

Published in IET Systems Biology
 Received on 8th January 2010
 Revised on 7th July 2010
 doi: 10.1049/iet-syb.2010.0009

Special issue on the Third *q-bio* Conference on Cellular Information Processing



Multivariate dependence and genetic networks inference

A.A. Margolin^{1,2} K. Wang^{2,3} A. Califano² I. Nemenman^{2,4}

¹Cancer Program, The Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

²Joint Centers for Systems Biology, Columbia University Medical Center, New York, NY 10032, USA

³Oncology Research Unit, Pfizer, San Diego, CA 92121, USA

⁴Departments of Physics and Biology, Computational and Life Sciences Strategic Initiative, Emory University, Atlanta, GA 30322, USA

E-mail: ilya.nemenman@emory.edu

Abstract: A critical task in systems biology is the identification of genes that interact to control cellular processes by transcriptional activation of a set of target genes. Many methods have been developed that use statistical correlations in high-throughput data sets to infer such interactions. However, cellular pathways are highly cooperative, often requiring the joint effect of many molecules. Few methods have been proposed to explicitly identify such higher-order interactions, partially due to the fact that the notion of multivariate statistical dependence itself remains imprecisely defined. The authors define the concept of dependence among multiple variables using maximum entropy techniques and introduce computational tests for their identification. Synthetic network results reveal that this procedure uncovers dependencies even in undersampled regimes, when the joint probability distribution cannot be reliably estimated. Analysis of microarray data from human B cells reveals that third-order statistics, but not second-order ones, uncover relationships between genes that interact in a pathway to cooperatively regulate a common set of targets.

1 Introduction

Reverse engineering molecular interaction networks is a critical challenge in modern systems biology [1]. High-throughput technologies allow simultaneous measurements of the concentrations of thousands of molecular species in a biological system, such as mRNA [2], microRNA [3], proteins [4] and metabolites [5]. Each such experiment may be treated as an observation from a joint probability distribution (JPD), and it is believed that statistical dependencies in this JPD provide clues about biochemical interactions among the species [6]. Thus, identifying dependencies in JPDs is an essential task for network reverse engineering, and this problem is also ubiquitous in other branches of systems biology [7–9] and in many other applications.

It is clearly understood [10] that statistical dependencies can be characterised by their order (i.e. by the number of

variables – molecular species – participating in them). Until recently, most network reverse engineering work focused on second-order (pairwise) dependencies. Their identification from data is now a common exercise. In particular, direct (irreducible) interactions can be disambiguated from indirect ones (e.g. two biochemical species correlated due to a common regulator) [11, 12]. However, combinatorial regulation [13], where multiple effectors combine to regulate a target gene, is prevalent in higher eukaryotes. Correspondingly, recent years have seen a surge in the use of high-throughput data to identify these higher-order structures [6, 14–19]. However, as described below, there has been little work to rigorously define the mathematical basis of the identified multivariate statistical dependencies and the structure of uncovered interactions (e.g. cooperative or independent regulation). For example, consider two transcription factors, TF_1 and TF_2 , that may regulate the expression of a target gene, T , in different ways, including, but not limited to (note that we use roman

characters to denote gene names and italic ones for gene expressions)

$$\frac{dT}{dt} = b_1(TF_1) + b_2(TF_2) - rT \quad (1)$$

$$\frac{dT}{dt} = b_1(TF_1)b_2(TF_2) - rT \quad (2)$$

Here b_i 's are single-effector activation terms, such as Hill functions, and rT is the first-order degradation. The first of these equations describes independent activation of the target. In the second equation, both transcription factors (TFs) act synergistically, for example, due to formation of a transcriptional complex. This type of dependency also applies in the case of a signalling molecule (SM) that post-translationally modifies a TF, influencing its ability to regulate the target. We expect T to be statistically dependent on TF_1 and TF_2 in both cases; however, clearly, there is a difference, since for (2) the effects of TF_1 and TF_2 on T cannot be studied in isolation from each other, forming a third-order dependency among the variables. With (1) and (2) infused with the usual Gaussian noise with variance σ^2 , the resulting steady-state equations are, respectively

$$P(T|TF_1, TF_2) \propto e^{-(1/2r^2\sigma^2)[T-b_1(TF_1)-b_2(TF_2)]^2} \quad (3)$$

$$P(T|TF_1, TF_2) \propto e^{-(1/2r^2\sigma^2)[T-b_1(TF_1)b_2(TF_2)]^2} \quad (4)$$

Thus, joint regulation involves a term that couples all three variables in the exponent of the JPD. A reasonable tool for statistical analysis of multivariate interaction patterns should distinguish such high-order structures from additive pairwise interactions, as in (1).

This is a non-trivial task since, even now, there is no consensus definition of an interaction in the multivariate setting. For example, standard statistical methods [20, 21] introduce many specialised dependence concepts applicable in restricted contexts, such as normal noise, binary, bivariate, or metric data, and so on. Alternatively, contingency tables literature associates interactions with deviations of the number of observed counts from their expectations under various independence assumptions [22–24]. Unfortunately, this is limited to categorical data and confounds the definition of dependence with sampling issues. In information theory [25, 26], continuous and categorical data can be treated uniformly [27, 28] and dependencies are defined based on distributions rather than counts, but none of the information theoretic interaction measures [10, 18, 19, 29–34] have become universally accepted either.

In the context of systems biology, multivariate dependencies have been studied traditionally [6] using probabilistic graphical models [35], such as Bayesian networks (BNs) or Markov networks, also known as

Markov random fields (MRFs). However, these models are generically unable to disambiguate different types of regulation, such as in (1) and (2) [36]. This limitation arises from relying on the notion of conditional (in)dependence rather than providing a precise definition of statistical dependency among subsets of variables (see below for more details). That is, many different interaction patterns can give rise to the same conditional independence structure in an MRF.

In this work, we provide a rigorous definition of multivariate statistical dependence. Section 2 motivates our definition by extending recently developed information theoretic concepts, such as the definition of connected interactions proposed by Schneidman *et al.* [10]. We refer the readers to [10] for additional background reading on some of the concepts used in our work. In Section 3, we describe the connection between our definition of multivariate dependence and commonly used graphical models. We use this framework to suggest how local tests on triplets of variables can be used in the context of genetic network inference to identify proteins that cooperatively interact in a pathway to control the expression of a target gene. We use a simple synthetic model in Section 4 to demonstrate the method's ability to infer interactions even for undersampled distributions, which are common in practical applications. In Section 5, we apply the method to identify combinatorially regulated pathways using a biological data set from human B cells. To promote the suitability of the method, we design a computationally efficient proxy test for our definition of multivariate dependence that provides a theoretical basis for methods used in previous work [14, 37]. This approach enables identification of bona-fide regulatory cofactors that modulate the activity of the MYC proto-oncogene.

2 Definition of multivariate dependence

Information theoretic concepts have successfully provided precise, and often unique, definitions of various types of statistical dependence, but have not yet converged on a unique definition of multivariate statistical dependence among subsets of variables within a group of covariates. In this section, we review recently developed information theoretic concepts, such as mutual, multi and connected information, and demonstrate how each is a special case of the principle of evaluating the Kullback–Leibler (KL) divergence [38] between the maximum entropy distributions constrained by a set of marginals and a subset of these marginals. We then generalise this concept to provide a precise definition of multivariate statistical dependence.

