Reconstructing regulatory networks from the
dynamic plasticity of gene expression by mutual
information

Jianxin Wang1,2,3, Bo Chen2, Yaqun Wang3, Ningtao Wang3, Marc Garbey4,
Roger Tran-Son-Tay5, Scott A. Berceli6 and Rongling Wu1,3,*

1Center for Computational Biology, Beijing Forestry University, Beijing 100083, China, 2School of Information,
Beijing Forestry University, Beijing 100083, China, 3Center for Statistical Genetics, The Pennsylvania State
University, Hershey, PA 17033, USA, 4Department of Computer Science, University of Houston, Houston,
TX 77204, USA, 5Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville,
FL 32610, USA and 6Department of Surgery, University of Florida, Gainesville, FL 32610, USA

Received January 6, 2013; Revised February 13, 2013; Accepted February 14, 2013

ABSTRACT

The capacity of an organism to respond to its environ-
ment is facilitated by the environmentally induced
alteration of gene and protein expression, i.e. expression plasticity. The reconstruction of gene
regulatory networks based on expression plasticity
can gain not only new insights into the causality of
transcriptional and cellular processes but also the
complex regulatory mechanisms that underlie biol-
ogical function and adaptation. We describe an
approach for network inference by integrating
expression plasticity into Shannon’s mutual infor-
mation. Beyond Pearson correlation, mutual inform-
ation can capture non-linear dependencies and
 topology sparseness. The approach measures the
network of dependencies of genes expressed in
different environments, allowing the environ-
ment-induced plasticity of gene dependencies to be
tested in unprecedented details. The approach is
also able to characterize the extent to which the
same genes trigger different amounts of expression
in response to environmental changes. We
demonstrated the usefulness of this approach
through analysing gene expression data from a
rabbit vein graft study that includes two distinct
blood flow environments. The proposed approach
provides a powerful tool for the modelling and
analysis of dynamic regulatory networks using
gene expression data from distinct environments.

INTRODUCTION

Network analysis using gene expression data has been
widely applied as an approach to studying the regulatory
causality of transcriptional processes involved in cell
survival and proliferation (1–4). In responding to
changes in environmental conditions, a functional cell
would modify the expression of particular genes through
signalling regulation to make it possible to preserve the
robustness of cellular processes (5). A comprehensive
characterization of regulatory networks behind such an
environment-induced response becomes essential in
studying how cells adapt and survive under non-ideal con-
ditions. However, current strategies for network construc-
tion from gene expression data in a single environment
are inadequate for our understanding of the complex regul-
atory mechanisms that underlie biological adaptation
and function. Furthermore, the static feature of these
strategies assumes that genes are expressed in a steady
state, making it infeasible to describe the dynamic
patterns of an evolving process (6).

The purpose of this article is to develop a computational
model for constructing regulatory networks of dynamic
gene expression in response to environmental changes.
The difference of expression for the same gene between
different environments is called expression plasticity
(7,8). As a new concept, expression plasticity has emerged
to be useful for studying the constraints for the evolution of
gene expression in fluctuating environments (9–11). Our
model for network construction capitalizes on gene expres-
sion plasticity, aimed at gleaning a better insight into the
regulatory mechanisms for an organism’s adaptation to
environmental changes. The model is founded on mutual information, a quantity that measures the mutual dependence of the two random variables, particularly in terms of positive, negative and non-linear correlations (12).

The approach for gene expression analysis with mutual information is not entirely new. Michaels et al. (13) attempted to cluster dynamic gene expression profiles according to information theory. Butte and Kohane (14) computed pair-wise mutual information for the expression of all genes using a method of discretizing variable domains. Steuer et al. (1) described the basic theory of mutual information and pioneered its usage to detect dependencies of different genes. Priness et al. (15) compared the properties of different methods for clustering gene expression profiles based on mutual information and classic Euclidean distance and Pearson correlation measures. A path consistency algorithm has been developed to reconstruct gene regulatory networks based on conditional mutual information (16). There are several applications of information-theoretic approaches for network reconstruction in a mammalian cellular context (17) and Escherichia coli transcriptional studies (18). Meyer et al. (19) packed mutual information into an R package minet for inferring transcriptional networks from microarray data. Rajapakse et al. (20,21) used information theory to reconstructing gene regulatory networks during the differentiation of a multipotential haematopoietic progenitor.

