1 2

A Statistical Model for Functional Characterization of Regulatory Pathways

Jayesh Pandey [‡] Mehmet Koyuturk [§] Wojciech Szpankowski [¶] Ananth Grama [‖]

Abstract

Standardized annotations of biomolecules in interaction networks (*e.g.*, Gene Ontology) provide comprehensive understanding of the function of individual molecules. Extending such annotations to pathways is a critical component of functional characterization of cellular signaling at the systems level. We propose a framework for projecting gene regulatory networks onto the space of functional attributes using multigraph models, with the objective of deriving statistically significant pathway annotations. We first demonstrate that annotations of pairwise interactions do not generalize to indirect relationships between processes. Motivated by this result, we formalize the problem of identifying statistically over-represented pathways of functional attributes. Then, we propose a statistical model that emphasizes the modularity of a pathway, evaluating its significance based on the coupling of its building blocks. We complement the statistical model by a comprehensive software infrastructure, NARADA, with an intuitive query interface. Comprehensive results from our methods on the *E. coli* transcription network demonstrate that our approach is effective in identifying known, as well as novel biological pathway annotations.

1 Introduction

Gene regulatory networks represent powerful formalisms for modeling cell signaling through regulation of cellular processes. These networks are inferred from gene expression, as well as other sources of data, using various statistical and computational methods [5]. Recent studies on networks of specific organisms show that interactions between genes that take part in certain pairs of biological processes are significantly overrepresented [8, 12]. Lee *et al.* [8] study the *S. cerevisiae* transcription regulation network with a view to understanding relationships between functional categories. They observe that many transcriptional regulators within a functional category bind to transcriptional regulators that play key roles in the control of other cellular processes. Similarly, Tong *et al.* [12] identify putative genetic interactions in yeast via synthetic genetic array (SGA) analysis and investigate the functional relevance of their results in the context of Gene Ontology (GO) annotations. They construct a network of GO terms by inserting an edge between any pair of terms that are *bridged* by a significant number of interacting gene pairs. Here, two GO terms are said to be bridged by an interaction if one of the interacting genes is associated with one of the terms, and the other gene with the second term, but neither is associated with both terms. They show that the resulting network is clustered according to underlying biological processes, while some biological processes buffer one another.

Generalizing such observations to pathways of arbitrary length may allow identification of standardized pathways, enabling creation of reference databases of direct and indirect interactions between various processes. Knowledge of such pathways is useful, not only in general understanding of the relationship between cellular processes at the systems level, but also in projecting existing knowledge of cellular organization of model organisms to other species. Increasing availability of species-specific interaction data, coupled with attempts aimed at creating standardized dictionaries of functional annotation for biomolecules, provide the knowledge base that can be effectively used for this purpose. What is lacking is a comprehensive set of tools that combine these two sources of data to identify significantly over-represented patterns of interaction through reliable statistical modeling with a formal computational basis.

In this paper, we introduce the notion of *functional network characterization*, derived from a gene regulatory network and associated functional annotations of genes. We use the Gene Ontology (GO) [1] for annotations, however, our methods themselves generalize to other networks and annotations. Functional network characterization is based on the *abstract* notion of regulatory interactions between pairs of functional attributes (as opposed to genes). In this context, we demonstrate that methods for identifying significant pairwise annotations do not generalize to pathway annotations. We introduce the problem of identifying statistically over-represented *pathways* of functional attributes, targeted at the identification of chains of regulatory interactions between functional attributes. Em-

*This paper appeared in the Data Mining for Biomedical Informatics Workshop which was held in conjunction with the 7th SIAM International Conference on Data Mining on April 28, 2007 in Minneapolis, MN.

†The work is supported by the National Institutes of Health and the National Science Foundation.

‡Department of Computer Science, Purdue University.

§Department of Computer Science, Purdue University.

¶Department of Computer Science, Purdue University.

‖Department of Computer Science, Purdue University.

phasizing the modularity of a pathway to assess its significance, we propose a statistical model that focuses on the coupling of the building blocks of a pathway. We use this statistical model to derive efficient algorithms for solving the pathway annotation problem. Our methods are implemented in a web-based tool, NARADA which provides an intuitive user and data interface. Comprehensive evaluation of NARADA on an *E. coli* transcription network from RegulonDB [11] shows that our method identifies several known, as well as novel pathways, at near-interactive query rates.

2 Multigraph Model for Networks of Functional Attributes

The basic approach for integrating existing knowledge of gene networks and functional annotations is to project the network in the *gene space* onto the *functional attribute space* through mapping of genes to attributes as specified by the annotation. A simple method for achieving this annotates each gene with its function and identifies overrepresented interacting annotations. This simple method yields interesting insights, as illustrated by [12] in the context of synthetic genetic arrays. This model, however, does not generalize beyond pairwise interactions since each interaction between a pair of functional attributes is within a specific context (a different pair of genes) in the network. For this reason, a pathway of functional attributes composed from pairwise interactions may not itself be significant, or even exist.

We develop a formal framework for projecting a gene network on a network of functional attributes, using *multigraph* models that accurately capture the context in which an interaction occurs. Through this framework, we generalize pairwise interactions between functional attributes to the identification of regulatory pathways of functional attributes. In our framework, A *gene regulatory network* is modeled by a directed graph $G(V_G, E_G)$. In this network, nodes $g_i \in V_G$ represent genes. Directed edge $g_i g_j \in E_G$, where $g_i, g_j \in V_G$, represents a regulatory interaction between genes g_i and g_j. A sample gene regulatory network is shown in Figure 1(a).