For two variables, X_1 and X_2 , independence is well defined via decomposition of the bivariate JPD, $P(X_1, X_2) = P(X_1)P(X_2)$, and mutual information $I(X_1; X_2) \equiv I_p(X_1; X_2) \equiv \langle \log_2 P(X_1, X_2) / [P(X_1)P(X_2)] \rangle$ is the unique

measure of dependence [26]. Similarly, the total interaction (i.e. the deviation from independence) in a multivariate JPD, $P(\{X_i\})$, $i = 1, \dots, N$, can be measured by the multi-information [33]

$$I_P(X_1; \dots; X_N) \equiv D_{KL}[P||P^*] = \left\langle \log_2 \frac{P(X_1, \dots, X_N)}{P(X_1) \dots P(X_N)} \right\rangle_P \quad (5)$$

which assigns a specific number of bits to the union of all interactions among the studied variables. Here D_{KL} is the KL divergence between the full JPD, $P(X_1, \dots, X_N)$, and its approximation under the independence assumption, $P^*(X_1, \dots, X_N) = P(X_1) \dots P(X_N)$. In order to define multivariate statistical dependence, we seek to partition the total deviation from independence into contributions from interactions among various variable subsets (specific pairs, triplets etc.), and a non-zero contribution from a subset would indicate an interaction among its members.

We first note that P^* is the maximum entropy (MaxEnt) distribution [39, 40] that has the same univariate marginals as P but introduces no statistical dependencies among the variables [10, 32, 41]. Thus, the multi-information is the KL divergence between the JPD and its MaxEnt approximation with univariate marginal constraints, and it measures the gain in information by knowing the complete JPD against assuming total independence. Similarly, MaxEnt distributions consistent with various multivariate marginals of the JPD introduce no statistical interactions beyond those in the said marginals. Thus, by comparing the JPD to its MaxEnt approximations under various marginal constraints, one can hope to separate dependencies included in the low-order statistics from those not present in them [32, 41–44].

Specifically, one can define connected interactions of a given order, that is, the interactions that need, at least, the full set of marginals of this order to be captured. Following Schneidman *et al.* [10], suppose that we have a network of N variables and we know a set of marginal distributions of all variable subsets of size $k \geq 1$, so that $P(X_{i_1}, \dots, X_{i_k}) = \sum_{X \notin \bigcup_{j=1}^k X_{i_j}} P(\{X\})$ is specified. One can ask what is the JPD $P^{(k)}$ that captures all multivariate interactions prescribed by these marginals, but introduces no additional dependencies. That is, one searches for a distribution $P^{(k)}$ with a minimum $I_{P^{(k)}}$ (or, alternatively, with the maximum entropy – MaxEnt – $H_{P^{(k)}}$) such that the constraints $C_{i_1, \dots, i_k}(P^{(k)}, P) \equiv P^{(k)}(X_{i_1}, \dots, X_{i_k}) - P(X_{i_1}, \dots, X_{i_k}) = 0$ are satisfied. This is given by the MaxEnt, or minimum multi-information, problem [10, 39, 41]

$$P^{(k)} \equiv \arg \max_{P' \in \Lambda} \left\{ H[P'] - \sum_{M \in \Omega} \lambda_M \cdot C_M(P', P) \right\} \quad (6)$$

where M 's are sets of constrained variables, such as

$M_{i_1, \dots, i_k} = \{X_{i_1}, \dots, X_{i_k}\}$ and $\Omega = \cup M$. Further, λ s are the Lagrange multipliers that enforce the marginal constraints. They are matrices of the same dimensionality as the constraints they enforce, and “ \cdot ” denotes element-by-element matrix multiplication (following Matlab notation). We do not write out the indices of JPDs or λ s explicitly.

The solution of a MaxEnt problem with marginal constraints has the form of a product of terms dependent on the constrained variables [45]. In particular, for (6)

$$P^{(m)} = \frac{1}{Z} \prod_{i_1 < \dots < i_m} \psi(M_{i_1 \dots i_m}), \quad \psi \geq 0 \quad (7)$$

where Z is the normalisation term and each ψ is a different function, known as a potential, which is determined implicitly by the marginal constraints. In general, no analytical solution for the ψ 's exists. However, an algorithm called the iterative proportional fitting procedure (IPFP) [46], which iteratively adjusts a trial solution to satisfy each of the constraints in turn, converges to the true solution [45]. The connected information of order k is then

$$I_C^{(k)}(X_1; \dots; X_N) \equiv D[P^{(k)}||P^{(k-1)}] = I_{P^{(k-1)}}(X_1; \dots; X_N) - I_{P^{(k)}}(X_1; \dots; X_N) \quad (8)$$

This characterises the increase in information by knowing all marginals of order k , as opposed to all marginals of order $k - 1$. Note that the multi-information can be decomposed into a series of connected information, $I_P = \sum_{k=2}^N I_C^{(k)}$.

While appealing, the connected interaction construction assigns interaction bits to a particular interaction order. We need to refine the approach to instead assign the bits to a particular combination of variables within this order, which has not yet been done.

To localise (connected) interactions to particular sets of covariates, we note that the definitions provided in this section illustrate that mutual, multi and connected information all rely on the general principle of evaluating the KL divergence between the MaxEnt distributions constrained by a set of marginals and a subset of these marginals (or, alternatively, the difference of entropies of these two MaxEnt distributions or the negative difference of the multi-information). If the divergence is positive, then the extra marginal constraints correspond to a non-zero interaction. Thus, to determine if interactions within a particular set V of variables contributes to I_P , we may check if enforcing the corresponding constraint C_V recovers any additional dependencies not already contained in a reference MaxEnt distribution, $P^{(\Omega)}$, constrained by some set of other marginal constraints in Ω . That is, we define

the interaction information

$$\Delta_{\Omega}^{(V)} \equiv \left\langle \log_2 \frac{P^{(\Omega \cup V)}}{P^{(\Omega)}} \right\rangle = I_{P^{(\Omega \cup V)}} - I_{P^{(\Omega)}} \equiv I^{(V)} - I^{*(V)} \quad (9)$$

Here, similar to (6), $P^{(\Omega)}$ is the MaxEnt distribution satisfying all constraints in Ω [45], as in

$$P^{(\Omega)} = \frac{1}{Z} \prod_{M \in \Omega} \psi(M) \equiv \frac{1}{Z} \exp \left[- \sum_{M \in \Omega} \varphi(M) \right] \quad (10)$$

By positivity of the Kullback–Leibler divergence, $\Delta_{\Omega}^{(V)} \geq 0$. Thus if $\Delta_{\Omega}^{(V)} > 0$, accounting for the constraint C_V recovers more information, and we say that the variables in V interact with respect to $P^{(\Omega)}$.

Note that $\Delta_{\Omega}^{(V)}$ is Ω -dependent, and to test for dependencies we must first select the reference set of constrained variables Ω . To define an irreducible interaction among variables in V , we choose Ω that minimises the interaction information

$$\Omega_V = \arg \min_{\Omega} \Delta_{\Omega}^{(V)} \quad (11)$$

$$\Delta^{(V)} \equiv \Delta_{\Omega_V}^{(V)} \quad (12)$$

This guarantees that interactions are defined only if they cannot be explained away by confounding effects of other statistical dependencies in the network. Then, in particular

$$I_P \geq \sum_{V \in \wp(\{X_1, \dots, X_N\})} \Delta^{(V)} \quad (13)$$

where $\wp(\{X_1, \dots, X_N\})$ is the power set (the set of all subsets) of the analysed variables.

Intuitively, adding additional constraints to Ω should not increase the information provided by V , and should only cause $\Delta_{\Omega}^{(V)}$ to decrease or remain the same. Stated formally:

Conjecture 1: Let $\Omega_1 \subset \Omega_2$ be sets of non-contradictory marginal constraints, and $P^{(\Omega_1)}$ and $P^{(\Omega_2)}$ be the corresponding MaxEnt distributions. Let V be an additional marginal constraint, possibly a subset of either Ω_1 or Ω_2 . Then

$$\Delta_{\Omega_1}^{(V)} \geq \Delta_{\Omega_2}^{(V)} \quad (14)$$

This says that interaction information depends on the order in which the interactions are considered. Dependency bits will be accounted for by the first marginal able to explain them, attributing less bits to later constraints. We have extensively tested this conjecture numerically (not shown), but the proof is not yet available.