Despite these developments, the use of mutual information to reconstruct regulatory networks based on the environment-induced plasticity of time-series expression profiles has not been explored. The model presented in this article will take advantage of mutual information in measuring the non-linear dependency of different variables to unravel the dynamic changes of network architecture in a response to the environment. The model was used to analyse experimental gene expression data obtained from rabbit vein grafts exposed to two different wall shear conditions, where these different environments resulted in two distinct adaptation phenotypes (22,23). The model has been validated through a simulation study. By extending the model to reconstruct a web of mutual relationships among genes and the target phenotype, it provides a useful tool for inferring the causality of gene regulation.

**MUTUAL INFORMATION**

Shannon (24) provided a mathematical theory of measuring the amount of uncertainty and quantifying the theoretical maximum capacity of information by a communication system to eliminate such uncertainty. This theory, called information theory, has been widely applied in a variety of fields. In what follows, we implement Shannon’s information theory to reconstruct a regulatory network with gene expression plasticity data.

**Expression plasticity entropy**

In mutual information, we view gene expression profiles as a discrete random variable. Suppose that expression profiles of genome-wide transcriptional genes are measured at the same series of time points for the same organism that receives two different treatments. For a particular gene, the difference of its time-dependent expression curve between the two treatments describes the pattern of how this gene responds to the change in the treatment’s environment. Wang et al. (23) have developed a dynamic model for clustering genes into distinct groups based on the temporal patterns of their expression profiles in a relation to specific biological functions.

Our model being developed here is to construct a regulatory network of these genes in terms of their dynamic relationships formed in response to environmental change. Mutual information allows the non-linear dependence among different genes to be characterized. We define the difference of the expression value of the same gene at the same time point between the two environments as the expression plasticity of this gene (7,8). Let \( \Delta X \) denote the time-dependent expression plasticity variable of a gene at time points \( \{1,\ldots,T\} \), expressed as \( \Delta X = \{\Delta X_1, \ldots, \Delta X_T\} \).

Suppose that \( D \) is the value range of \( \Delta X \), and the subinterval set \( \{D_i\} \), \( j = 1, 2, \ldots, M \), is a partition of \( D \), satisfying that \( \cup_j D_j = D \), and \( D_j \cap D_k = \emptyset \) if \( j \neq k \). Note that \( M \) is the number of subsections partitioned from the domain \( D \). For convenience, we denote the partition \( \{D_j\} \) simply as \( D \). Define the delta function as follows,

\[
\delta(\Delta x_i, D_j) = \begin{cases} 
1, & \text{if } \Delta x_i \in D_j, \\
0, & \text{else},
\end{cases}
\]

where \( i = 1, 2, \ldots, T \), and \( j = 1, 2, \ldots, M \). Then the probability of \( D_j \) according to the expression plasticity variable \( \Delta X \) is defined as,

\[
p_{\Delta X}(D_j) = \frac{1}{T} \sum_{i=1}^{T} \delta(\Delta x_i, D_j), \quad j = 1, 2, \ldots, M.
\]

Based on the probability defined above, in accordance with Shannon (24), the entropy of \( \Delta X \) with a given partition \( D \) is defined as

\[
H^D(\Delta X) = - \sum_{j=1}^{M} p_{\Delta X}(D_j) \log p_{\Delta X}(D_j), \quad (1)
\]

where the bottom of the logarithmic function, usually 2 or \( e \), could be any positive number without changing the properties of entropy. In this article, we will use 2, in accordance with the definition based on bits by Shannon. If \( p_{\Delta X}(D_j) = 0 \) in Equation (1), the expression \( p_{\Delta X}(D_j) \log p_{\Delta X}(D_j) \) is mathematically undefined. But it can be redefined to be its limit 0 when \( p_{\Delta X}(D_j) \) goes to 0 from the right side of 0.

