, *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, (2007).
6. J. Fong, F. Yildiz, *J Bacteriol.*, **189:6** 2319 (2007).
7. N. Friedman, M. Linial, I. Nachman, and D. Pe'er, *Journal of Computational Biology.* **7**, 601 (2000).
8. L.S. Huang and P.W. Sternberg, *WormBook*, **14:1-19** (2006).
9. F. Markowetz and R. Spang, *BMC Bioinformatics* **8** (2007).
10. J. Pearl, *Causality: Models, Reasoning, and Inference*, Cambridge University Press, Cambridge (2000).
11. D. Pe'er, A. Regev, G. Elidan, and N. Friedman, *Bioinformatics* **17S1** S215 (2001).
12. K. Sachs, O. Perez, D. Pe'er, D. Lauffenburger, and G. Nolan, *Science* **308** 523 (2005).
13. E. Segal, R. Yelensky, and D. Koller, *Bioinformatics* **19** i273–i282 (2003).
14. N. Van Driessche, J. Demsar, E. O. Booth, P. Hill, P. Juvan, B. Zupan, A. Kuspa, and G. Shaulsky, *Nature Genetics* **37:5** (2005).
15. F. Yildiz, X. Liu, A. Heydorn, G. Schoolnik, *Molecular Microbiology*, **53** 497 (2004).
16. J. Yu, V.A. Smith, P.P. Wang, A.J. Hartemink, E.D. Jarvis, *Bioinformatics* **20:18** 3594 (2004).
17. http://sysbio.soe.ucsc.edu/projects/knockoutnets