According to the conjecture, the reference set of constraints Ω_V to test for the existence of irreducible interactions within V is

$$\Omega_V = \bigcup_{M \subset \wp(\{X_1, \dots, X_N\}), M \not\supset V} M \quad (15)$$

Thus, $P^{(\Omega_V)}$ preserves all marginals of the original JPD except those that involve all covariates in V simultaneously. This is similar to the Type III sum of squares analysis of variance for the testing significance of predictors. In fact, since D_{KL} is equal to χ^2 asymptotically, the similarity is not accidental. Dependence defined by this choice of $P^{(\Omega_V)}$ is a generalisation of conditional dependence with the rest of the network as a condition. This extends the analysis of Schneidman *et al.* [10] and defines an interaction among a particular set of variables, rather than within all variable subsets of the same cardinality.

While this formulation gives a precise definition of multivariate statistical dependence, computational issues arise in applying it to large networks. First, searching through the space of all possible multivariate dependencies is exponential in the number of variables as, for N variables, there are 2^N possible subsets of the variables. Moreover, each test for an irreducible interaction

$$\Delta^{(V)} > 0 \quad (16)$$

requires computing two large MaxEnt distributions, which is not trivial, especially since empirical distributions computed for large-dimensional marginals will be severely undersampled. Finally, in many cases, some of the variables in the network will be unmeasurable (hidden), influencing the interaction structure derivable from the visible variables [10, 47]; this is clearly prevalent in systems biology applications, where we are still far from measuring the concentrations of all chemical species in a cell. We will address these issues partially in Section 3.

Complications aside, the MaxEnt formulation resolves the problem of disambiguating dynamics arising from different dependency structures, such as in (1) and (2). Indeed, independent regulation, as in (1), produces a JPD with only pairwise potentials, while joint regulation requires a third-order potential and will, therefore, result in a third-order interaction.

3 Graphical models

Graphical models [35] are widely used to provide a visual representation of the factorisation of a JPD and to motivate efficient inference algorithms based on graph theoretic considerations. This framework has been applied often in genetic network inference applications [6]. Thus, it is informative to illustrate statistical interactions that can be identified using our approach, but not by commonly used graphical models, as we describe below and in Section 3.1.

Moreover, in cases where computation of irreducible dependencies is intractable, the graphical models framework provides insights into the interpretation of local dependency tests, as we describe in Section 3.2.

The maximum entropy formulation is strongly related to undirected graphical models. In particular, (10) has the form of a Markov network, which is visually represented by associating each variable with a node and drawing an edge between each pair of variables that appear together in a potential. However, this network representation is insufficient to distinguish between potentials that are fully parameterised, or only parameterised by functions on subsets of variables, which is a major goal of this work. A more general graphical model, known as a factor graph, is able to represent this distinction. The factor graph representation of a JPD contains two types of nodes. Each factor (potential) $\psi(V)$ is explicitly represented as a factor node, with an edge connected to each variable in V , which are represented as variable nodes (Fig. 1). However, in traditional factor graph literature, the factors cannot be defined uniquely once the JPD is known. For example, if a three-variable factor $\psi(X_1, X_2, X_3)$ is present, then any two-variable factor $\psi(X_i, X_j)$, $i, j = 1, 2, 3$, can be subsumed into it. Therefore traditional factor graphs blur the distinction between columns (d) and (e) in Fig. 1. Conversely, while our formulation can also be represented using factor graph notation, each factor is defined uniquely, so that the factor structure of the JPDs in columns (d) and (e) is materially different. Therefore one can talk about existence or non-existence of a lower-order factor uniquely and independently of whether the higher-order factor involving the same variables exists.

3.1 Examples and properties

We consider a few examples of different graphical model representations for networks of size $M = 3$ (larger M is

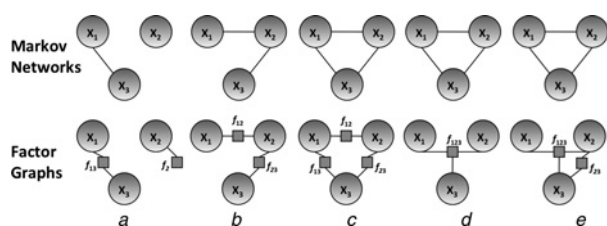


Figure 1 Examples of Markov networks and corresponding factor graphs for three-variable networks

- a A simple second-order interaction
- b Two second-order interactions without loops
- c Three second-order interactions forming a loop
- d A Third-order interaction
- e A third-order and a second-order interaction

In the factor graphs, variable nodes are represented by circles and factor nodes are represented by squares. Note that a three-way interaction, three two-way interactions or combinations of the two are represented by the same Markov network

analysed similarly). First, for a regulatory cascade, or a Markov chain, $X_1 \rightarrow X_2 \rightarrow X_3$, $P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_2)$, as shown in Fig. 1b. Consider the test for X_1X_2 dependence. Following the notation of (9), we let $I^{(12)}$ and $I^{*(12)}$ be the multi-informations of the distributions used to test for dependency in X_1X_2 . That is, $\Delta^{(X_1X_2)} \equiv \Delta^{(12)} = I^{(12)} - I^{*(12)}$. Then, we have $I^{*(12)} = I[X_1, X_3] + I[X_2, X_3] \leq I^{(12)} = I[X_1, X_2] + I[X_2, X_3]$, where the inequality is due to the information processing inequality, and the bound is reached only in special cases. Thus, X_1 and X_2 are (generically) dependent. Similarly, X_2 and X_3 are dependent. However, $\Delta^{(13)} = 0$, and X_1, X_3 are not dependent (even though their marginal mutual information, induced by other interactions, is not zero). Checking for the triplet interactions, we find $I^{*(123)} = I[X_1, X_2] + I[X_2, X_3] = I^{(123)}$, thus no such dependencies are present. If instead X_2 regulates X_1 and X_3 , one sees that the dependence structure is the same. Both networks correspond to the graph in Fig. 1b.

A more interesting case is when X_1 and X_3 regulate X_2 jointly. Here many possibilities exist, not all of them realisable in terms of Bayesian or Markov Network modelling. First, consider independent regulation: to predict X_2 , one does not need to know the values of X_1 and X_3 simultaneously, $P(x_2|x_1, x_3) = f_{12}f_{23}$, for example, $P(x_2|x_1, x_3) \propto \exp[-a(x_2 - x_1)^2 - b(x_2 - x_3)^2]$ (this corresponds to probabilistic analogues of OR and AND gates [10], to the Lac-repressor [13], and to all regulatory models based on independent binding of TFs to the DNA [8]). If $P(X_1X_3) = P(X_1)P(X_3)$, then the dependency structure is again as in Fig. 1b. If, in addition, there is a regulation $X_1 \rightarrow X_3$, so that $P(X_1X_3) \neq P(X_1)P(X_3)$, then $\Delta^{(13)} \geq 0$, and $\Delta^{(123)} = 0$. The dependency graph now has a loop in it, as in Fig. 1c. However, in the case of joint (e.g. cooperative) regulation, $P(X_1X_2X_3)$ is non-factorisable, $\Delta^{(123)} > 0$, and the dependence structure is as in Figs. 1d or e.