According to Faser and Swinney (25), when the measurement is expressed as \( \Delta X \), we can describe the expression plasticity entropy \( H^D(\Delta X) \) as the degree of surprise, i.e. the elimination of uncertainty about \( \Delta X \). Information entropy has many properties, several of which are listed as follows:

(i) The entropy \( H^D(\Delta X) \) reaches its minimum 0 if the expression plasticity \( \Delta X \) as a variable is determined, i.e. \( \Delta X \) is no longer random. In this case, the
probability of one element in \( \{D_1, \ldots, D_M\} \) is 1 and that of each of the other elements is 0. (ii) If \( \{D_1, \ldots, D_M\} \) are equiprobable, then entropy \( H^D(\Delta X) \) is maximized to the value \( \log M \). In this case, the entropy \( H^D(\Delta X) \) is the most uncertain; i.e. \( H^D(\Delta X) \) is the hardest to predict.

Conditional entropy of expression plasticity

Analogously to the delta function defined earlier in the text, we can also define joint-delta function as follows.

\[
\delta(\Delta x_i, \Delta y_j, D_j, D_k) = \begin{cases} 1, & \text{if } \Delta x_i \in D_j \text{ and } \Delta y_j \in D_k \\ 0, & \text{else} \end{cases}
\]

where \( i = 1, 2, \ldots, T \), and \( j, k = 1, 2, \ldots, M \). Then the joint probability and the conditional probability of \( \{D_j, D_k\} \) according to the expression plasticity variable \( \Delta X \) and \( \Delta Y \) are defined as follows, respectively.

\[
p_{\Delta X, \Delta Y}(D_j, D_k) = \frac{1}{T} \sum_{t=1}^{T} \delta(\Delta x_i, \Delta y_j, D_j, D_k), j, k = 1, 2, \ldots, M.
\]

\[
p_{\Delta X|\Delta Y}(D_j, D_k) = \frac{1}{\sum_{i=1}^{T} \delta(\Delta y_i, D_k)} \sum_{i=1}^{T} \delta(\Delta x_i, \Delta y_j, D_j, D_k), j, k = 1, 2, \ldots, M.
\]

According to information theory (24), we can calculate the conditional entropy of the expression plasticity of one gene \( \Delta X \), given the expression plasticity of another gene \( \Delta Y \) with time-series values \( \{\Delta y_1, \ldots, \Delta y_T\} \), which is defined as

\[
H^D(\Delta X|\Delta Y) = -\sum_{j=1}^{M} \sum_{k=1}^{M} p_{\Delta X, \Delta Y}(D_j, D_k) \log p_{\Delta X|\Delta Y}(D_j, D_k),
\]

where \( H^D(\Delta X|\Delta Y) \) is the conditional entropy measuring the remaining uncertainty of \( \Delta X \) if \( \Delta Y \) is determined, which has the following property.

\[
H^D(\Delta X|\Delta Y) \leq H^D(\Delta X).
\]

If \( \Delta X \) and \( \Delta Y \) are statistically independent of each other, we have

\[
H^D(\Delta X|\Delta Y) = H^D(\Delta X).
\]

Joint entropy of expression plasticity

The joint entropy of expression plasticity for the two genes, \( H^D(\Delta X, \Delta Y) \), is defined, analogously to \( H^D(\Delta X) \), as

\[
H^D(\Delta X, \Delta Y) = -\sum_{j=1}^{M} \sum_{k=1}^{M} p_{\Delta X, \Delta Y}(D_j, D_k) \log p_{\Delta X, \Delta Y}(D_j, D_k),
\]

where \( p_{\Delta X, \Delta Y}(D_j, D_k) \) is defined based on the expression plasticity variables \( \Delta x \) and \( \Delta y \) for the two genes. The joint entropy is not greater than the sum of the entropies of two expression plasticity variables, i.e.

\[
H^D(\Delta X, \Delta Y) \leq H^D(\Delta X) + H^D(\Delta Y).
\]

If \( \Delta X \) and \( \Delta Y \) are statistically independent, we have

\[
H^D(\Delta X, \Delta Y) = H^D(\Delta X) + H^D(\Delta Y).
\]

The relationship among the entropy, conditional entropy and joint entropy is expressed as

\[
H^D(\Delta X, \Delta Y) = H^D(\Delta X|\Delta Y) + H^D(\Delta Y).
\]

Equation (8) implies that the uncertainty of the joint system (\( \Delta X, \Delta Y \)) is the uncertainty of \( \Delta Y \), plus the conditional uncertainty of \( \Delta X \) given \( \Delta Y \).