3.2 Local tests

While the previous section described precise tests for three-variable networks, computing irreducible statistical dependencies for large networks is computationally intractable. The graphical models framework provides an intuitive interpretation of statistical tests performed on subsets of variables. For example, consider a network with $N \gg 3$ nodes and define $V_3 = \{(1, 2, 3)\}$ and $V_2 = \{(1, 2), (2, 3), (1, 3)\}$. Evaluation of $\Delta^{(V_2)}$ or $\Delta^{(V_3)}$ using (16) is unrealistic since it requires computing MaxEnt distributions with factors over $N - 2$ and $N - 3$ variables. Instead, one may need to marginalise over many X_i , $i > 3$, and search for dependencies in the JPD with three variables only. In general, with marginalised (hidden) variables, an irreducible dependency cannot be inferred by MaxEnt tests, but it is informative to understand the meaning of a difference in MaxEnt entropies even in this case.

Owing to the factor structure of the JPD in (10), marginalising over a variable will couple all of its neighbours (nodes with which it participates in a potential) into a single factor. If any of those nodes are marginalised out, its neighbours will further be coupled into this factor, and so on. As a consequence, for any three variables remaining in a marginalised graph, if, in the full factor graph, there exists a factor node such that there is a direct path between it and each of the remaining three variables that does not pass through the other two, then marginalisation over the hidden variables will produce an effective third-order interaction among the remaining three variables. As discussed in Section 5, this observation has important consequences in genetic network inference and indicates that the proposed multivariate dependency framework can be used to identify proteins that cooperatively interact in a pathway to regulate the expression of a target gene.

4 Synthetic data

A major advantage of our definition of statistical dependencies in terms of MaxEnt approximations is that it can be applied even when the underlying distributions are undersampled and the corresponding factorisations cannot be readily observed. If the cardinality of the sample space,

$K = \prod K_i$, where K_i are cardinalities of individual variables, is larger than the number of samples, s , we cannot estimate the distributions reliably (we note that because continuous gene expression values are routinely discretised in genomics, we focus on the discrete case in view of its conceptual simplicity). However, entropic quantities, and therefore the interactions are inferable. In genomics, continuous expression levels are routinely discretised. Thus, we focus on the discrete case in view of its relevance and conceptual simplicity. Some progress is possible even for $s \sim \sqrt{K}$ [48, 49]. To show this, we used Dirichlet priors [49] to generate random probability distributions with different interaction structures for $N = 3$, and with marginal cardinalities $K_i = 50$. We generated random samples of different sizes, $s = 50, \dots, 125\,000$, from these distributions and tested the quality of inference of the dependencies as a function of s . As an evaluation metric, we used the evidence for an interaction, $E^{(V)} \equiv \Delta^{(V)} / \delta \Delta^{(V)}$, where $\delta \Delta^{(V)}$ is the statistical error of the interaction information. If $E^{(V)}$ is large, the dependency is present. According to Fig. 2, proper recovery is possible for $s = K = K_1 K_2 K_3$ with few assumptions about the distributions.

With modern entropy estimation techniques [49], our approach will work even for severely undersampled JPDs.

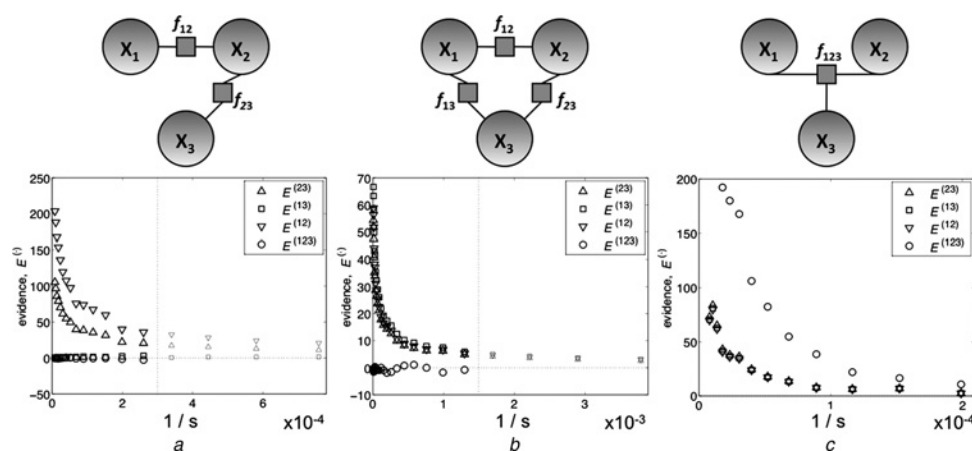


Figure 2 Inferring regulatory networks from sample size, s

a Network with $P \propto \psi(X_1, X_2)\psi(X_2, X_3)$. To the left of the vertical dotted line, $s \approx 3000 > 2^{H_{p(\Omega)}} \ll K \approx 125\,000$, the sample size corrections are reliable, and all entropies are known well. Here, H_p stands for the entropy of the distribution P . There is evidence only for X_1X_2 and X_2X_3 interactions, just as it should be. For smaller s , the method of [50] fails, but the NSB entropy estimation method [49] works until $s \sim 2^{1/2H_p} \approx 60$. For pairwise interactions, we may replace $H_{p(\Omega)}$ by H_p (denoted by smaller markers on the plot) and, since $E^{(13)}$ stays zero nonetheless, and $I(X_1; X_2) + I(X_2; X_3) = I_p$, we still recover the interactions correctly

b Network with three pairwise interactions. Again, to the left of the line, $s > 2^{H_{p(\Omega)}}$, all entropies are determined reliably, and there is evidence for all three pairwise interactions, but not for the triplet interaction. To the right of the line, NSB still works, but now we cannot disentangle the loop from the three-way dependence without estimating $H_{p(\Omega)}$

c Network with three pairwise interactions and a third-order interaction. Only the regime $s > 2^{H_{p(\Omega)}}$ is shown. The evidence for all three pairwise interactions and for the triplet interaction is barely significant for small s but grows fast

We used the NSB method to estimate the entropies (with error bars) of the JPD and its marginals directly. The method does not work for the entropy of $P^{(\Omega)}$ for $\Omega \equiv (123) = \{(X_1, X_2), (X_2, X_3), (X_1, X_3)\}$. Thus, the IPFP [46] was applied to the counts and the entropy $H_{p(\Omega)}$ in the solution was evaluated and extrapolated for $1/s \rightarrow 0$, following [50, 51], to account for the sample-size dependent bias. The statistical error for each sample size, s , was determined by bootstrapping, and the resulting extrapolation error was used for $\delta H_{p(\Omega)}$. This approach works since the MaxEnt constraints, like those in (6), are linear in the unknown JPD, P , making the biases of H_p and $H_{p(\Omega)}$ behave similarly. Finally, $\Delta^{(V)}$ was calculated as the differences of the appropriate entropies, and $\delta^2 \Delta^{(V)}$ as the sums of squares of the entropy errors. Network models are displayed above each plot

The bottleneck is estimation of the maximum entropy consistent with the marginals, which currently requires substantial sampling of the marginals, requiring $s \sim \max(K_1K_2, K_2K_3, K_1K_3)$, similarly to the jackknifing method used in [50, 51]. This is encouraging, since the marginals may be well-sampled when the JPD is not. However, it is still essential to develop techniques to infer maximum entropies directly. Further, the interaction information is the difference of entropies. It may be small when its error, which is a quadratic sum of the entropy errors, is large. This leads to uncertainties about dependencies even for reliably estimated entropies. Therefore a method that directly estimates Δ will be preferred over another entropy-based technique.

5 Genetic networks inference

5.1 Inferring regulatory pathways

The proposed method for identifying multivariate dependencies has important applications for cellular networks inference. Cellular networks are composed of a complex system of interacting and diverse molecular species. For example, consider the task of inferring genetic regulatory interactions using statistical correlations between mRNA concentrations measured using gene expression microarrays. Generically, genes encode mRNAs, which are translated into proteins. Some of the latter encode TFs, which in turn can bind to DNA and influence the expression of other genes. However, mRNA abundance data only probes a small percentage of the regulatory network. For example, the translation of mRNA into protein is dynamically regulated at many levels, including by regulating mRNA stability, nuclear export and cytoplasmic localisation and translation initiation. Once translated, proteins engage in a vast network of interactions, being regulated, for example, by complex formation as well as a variety of post-translational modifications, such as (de)phosphorylation, (de)acetylation and (de)ubiquitination. Finally, the ability of a gene to be transcribed is strongly affected by modifications of the DNA, such as methylation, chromatin accessibility (which is influenced by histone modifications such as acetylation), as well as other genetic factors including mutations, single nucleotide polymorphisms and chromosomal alterations. Many of these regulatory processes are carried out by proteins, but there is also a critical and ever increasingly appreciated role for other regulatory factors such as non-coding RNAs and metabolites. The combined effect of these considerations is to create a vast network of hidden variables, while we only probe a small percentage of the system with current technologies. For such complicated regulatory systems, it is difficult to understand the effect of the unobserved variables and thus to interpret the meaning of statistical dependencies between mRNAs.

Section 3.2 provides some insight into this question and suggests that irreducible multivariate statistical

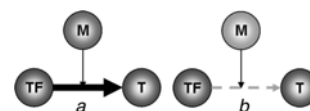


Figure 3 TF regulates a target gene, T

a When M is abundant, TF strongly regulates T
b When M is scarce, TF only weakly regulates T
 Modulator gene, M, influences this interaction, for example, by forming a complex with TF, by phosphorylating TF or by regulating another protein that interacts with TF

dependencies between mRNAs may be used to identify genes that interact in a pathway to jointly regulate the expression of a downstream target. Consider, for example, a TF, that regulates a target gene, T. This interaction is influenced by a (possibly large) number of other proteins, which we call modulators [14], denoted by M (Fig. 3). The modulators may interact directly with TF, for example, via post-transcriptional modification, creating a third-order dependency between TF, M and T. However, as discussed, effective third-order dependencies are also created between variables that interact indirectly, for example, if the modulator regulates another gene or protein that subsequently interacts with TF downstream. This type of series of interactions in which multiple genes jointly control a cellular process (e.g. expression of a target gene) is called a pathway, and is a principle mechanism by which a cell regulates gene expression.

To identify such third-order dependencies, we test for a reduction in entropy by constraining $P(TF, M, T)$ compared to constraining $P(TF, T)$, $P(M, T)$ and $P(TF, M)$. The MaxEnt distribution constrained by all three pairwise marginals must be computed by an iterative algorithm. However, a much more computationally efficient procedure can be derived under the simplifying assumption that TF and M are not (irreducibly) statistically dependent, which is a common occurrence in biology [37]. That is, the factorisation of the JPD produced by the MaxEnt formalism does not contain the $\psi(TF, M)$ potential. Note that this is less stringent than requiring $I(TF; M) = 0$, and only means that we do not need to constrain (TF, M) in the MaxEnt construction. Then the test for the difference in entropy of MaxEnt distributions constrained by $[(TF, T), (M, T)]$ against that constrained by $[(TF, T), (TF, M)]$ reveals a simplified equation based on conditional mutual information. In particular, the MaxEnt distribution constrained by the two pairwise marginals has the form $P^*(TF, T, M) = [(P(TF, T)P(M, T))/P(T)]$, whereas the distribution constrained by the three-way interaction has the form $P^*(TF, T, M) = P(TF, T, M)$. Therefore letting Δ denote the difference in multi-information of the two distributions, we have

$$\begin{aligned} \Delta &= H(TF, T) + H(M, T) - H(T) - H(TF, M, T) \\ &= H(T|M) - H(T|TF, M) + H(TF, T) \\ &\quad + H(M) - H(T) - H(M|TF) - H(TF) \\ &= I(TF; T|M) - I(TF; T) + I(TF; M) \end{aligned} \quad (17)$$

We implemented this form of the equation, considering cases where $I(TF, M) = 0$, ensuring that the simplifying assumption of no statistical interaction between TF and M holds true. This form of the equation was used in [14, 37], but its theoretical basis has not been developed until the current work. This procedure relies on computing whether the mutual information between TF and T increases when conditioned on M under the $I(TF, M) = 0$ assumption. Since expression data are continuous, to overcome the undersampling issue, we use Gaussian kernels to estimate conditional information [12]. Further, following [37], we discretise M into M^+ and M^- , representing high and low-modulator expressions, and test for $I(TF; T|M^+) \neq I(TF; T|M^-)$ as a proxy for (17). Below we consider how this simplified version of the general framework can be used to identify cooperative regulation, and compare it to using pairwise dependencies only. The main contribution of this work is to formalise the concept of multivariate dependence, and thus we do not claim to exhaustively test its application to biological networks, but rather provide initial evidence of the method's effectiveness for this purpose.

5.2 Results for biological networks

The MYC proto-oncogene is a critical regulator of oncogenic onset and progression, and is estimated to be overexpressed in at least one-seventh of all human cancers [52], including a large percentage of B cell lymphomas. The pluripotent nature of MYC's interactions make it difficult to characterise the critical pathways that are affected by aberrant MYC expression, and it is thus important to characterise the network of interactions associated with MYC. In addition, MYC provides a convenient test case for reverse engineering algorithms due to a public database containing a large number of biochemically validated MYC transcriptional targets [53]. Moreover, MYC is known to be regulated by the B cell receptor (BCR) pathway in B cells [54], and has over

60 known protein–protein interaction (PPI) partners in the human protein reference database [55]. Thus, while far from a perfect test, comparing predicted modulators against these two data sources provides a level of validation.

We have recently taken steps towards characterising the genetic network associated with MYC by analysing a data set of 254 microarrays derived from normal and tumour-related human B lymphocyte populations [56]. In particular, we have developed a method [12, 57] that has been used to accurately identify downstream MYC targets [11], and has led to insights into the relationship between the part of the cellular interaction network regulated by MYC, and those regulated by other proto-oncogenes such as NOTCH1 [58]. Further, we have identified a variety of modulators of MYC [14, 37]. In this work, we take another, more principled look at the identification of the cellular network that works cooperatively with MYC to jointly regulate sets of target genes.

After filtering out all genes from the microarray exhibiting low expression or insufficient dynamic range, following [37], we compiled two sets of potential modulator genes. The first, which we call SMs, contains genes that are annotated as protein kinase, protein phosphatase, acetyltransferase or acylase in the gene ontology database, and may potentially post-translationally regulate MYC or another gene that acts in the same pathway as MYC. The second group contains genes with the gene ontology annotation of TF activity, which may serve as co-TFs associated with MYC. We also compiled a set of experimentally validated MYC targets from the www.mycancer.org database [53]. In order to apply (17), we removed potential modulators that had significant MI with MYC, leaving a total of 1128 Affymetrix probe sets as potential modulators (542 SMs and 598 TFs), which were tested for their ability to modulate MYC interactions with the 340 probe sets associated with MYC targets.

Table 1 Third-order, but not pairwise, statistics effectively identified putative modulators

		Putative	Total	Percent, %	p -Value
SMs	all genes	12 580	74 800	16.8	
	inferred (three-way)	1015	3586	28.3	$3.3 \cdot 10^{-11}$
	inferred (pairwise)	432	3586	12.1	
TFs	all genes	9520	87 040	10.9	
	inferred (three-way)	771	4343	17.8	$4.0 \cdot 10^{-11}$
	inferred (pairwise)	380	4343	8.8	

Putative modulators were defined as those contained in the BCR pathway or participating in a PPI with MYC. We considered the percent of putative modulators associated with significant three-way interactions against a background of all tested triplets, as well as triplets with the highest total MI, $I(TF; T) + I(M; T)$. We considered separate statistics for SMs and TFs. Because pairwise statistics identified a lower percent of putative modulators than background, we assessed statistical significance of the third-order interactions against the background. As shown, third-order statistics identified interactions highly enriched in putative MYC modulators, whereas pairwise statistics performed slightly worse than background.