Mutual information

The mutual information between two variables of expression plasticity \( \Delta X \) and \( \Delta Y \) according to a domain partition \( D \) is defined as

\[
I^D(\Delta X, \Delta Y) = H^D(\Delta X) + H^D(\Delta Y) - H^D(\Delta X, \Delta Y).
\]

From Equation (6), we have

\[
I^D(\Delta X, \Delta Y) \geq 0.
\]

Furthermore, from Equation (7), we obtain the conclusion that, if \( \Delta X \) and \( \Delta Y \) are statistically independent, their mutual information is 0.

Mutual information is symmetrical, i.e.

\[
I^D(\Delta X, \Delta Y) = I^D(\Delta Y, \Delta X).
\]

In sum, mutual information shown by Equation (9) measures the dependency between the expression plasticity of two arbitrary genes, no matter the dependency is linear or non-linear.

DISCRETIZATION

To apply mutual information of expression plasticity, the random variable domain must first be partitioned into discrete bins. Butte and Kohane (14) used a straightforward method of evenly dividing a domain interval into a certain number of sub-intervals and then approximating the probabilities by the corresponding relative frequencies of occurrence. The mutual information by this approach depends much on the distribution type of the expression plasticity variables and the distribution parameters. Schreiber and Schmitz (26) proposed an adaptive partitioning method. Per this method, each resultant sub-interval for a random variable contains approximately equal number of occurrences. This method is more precise than the straightforward one in finding the mutual information. In Supplementary Text S1, we illustrate the procedure of bin characterization by these two methods.

The two methods described earlier in the text may not produce ideal results when the variable distribution types are the same but the distribution parameters are different. This is common, especially for gene expression data. To improve the calculation of mutual information by
Schreiber and Schmitz's (26) method, we partition the domains of the two random variables of expression plasticity under consideration according to a common standard, while simultaneously making the intervals adaptive to the respective data. We call this process 'common adaptive partitioning'. Let $\Delta x_i$ and $\Delta y_i$ denote two random variables of expression plasticity measured at time $t$ and $t'$, respectively, expressed as

$$
(\Delta x_i, \Delta y_i), t, t' = 1, \ldots, T.
$$

whose means are denoted as $\langle \mu x, \mu y \rangle$ and standard deviations denoted as $\langle \sigma x, \sigma y \rangle$. Suppose $\lbrace q_0, q_1, \ldots, q_T, q_{T+1} \rbrace$ is a sequence of real numbers, $q_0 = -\infty$, $q_{T+1} = \infty$ and $q_i < q_{i+1}$ for $1 < t < \tau - 1$. Except for the two infinities, the other $\tau$ parameters are to be determined later, which are denoted as $\Theta = \lbrace q_1, q_2, \ldots, q_t \rbrace$.

The domains of $\Delta X$ and $\Delta Y$ are partitioned by a transformation of the sequence into the following intervals, expressed, respectively, as

$$
\eta_i^X = \left( \mu x + \sigma x q_{t-1}, \mu x + \sigma x q_t \right), \quad t = 1, \ldots, \tau, \tau + 1
$$

$$
\eta_i^Y = \left( \mu y + \sigma y q_{t-1}, \mu y + \sigma y q_t \right), \quad t = 1, \ldots, \tau, \tau + 1
$$

Let $k^X$, $k^Y$ and $k^XY$ denote the numbers of time-dependent expression plasticity values from Equation (12) located in the $rth$ interval of $X$, in the $tth$ interval of $Y$ and in the $rth$ interval of $X$ while simultaneously in the $tth$ interval of $Y$, respectively.