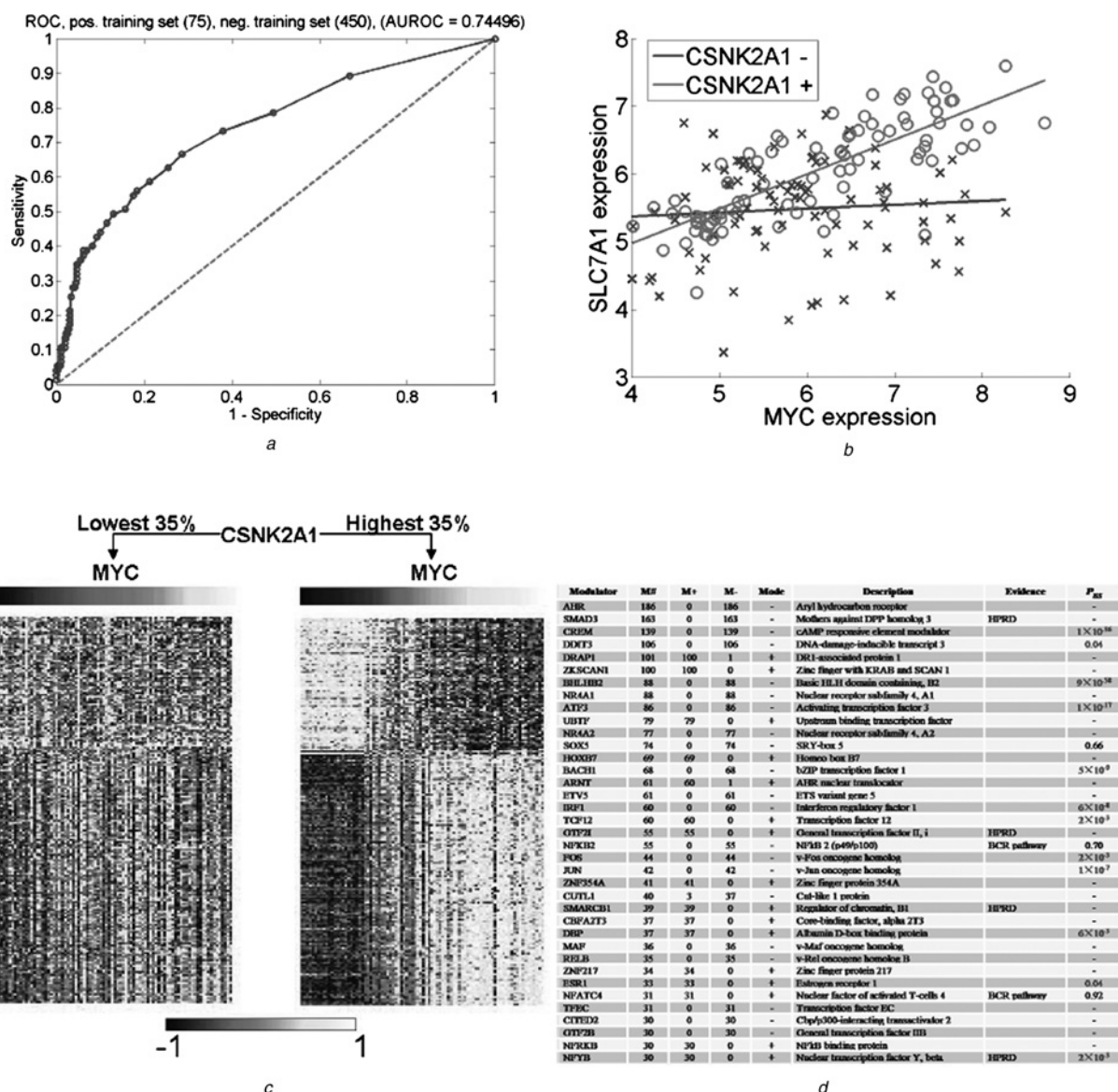


Figure 4 Analysis of third order interactions in the MYC transcriptional network

a Set of 75 putative modulators was compiled, including probe sets from BCR pathway genes and known PPI partners of MYC, together with 450 negative instances, including those not in the positive set and not correlated with any probe sets in the positive set. Each probe set was ranked based on the number of MYC interactions that it modulated. By varying this number as the threshold, an ROC curve was produced. The area under the curve was calculated to be 0.74

b Example of a scatter plot of an interaction modulated by CSNK2A1, a bona-fide positive modulator of MYC. Expression levels (in log) of MYC and SLC7A1 (a known MYC target) were plotted on the X and Y-axes, respectively. Data were partitioned into the 35% of samples with the highest CSNK2A1 expression and the 35% of samples with the lowest CSNK2A1 expression (circles and crosses, respectively), and a line was fitted to the data points in each subset. As shown, when CSNK2A1 was highly expressed, MYC strongly regulated SLC7A1, whereas this interaction was not apparent at low CSNK2A1 expression levels

c MYC target gene expression modulated by CSNK2A1. Two microarray images (modulated MYC target genes on rows and samples on columns) are shown for each subset of high and low CSNK2A1 expression. Samples in each subset were sorted by MYC expression and expression values of target genes were rank transformed, scaled between -1 and 1, and displayed using the colour scheme indicated by the colour bar shown at the bottom of the plot. At high CSNK2A1 expression, MYC was highly correlated with these targets, but not at low CSNK2A1 expression

d TF binding site enrichment analysis for MYC modulators functioning as potential co-TFs. For each modulator with an available scoring matrix in the TRANSFAC database [59], its binding sites were searched for in the promoter regions (2 K upstream and 2 K downstream from transcription start site) of each modulated MYC target gene. Binding site enrichment for each modulator was assessed using Fisher's exact test and comparing to 13 000 random human promoters. M#: number of modulated MYC targets; M+/- -: number of MYC targets positively/negatively affected by the modulators; P_{BS}: p-value of the binding site enrichment test. Twelve of the top 15 inferred co-TF modulators that had available scoring matrices in TRANSFAC displayed statistically significant enrichment of their DNA binding site in the promoters of the inferred target genes

We applied (17) to all combinations of modulators and target genes, with MYC fixed as the TF variable. Statistical significance was assessed using the permutation test described in [37]. This creates a matrix of interaction p -values with all modulators on the columns and all genes on the rows. Significant interactions were defined as those having a Bonferroni corrected p -value less than 0.05.

We sought to test two specific claims made in the preceding papers [14, 37]. First, that third-order statistics can be used to identify genes that interact in a pathway to directly or indirectly cooperate with a TF to control a set of target genes. Second, that such interactions may be identified by third-order statistics but not by second-order ones. To this end, we considered all significant third-order interactions and analysed the number of associated modulators either annotated as belonging to the BCR pathway, or as a PPI partners with MYC. We call genes meeting these criteria putative modulators. Overall, there were 3586 and 4343 significant interactions for the SM and TF data sets, respectively. As shown in Table 1, modulators associated with inferred three-way interactions were significantly enriched with putative modulators. We note that the p -values may be overestimated because samples are not independent.

To test against the hypothesis that modulators can be identified by second-order statistics alone, for each data set we ranked each interaction based on the total pairwise mutual information, $I(TF; T) + I(M; T)$, and, to compare with third-order tests, considered the top-ranking 3586 and 4343 triplets for SMs and TFs, respectively. Only 432 (12.1%) SMs and 380 (8.8%) TFs were putative modulators, indicating that modulators could not be identified using pairwise statistics alone. In fact, the top-ranked interactions based on MI contained a slightly lower percent of putative modulators than the background, likely because the activity of a modulator affects the strength of coupling between the TF and target, diluting the MI. Thus, gene triplets with high MI are likely to preferentially not include third-order interactions.

Next, reasoning that important modulators may affect MYC's interaction with a large number of target genes, we tested whether putative modulators could be identified by ranking them based on the number of MYC interactions that they affect. Using this procedure, we can simultaneously identify the modulators of MYC and the lists of target genes that they modulate. Receiver operating characteristic (ROC) analysis (Fig. 4a) showed that the top-ranking genes by this procedure were significantly enriched for putative modulators. The top-ranking gene, casein kinase 2 alpha 1 (CSNK2A), showed a strong pattern of positive modulation of MYC (Figs. 4b and c), and has been experimentally validated in vivo to directly phosphorylate MYC and positively modulate its DNA binding kinetics [60, 61]. Finally, the binding sites for the top-ranking TF modulators were significantly enriched in

the promoters of inferred target genes (Fig. 4d), providing evidence that these co-TFs cooperate with MYC by binding to the promoters of common targets. Together, these results indicated that our procedure could effectively identify genes that interact in a cellular pathway of interest.

6 Conclusions

In this article, we revisited the concept of multivariate dependence using information theoretic, maximum-entropy-based techniques. We provided a definition of a higher-order statistical interaction that is able to measure the interaction strength, in bits, and assign it to a specific set of statistical covariates. This extends earlier results of Schneidman *et al.* [10], which allowed for identification of the existence of a higher-order interaction, but could not identify which specific variables participated in it.

Clearly, the method is still in the early stages of development. To complete the definition, the conjecture that allowed us to define the interaction information uniquely needs to be proved. Further, for applications, development of techniques for dealing with undersampling for identification of higher-order dependencies is likely the largest obstacle to a wide adoption of the method. Finally, additional testing is required to validate the applicability of the approximate test to various biological data. We will return to all of these questions in future work.

However, in its present form, we believe that our work represents both a theoretical and practical advance. Our formal definition of multivariate dependence will allow more precise consideration, in statistics and information theory disciplines, of irreducible statistical dependencies between subsets of covariates. From a practical standpoint, our definition is only useful to the extent that it can be used in real-world applications. Thus we demonstrated the robust performance of our method even for undersampled distributions that are typical of statistical inference problems. Further, we argued that the definition allows a new, principled treatment of reverse-engineering transcriptional regulatory networks, in particular to identify combinatorially regulated pathways in transcriptional data. To facilitate the application to large-scale data sets, we designed a computationally efficient test that well-approximates our definition of multivariate dependence in typical transcriptional regulation data. The test enabled specific, verifiable and literature-supported predictions about regulatory cofactors, also called modulators, operating together with MYC to regulate its targets. Our results, taken together, suggest that the definition of multivariate dependence introduced in this work provides an important theoretical advance in the field of statistical inference, with applications to systems biology and related disciplines.

7 Acknowledgments

A.A.M. was supported by an IBM PhD fellowship during earlier stages of this work. I.N. acknowledges support from Columbia University, KITP/UCSB, DOE under contract no. DE-AC52-06NA25396 and NSF under grant no. ECS-0425850 during earlier stages of this work.

8 References

- [1] MARGOLIN A.A.: 'Computational inference of genetic networks in human cancer cells' (VDM Verlag, Saarbrücken, Germany, 2009)
- [2] SCHENA M., SHALON D., DAVIS R.W., BROWN P.O.: 'Quantitative monitoring of gene expression patterns with a complementary DNA microarray', *Science*, 1995, **270**, (5235), pp. 467–470
- [3] LU J., GETZ G., MISKA E.A., ET AL.: 'MicroRNA expression profiles classify human cancers', *Nature*, 2005, **435**, (7043), pp. 834–838
- [4] PEREZ O.D., NOLAN G.P.: 'Simultaneous measurement of multiple active kinase states using polychromatic flow cytometry', *Nat. Biotechnol.*, 2002, **20**, (2), pp. 155–162
- [5] LU W., KIMBALL E., RABINOWITZ J.D.: 'A high-performance liquid chromatography-tandem mass spectrometry method for quantitation of nitrogen-containing intracellular metabolites', *J. Am. Soc. Mass. Spectrom.*, 2006, **17**, (1), pp. 37–50
- [6] FRIEDMAN N.: 'Inferring cellular networks using probabilistic graphical models', *Science*, 2004, **303**, (5659), pp. 799–805
- [7] BEER M.A., TAVAZOIE S.: 'Predicting gene expression from sequence', *Cell*, 2004, **117**, (2), pp. 185–198
- [8] BUSSEMAKER H.J., LI H., SIGGIA E.D.: 'Regulatory element detection using correlation with expression', *Nat. Genet.*, 2001, **27**, (2), pp. 167–171
- [9] SLONIM N., ELEMENTO O., TAVAZOIE S.: 'Ab initio genotype-phenotype association reveals intrinsic modularity in genetic networks', *Mol. Syst. Biol.*, 2006, **2**, p. 0005
- [10] SCHNEIDMAN E., STILL S., BERRY M.J. II, BIALEK W.: 'Network information and connected correlations', *Phys. Rev. Lett.*, 2003, **91**, (23), p. 238701
- [11] BASSO K., MARGOLIN A.A., STOLOVITZKY G., KLEIN U., DALLA-FAVERA R., CALIFANO A.: 'Reverse engineering of regulatory networks in human B cells', *Nat. Genet.*, 2005, **37**, (4), pp. 382–390
- [12] MARGOLIN A.A., NEMENMAN I., BASSO K., ET AL.: 'ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context', *BMC Bioinf.*, 2006, **7**, (Suppl 1), p. S7
- [13] BUCHLER N.E., GERLAND U., HWA T.: 'On schemes of combinatorial transcription logic', *Proc. Natl. Acad. Sci. USA*, 2003, **100**, (9), pp. 5136–5141
- [14] WANG K., SAITO M., BISIKIRSKA B.C., ET AL.: 'Genome-wide identification of post-translational modulators of transcription factor activity in human B cells', *Nat. Biotechnol.*, 2009, **27**, (9), pp. 829–839
- [15] DE LA FUENTE A., BING N., HOESCHELE I., MENDES P.: 'Discovery of meaningful associations in genomic data using partial correlation coefficients', *Bioinformatics*, 2004, **20**, (18), pp. 3565–3574
- [16] HARTEMINK A.J., GIFFORD D.K., JAAKKOLA T.S., YOUNG R.A.: 'Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks'. Pacific Symp. Biocomputing, 2001, pp. 422–433
- [17] PE'ER D., REGEV A., ELIDAN G., FRIEDMAN N.: 'Inferring subnetworks from perturbed expression profiles', *Bioinformatics*, 2001, **17**, (Suppl 1), pp. S215–S224
- [18] ANASTASSIOU D.: 'Computational analysis of the synergy among multiple interacting genes', *Mol. Syst. Biol.*, 2007, **3**, p. 83
- [19] WATKINSON J., LIANG K.C., WANG X., ZHENG T., ANASTASSIOU D.: 'Inference of regulatory gene interactions from expression data using three-way mutual information', *Ann. NY Acad. Sci.*, 2009, **1158**, pp. 302–313
- [20] AGRESTI A.: 'Categorical data analysis' (Wiley, New York, 1990)
- [21] JOE H.: 'Multivariate models and dependence concepts' (Boca Raton, 1997)
- [22] DARROCH J.N.: 'Interactions in multi-factor contingency tables', *J. Roy. Stat. Soc. Ser. B (Methodol.)*, 1962, **24**, (1), pp. 251–263
- [23] LANCASTER H.O.: 'Complex contingency tables treated by the partition of chi square', *J. Roy. Stat. Soc. Ser. B (Methodol.)*, 1951, **13**, (2), pp. 242–249
- [24] ROY S.N., KASTENBAUM M.A.: 'On the hypothesis of no "interaction" in a multi-way contingency table', *Ann. Math. Stat.*, 1956, **27**, (3), pp. 749–757
- [25] COVER T.M., THOMAS J.A.: 'Elements of information theory' (Wiley, New York, 1991)