Our purpose is to select an optimal parameter set described in Equation (13) that makes the time-dependent expression plasticity values from Equation (12) located in the $rth$ interval of $X$, in the $tth$ interval of $Y$ while simultaneously in the $rth$ interval of $Y$, respectively.

Our purpose is to select an optimal parameter set described in Equation (13) that makes the time-dependent expression plasticity values from Equation (12) located in the $rth$ interval of $X$, in the $tth$ interval of $Y$ while simultaneously in the $rth$ interval of $Y$, respectively.

Our purpose is to select an optimal parameter set described in Equation (13) that makes the time-dependent expression plasticity values from Equation (12) located in the $rth$ interval of $X$, in the $tth$ interval of $Y$ while simultaneously in the $rth$ interval of $Y$, respectively.

Our purpose is to select an optimal parameter set described in Equation (13) that makes the time-dependent expression plasticity values from Equation (12) located in the $rth$ interval of $X$, in the $tth$ interval of $Y$ while simultaneously in the $rth$ interval of $Y$, respectively.

Our purpose is to select an optimal parameter set described in Equation (13) that makes the time-dependent expression plasticity values from Equation (12) located in the $rth$ interval of $X$, in the $tth$ interval of $Y$ while simultaneously in the $rth$ interval of $Y$, respectively.

Our purpose is to select an optimal parameter set described in Equation (13) that makes the time-dependent expression plasticity values from Equation (12) located in the $rth$ interval of $X$, in the $tth$ interval of $Y$ while simultaneously in the $rth$ interval of $Y$, respectively.

Our purpose is to select an optimal parameter set described in Equation (13) that makes the time-dependent expression plasticity values from Equation (12) located in the $rth$ interval of $X$, in the $tth$ interval of $Y$ while simultaneously in the $rth$ interval of $Y$, respectively.

The mutual information between the expression profiles in the two environments can be compared for the extent of such dependence. Accumulating all genes within groups, different groups can be used to study the dependence of the expression of individual genes between different environments. By accumulating all genes within groups, different groups can be compared for the extent of such dependence.

### Working example

In previous work by Wang et al. (23), a dynamic model was developed and used to identify unique groups of genes based on their differential response to the local environment. Specifically, vein bypass grafts, exposed to either...
high or low flow, were harvested at 2 h, 1 day, 7 days or 28 days after implantation (20). Microarray analysis of 14,958 genes was used to define and cluster the temporal response of the transcriptional profile induced by the local flow environment. Wang et al.'s model identified eight groups, symbolized by A (0.0116), B (0.0123), C (0.3535), D (0.3831), E (0.1134), F (0.0359), G (0.0100) and H (0.0083), where the numbers in parentheses are the proportions of genes belonging to a particular group. These groups display different patterns of environment-induced changes in gene expression trajectories. Our mutual information approach was applied to reconstruct a regulatory network that encompassed the dynamics of gene expression. Our analysis was based on three scenarios: (i) reconstructing an overall network by jointly using time-series gene expression data from the two flow environments; (ii) reconstructing a network by using the expression plasticity between high and low flows; and (iii) reconstructing two networks by using time-series gene expression data separately for two flows.

According to scenario (i), a sparse network of gene expression was obtained (Figure 1), in which a few pairs of gene groups have regulatory connections. Of all pair-wise relationships, group A shares the highest level of mutual information with group H, followed by the level of mutual information between groups H and F, groups B and E, groups A and F, groups B and F and so forth. Several pairs of groups share very low mutual information. It appears that groups C, D and G are substantially dissimilar to the rest of the groups, with each of these groups only weakly connected with two other clusters.

Scenario (ii) emphasizes the similarities of gene groups in terms of their pattern of differential expression over two different flows. Figure 2a provides a quantitative description of the level of regulatory connections among eight gene groups identified by Wang et al.'s (23) dynamic model. Although many connections are observed, the levels of mutual information are highly variable. To respond to environmental changes from one flow to next, groups A and H, groups H and F, and groups F and D would adjust their expression profiles in a highly similar way. As such, we conclude that groups A, H, F and D share substantial overlapping information, compared with other clusters in the network. The significant overlap and network autonomy among these four groups is further underscored by the configuration of group D, which, except for weak connections with groups C and B, only demonstrates the dominant connection to group F.

To reconstruct the networks per scenario (iii), we calculated with the common partition $D$ the mutual information of expression dynamics of genes $X_j$ and $Y_j$ from two groups $G_1$ and $G_2$, respectively, in a particular environment $j$ using

$$I_j^D(G_1, G_2) = \frac{1}{|G_1| \cdot |G_2|} \sum_{X_j \in G_1} \sum_{Y_j \in G_2} I^D(X_j, Y_j).$$

Figure 1. An overall regulatory network of eight groups of rabbit genes constructed by jointly using expression data from high and low flows.

Figure 2. Regulay network of eight groups of rabbit genes expressed in high and low flows. (a) Between-group network reconstruction based on average value of expression between the two treatments. (b) Within- and between-group network reconstruction based on gene expression separately in low and high flows. The thickness of a circle represents the level of mutual information between two treatments within a group, whereas the thickness of lines represents the level of mutual information between two groups in low (green) and high flows (red). The lines representing mutual information below the average level are omitted.
This equation was used to calculate group-wise dependence separately for each different environment. It is interesting to see that the degree of dependence between groups is not identical for low and high flows (Figure 2b). For example, groups A is associated with groups C and D in the low flow, but this association does not occur in the high flow. Figure 2b provides a quantitative measure of the difference in the level of group-wise mutual information of gene expression between the two treatments. In addition, the amount of mutual information of the two treatments for the same group varies, depending on group type. Group G is most highly associated between low and high flows, followed by groups H, E and F. Within group D, the two flows are weakly associated. The results given in Figure 2 provide a comprehensive characterization of regulatory networks of genes related to vein graft remodeling, which are expressed differently in response to low and high flow environments.

**Computer simulation**

The basic idea of using mutual information to reconstruct networks for genes expressed in a single environment has been available in the literature. Some studies critically analysed the advantages of information-based approaches over those based on classic Euclidean distance and Pearson correlation measures (1,15). Thus, we will not focus on methodological comparisons in this article, rather than on the investigation of the advantage of our information-based approach in studying gene expression plasticity.

We simulated two data sets each of three equally sized groups of genes expressed in a time course. In the first data set, genes are measured in a single environment, whereas the second data set contains genes measured in two different treatments. Our model was used to analyse these two sets of data, having results to be in a good agreement with the actual case of each data set (Figure 3). However, it is impossible that a good result can be obtained for gene expression in two environments using a traditional single-environment approach.

**DISCUSSION**

Many biological processes including plant and animal development are coordinated by cell-to-cell communication regulated by genes (5). High-throughput measurement techniques have now led to the identification of tens of thousands of genes involved in sensing external cues. However, the dynamic interplay between genes is highly complex and cannot be understood by a simple approach (28). The reconstruction of gene regulatory networks can be a valuable tool for identifying the key mechanisms that shape the dynamics of cellular and transcriptional processes (6,29).

External stimuli or agents can alter the speed and direction of cellular processes through differential expression of the gene set. There exist specific mechanisms that shepherd the signal into the nucleus, where signal integration occurs by complex transcription factor networks. In this article, we describe a procedure for quantitative modelling of biological regulatory networks regulated by gene expression using mutual information. Beyond classic correlation parameters, mutual information can measure and evaluate the non-linear dependencies of random variables (12,14,24). We extended this information-based approach to assess and detect the non-linear dependencies of genes both between and within different gene groups of a particular function.

Our model has combined two complexities of network reconstruction. First, although much previous work focuses on static (steady-state) gene regulation, improved biotechnologies have allowed the measures of dynamic gene expression data during a biological process. The availability of dynamic data enables geneticists to better study the regulatory machineries underlying cellular processes (2) but, meanwhile, brings about a difficulty in analysing and interpreting expression data. Second, as gene expression is environment dependent (5), the reconstruction of regulatory networks by integrating environmental impact is crucial. By taking into account dynamic and environment-dependent complexities of gene expression, our model allows the reconstruction of more mechanistic and, therefore, more powerful regulatory networks.