- [26] SHANNON C.E.: 'A mathematical theory of communication', *AT&T Tech. J.*, 1948, **27**, (3), pp. 379–423
- [27] IRELAND C.T., KULLBACK S.: 'Contingency tables with given marginals', *Biometrika*, 1968, **55**, (1), pp. 179–188
- [28] KULLBACK S.: 'Probability densities with given marginals', *Ann. Math. Stat.*, 1968, **39**, (4), pp. 1236–1243
- [29] BELL A.J.: 'Co-information lattice'. Technical report RNI-TR-02–1, 2002
- [30] CHECHICK G., GLOBERSON A., ANDERSON M.J., ET AL.: 'Groups redundancy measures reveal redundancy reduction along the auditory pathway', in DIETTERICH T., BECKER S., GHAHRAMANI Z. (EDS.): 'Advance in Neural Information Processing System' (MIT Press, Cambridge, MA, 2002, vol. 14)
- [31] GARNER W.R., MCGILL W.J.: 'The relation between information and variance analysis', *Psychometrika*, 1956, **21**, (3), pp. 219–228
- [32] GOOD I.J.: 'Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables', *Ann. Math. Stat.*, 1973, **34**, (3), pp. 911–934
- [33] MCGILL W.: 'Multivariate information transmission', *IRE Trans. Inf. Theory*, 1954, **4**, pp. 93–110
- [34] WATANABE S.: 'Information theoretical analysis of multivariate correlation', *IBM J. Res. Dev.*, 1960, **4**, (1), pp. 66–82
- [35] PEARL J.: 'Probabilistic reasoning in intelligent systems: networks of plausible inference' (Morgan Kaufmann Publishers, Inc., San Francisco, CA, 1988)
- [36] JAKULIN A., BRATKO I.: Quantifying and visualizing attribute interactions. arXiv:cs/0308002, 2003
- [37] WANG K., NEMENMAN I., BANERJEE N., ET AL.: 'Genome-wide identification of modulators of transcriptional networks in human B lymphocytes'. Proc. Tenth Annual Int. Conf. on Research in Computational Molecular Biology (RECOMB), 2006, (LNCS, 3909), pp. 348–362
- [38] KULLBACK S., LEIBLER R.A.: 'On information and sufficiency', *Ann. Math. Stat.*, 1951, **22**, (1), pp. 142–143
- [39] JAYNES E.T.: 'Information theory and statistical mechanics', *Phys. Rev.*, 1957, **106**, pp. 620–630
- [40] LEWIS P.M.: 'Approximating probability distributions to reduce storage requirements', *Inf. Control.*, 1959, **2**, pp. 214–225
- [41] KU H.H., KULLBACK S.: 'Interaction in multidimensional contingency tables: an information theoretic approach', *J. Res. Natl. Bur. Stand (Math. Sci.)*, 1968, **72B**, (3), pp. 159–200
- [42] AMARI S.: 'Information geometry on hierarchy of probability distributions', *IEEE Trans. Inf. Theory*, 2001, **47**, (5), pp. 1701–1711
- [43] MARTIGNON L.: 'Neural coding: higher-order temporal patterns in the neurostatistics of cell assemblies', *Neural Comput.*, 2000, **12**, pp. 2621–2653
- [44] SOOFIE: 'A generalized formulation of conditional logit with diagnostics', *J. Am. Stat. Assoc.*, 1992, **87**, (419), pp. 812–816
- [45] CSISZAR I.: 'l-divergence geometry of probability distributions and minimization problems', *Ann. Probab.*, 1975, **3**, (1), pp. 146–158
- [46] DEMING W.E., STEPHAN F.F.: 'On a least squares adjustment of a sampled frequency table when the expected marginal totals are known', *Ann. Math. Stat.*, 1940, **11**, pp. 427–444
- [47] MARGOLIN A.A., CALIFANO A.: 'Theory and limitations of genetic network inference from microarray data', *Ann. NY Acad. Sci.*, 2007, **1115**, pp. 51–72
- [48] NEMENMAN I., BIALEK W., DE RUYTER VAN STEVENINCK R.R.: 'Entropy and information in neural spike trains: progress on the sampling problem', *Phys. Rev. E*, 2004, **69**, (5), article id 056111
- [49] NEMENMAN I., SHAFEE F., BIALEK W.: 'Entropy and inference revisited', in DIETTERICH T.G., BECKER S., GHAHRAMANI Z. (EDS.): 'Advances in neural information processing systems' (MIT press, Cambridge, MA, 2002, vol. 14) pp. 471–478
- [50] STRONG S.P., KOBERLE R., DE RUYTER VAN STEVENINCK R.R., BIALEK W.: 'Entropy and information in neural spike train', *Phys. Rev. Lett.*, 1998, **80**, pp. 197–200
- [51] NEMENMAN I., LEWEN G.D., BIALEK W., DE RUYTER VAN STEVENINCK R.R.: 'Neural coding of natural stimuli: information at sub-millisecond resolution', *PLoS Comput. Biol.*, 2008, **4**, (3), p. e1000025
- [52] FERNANDEZ P.C., FRANK S.R., WANG L., ET AL.: 'Genomic targets of the human c-Myc protein', *Genes Dev.*, 2003, **17**, (9), pp. 1115–1129
- [53] ZELLER K.I., JEGGA A.G., ARONOW B.J., O'DONNELL K.A., DANG C.V.: 'An integrated database of genes responsive to the Myc oncogenic transcription factor: identification of direct genomic targets', *Genome Biol.*, 2003, **4**, (10), p. R69
- [54] NIRO H., CLARK E.A.: 'Regulation of B-cell fate by antigen-receptor signals', *Nat. Rev. Immunol.*, 2002, **2**, (12), pp. 945–956
- [55] PERI S., NAVARRO J.D., AMANCHY R., ET AL.: 'Development of human protein reference database as an initial platform for approaching systems biology in humans', *Genome Res.*, 2003, **13**, (10), pp. 2363–2371

[56] KLEIN U., TU Y., STOLOVITZKY G.A., *ET AL.*: 'Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells', *J. Exp. Med.*, 2001, **194**, (11), pp. 1625–1638

[57] MARGOLIN A.A., WANG K., LIM W.K., KUSTAGI M., NEMENMAN I., CALIFANO A.: 'Reverse engineering cellular networks', *Nat. Protocols*, 2006, **1**, (2), pp. 662–671

[58] PALOMERO T., LIM W., ODOM D., *ET AL.*: 'NOTCH1 directly regulates MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth', *Proc. Natl. Acad. Sci.*, 2006, **103**, (48), pp. 18261–18266

[59] MATYS V., FRICKE E., GEFFERS R., *ET AL.*: 'TRANSFAC: transcriptional regulation, from patterns to profiles', *Nucl. Acids Res.*, 2003, **31**, (1), pp. 374–378

[60] LUSCHER B., KUENZEL E.A., KREBS E.G., EISENMAN R.N.: 'Myc oncoproteins are phosphorylated by casein kinase II', *Embo J.*, 1989, **8**, (4), pp. 1111–1119

[61] BOUSSET K., HENRIKSSON M., LUSCHER-FIRZLAFF J.M., LITCHFIELD D.W., LUSCHER B.: 'Identification of casein kinase II phosphorylation sites in Max: effects on DNA-binding kinetics of Max homo- and Myc/Max heterodimers', *Oncogene*, 1993, **8**, (12), pp. 3211–3220