The new model based on mutual information can effectively handle any dynamic relationships of genes, linear or non-linear, a characteristic better than classic Euclidean distance and Pearson correlation measures (15), and thereby should be able to find its broader application in computational biology. The model was used to analyse a time-series data set of gene expression measured for vein bypass grafts subjected to two distinct conditions, high and low blood flow, leading to the construction of genetic network that connects different groups of genes with different response trajectories to the local environment (23). The model can quantify the mutual dynamic
relationships of different genes in terms of their differential expression to environmental change. The model was validated through computer simulation, showing its practical usefulness. In practice, when the number of genes is large, some inference procedure for selecting important groups, such as some permutation procedure, may be helpful and can be implemented.

There is much room for the model to be improved. First, our model assumes a normal distribution of gene expression, which is reasonable for microarray data. However, an increasing body of expression data is being collected by high-throughput cDNA sequencing (RNA-Seq). The current model will need to be modified to accommodate the feature of Poisson distribution, which characterizes the data obtained from RNA-Seq (30). Second, the ultimate goal of network construction is to identify key genes or elements that can determine or alter the behaviour of an outcome, such as the critical stenosis that leads to vein bypass graft failure. Thus, the incorporation of outcome variables into the network and the estimation of direct or indirect effects of each gene on the outcome are essential for mechanistic characterization. Third, it is likely that the regulation of gene elements is under global genetic control (31). The integration of mutual information into genetic mapping will provide a powerful means of identifying expression quantitative trait loci that control regulatory networks. The characterization of expression quantitative trait loci will enable geneticists to gain a better understanding of the etiology underlying complex traits or diseases.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online: Supplementary Texts 1 and 2.

ACKNOWLEDGEMENTS
The contents of the Article are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

FUNDING
Funding for open access charge: [NSF/IOS-0923975, UL1 TR000127] from the National Center for Advancing Translational Sciences (NCATS) and NIH [R01-HL095508].

Conflict of interest statement. None declared.

REFERENCES
1. Steuer, R., Kthurts, J., Daub, C.O., Weise, J. and Selbig, J. (2002) The mutual information: detecting and evaluating dependencies between variables. *Bioinformatics*, 2, 231–240.
2. Luscombe, N.M., Babu, M.M., Yu, H., Snyder, M., Teichmann, S.A. and Gerstein, M. (2004) Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature*, 431, 308–312.
3. Legewie, S., Herzel, H., Westerhoff, H.V. and Blüthgen, N. (2008) Recurrent design patterns in the feedback regulation of the mammalian signaling network. *Mol. Syst. Biol.*, 4, 190.
4. Bleda, M., Medina, I., Alonso, R., De Maria, A., Salavert, F. and Dopazo, J. (2012) Inferring the regulatory network behind a gene expression experiment. *Nucleic Acids Res.*, 40, W168–W172.
5. Chen, D., Toone, W.M., Mata, J., Lyne, R., Burns, G., Kivinen, K., Brazma, A., Jones, N. and Bahler, J. (2003) Global transcriptional responses of fission yeast to environmental stress *Mol. Biol. Cell.*, 14, 214–229.
6. Zhu, H.L., Rao, R.S.P., Zeng, T. and Chen, L.N. (2012) Reconstructing dynamic gene regulatory networks from sample-based transcriptional data. *Nucleic Acids Res.*, 40, 10657–10667.
7. Ihmels, J., Friedlander, G., Bergmann, S., Sarig, O., Ziv, Y. and Barkai, N. (2002) Revealing modular organization in the yeast transcriptional network. *Nat. Genet.*, 31, 370–377.
8. Muers, M. (2011) Noise versus plasticity. *Nat. Rev. Genet.*, 12, 4.
9. Lehner, B. (2010) Conflict between noise and plasticity in yeast. *PLoS Genet.*, 6, e1001185.
10. Yampolsky, L.Y., Glazko, G.V. and Fry, J.D. (2012) Evolution of gene expression and expression plasticity in long-term experimental populations of Drosophila melanogaster maintained under constant and variable ethanol stress. *Mol. Ecol.*, 21, 4287–4299.
11. Evans, T.G. and Hofmann, G.E. (2012) Defining the limits of physiological plasticity: how gene expression can assess and predict the consequences of ocean change. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 367, 1733–1745.
12. Cover, T.M. and Thomas, J.A. (1991) *Elements of Information Theory*, New York, Wiley.
13. Machaels, G., Carr, J., Askenazi, M., Fuhrman, S., Wen, X. and Somogyi, R. (1998) Cluster analysis and data visualization of large scale gene expression data. *Pac. Symp. Biocomput.*, 2, 42–53.
14. Butte, A. and Kohane, I.S. (2000) Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pac. Symp. Biocomput.*, 5, 415–426.
15. Priness, I., Maimon, O. and Ben-Gal, I. (2007) Evaluation of gene-expression clustering via mutual information distance measure. * BMC Bioinformatics*, 8, 111.
16. Zhang, X.J., Zhao, X.M., He, K., Lu, L., Cao, Y., Liu, J., Hao, J.K., Liu, Z.P. and Chen, L.N. (2012) Inferring gene regulatory networks from gene expression data by PC-algorithm based on conditional mutual information. *Bioinformatics*, 28, 98–104.
17. Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R. and Califano, A. (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, 7, 7.
18. Faith, J.J., Hayete, B., Thaden, J.T., Mogno, I., Wierzbowski, J., Cottarel, G., Kasif, S., Collins, J.J. and Gardner, T.S. (2007) Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biol.*, 5, 54–66.
19. Meyer, P.E., Lafitte, F. and Bontempi, G. (2008) minet: a R/Bioconductor package for inferring large transcriptional networks using mutual information. *BMC Bioinformatics*, 9, 461.
20. Rajapakse, I., Perlmutter, M.D., Scalzo, D., Kooperberg, C., Grouin, M. and Kosak, S.T. (2009) The emergence of lineage-specific chromosomal topologies from coordinate gene regulation. *Proc. Natl Acad. Sci. USA*, 106, 6679–6684.
21. Rajapakse, I., Grouin, M. and Mesbah, M. (2012) What can systems theory of networks offer to biology? *PLoS Comput. Biol.*, 8, e1002543.
22. Jiang, Z., Wu, L., Miller, B.L., Goldman, D.R., Fernandez, C.M., Abouhamze, Z.S., Ozaki, C.K. and Berceli, S.A. (2004) A novel vein graft model: adaptation to differential flow environments. *Am. J. Physiol. Heart Circ. Physiol.*, 286, H240–H245.
23. Wang, Y., Xu, M., Wang, Z., Tao, M., Zhu, J., Wang, L., Li, R., Berceli, S.A. and Wu, R.L. (2012) How to cluster gene expression dynamics in response to environmental signals. *Brief Bioinformatics*, 13, 162–174.
24. Shannon, C.E. (1948) A mathematical theory of communication. *Bell Sys. Tech. J.*, 27, 379–423.
25. Faster, A.M. and Swinney, H.L. (1986) Independent coordinates for strange attractors from mutual information. *Phys. Rev. A.*, 33, 2318–2321.
26. Schreiber, T. and Schmitz, A. (2000) Surrogate time series. *Physica D.*, 142, 346–382.
27. D'haeseleer, P. (2005) How does gene expression clustering work? *Nat. Biotech.*, **23**, 1499–1501.
28. Sivriyer, J., Habib, N. and Friedman, N. (2011) An integrative clustering and modeling algorithm for dynamical gene expression data. *Bioinformatics*, **27**, i392–i400.
29. Hecker, M., Lambeck, S., Toepfer, S., van Someren, E. and Guthke, R. (2009) Gene regulatory network inference: data integration in dynamic models—a review. *BioSystems*, **96**, 86–103.

30. Huang, W., Umbach, D.M., Vincent Jordan, N., Abell, A.N., Johnson, G.L. and Li, L. (2011) Efficiently identifying genome-wide changes with next-generation sequencing data. *Nucleic Acids Res.*, **39**, e130.
31. Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S. *et al.* (2008) Genetics of gene expression and its effect on disease. *Nature*, **452**, 423–428.