Each gene in the network is associated with a set of *functional attributes*. These attributes describe a functional *annotation* of the gene, *i.e.*, they map an individual biological entity to known functional classes. Formally, given a set of genes V_G and a set of functional attributes V_F, let 2^{V_G} and 2^{V_F} denote the power set of V_G and V_F, respectively. Then, functional annotation $\mathcal{A}(V_G, V_F) = \{\mathcal{F}, \mathcal{G}\}$ defines mapping $\mathcal{F} : V_G \rightarrow 2^{V_F}$ and $\mathcal{G} : V_F \rightarrow 2^{V_G}$, such that $T_j \in \mathcal{F}(g_i)$ if and only if $g_i \in \mathcal{G}(T_j)$, for any $g_i \in V_G$ and $T_j \in V_F$. The frequency of T_j, $\phi(T_j) = |\mathcal{G}(T_j)|$, is equal to the number of genes that are mapped to T_j.

In Figure 1, each gene g_i is tagged with the functional attributes in $\mathcal{F}(g_i)$. For each T_j, $\mathcal{G}(T_j)$ is composed of the genes tagged by T_j. We use Gene Ontology (GO) [1] as a reference library for annotating genes. For each gene, GO specifies the *molecular functions* associated with it, *biological processes* it takes part in, and *cellular components* it may be part of. Based on this mapping between genes and functional attributes, we model networks of functional attributes using multigraphs. A multigraph is a generalized graph, where multiple edges are allowed between a single pair of nodes.

DEFINITION 1. **Functional Attribute Network.** *Given gene regulatory network $G(V_G, E_G)$, a set of functional attributes V_F, and functional annotation $\mathcal{A}(V_G, V_F) = \{\mathcal{F}, \mathcal{G}\}$, the corresponding functional attribute network $F(V_F, E_F)$ is a multigraph defined as follows. The set of functional attributes V_F is also the set of nodes in F. Each node $T_i \in V_F$ contains a set of ports corresponding to the set of genes associated with T_i,* i.e., *$\mathcal{G}(T_i)$. Each multiedge $T_i T_j$ is a set of ordered port pairs (edges) $g_k g_\ell$, such that $g_k \in \mathcal{G}(T_i)$, $g_\ell \in \mathcal{G}(T_j)$, and $g_k g_\ell \in E_G$.*

The functional attribute network corresponding to the gene regulatory network in Figure 1(a) is shown in Figure 1(b). This multigraph model captures the context of each interaction accurately through the concept of ports. This model is more powerful than a simple graph model, in which paths that do not exist in the gene network emerge in the functional attribute network. This is not possible in the multigraph model, since a *path* must leave a node from the port in which it enters to the node.

DEFINITION 2. **Path & Multipath.** *In functional attribute network $F(V_F, E_F)$, a path $\pi = \{(T_{i_1}, g_{j_1}), (T_{i_2}, g_{j_2}), ..., (T_{i_k}, g_{j_k})\}$ is an ordered set of node-port pairs such that (i) $T_{i_r} \neq T_{i_s}$ for $1 \leq r < s \leq k$ (nodes are not repeated), (ii) $g_{j_r} \in \mathcal{G}(T_{i_r})$ for $1 \leq r \leq k$, and (iii) $g_{j_r} g_{j_{r+1}} \in T_{i_r} T_{i_{r+1}} \in E_F$ for $1 \leq r < k$ (consecutive edges are connected through the same port). The length of π is $|\pi| - 1 = k - 1$. A multipath $\Pi = \{T_{i_1}, T_{i_2}, ..., T_{i_k}\}$ is an ordered set of nodes such that (i) $T_{i_r} \neq T_{i_s}$ for $1 \leq r < s \leq k$, and (ii) there exist $g_{j_r} \in T_{i_r}$ for $1 \leq r \leq k$, such that $\{(T_{i_1}, g_{j_1}), (T_{i_2}, g_{j_2}), ..., (T_{i_k}, g_{j_k})\}$ is a path. The occurrence set $\mathcal{O}(\Pi)$ of Π consists of all distinct paths that satisfy (ii) and each such path is called an occurrence of Π. The frequency of Π, $\phi(\Pi) = |\mathcal{O}(\Pi)|$, is equal to the number of occurrences of Π.*

In Figure 1(b), $\{(T_1, g_1), (T_2, g_3), (T_4, g_6)\}$ is a path but $\{(T_1, g_1), (T_2, g_4), (T_4, g_6)\}$ is not, since multiedge

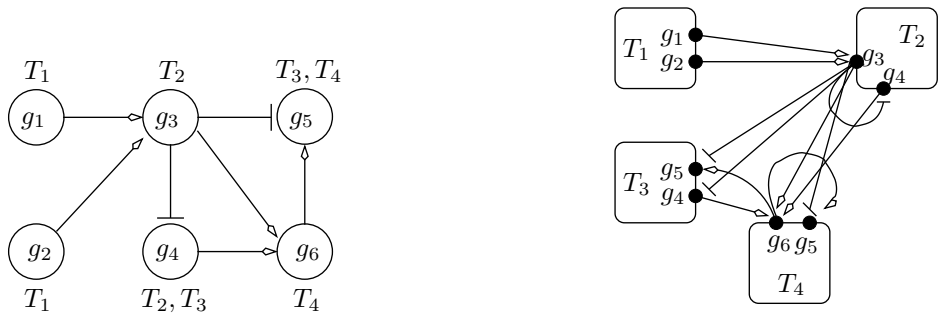


Figure 1: (a) A sample gene regulatory network and the functional annotation of the genes in this network. Each node represents a unique gene and is tagged by the set of functional attributes attached to that gene. Activator interactions are shown by regular arrows, repressor interactions are shown by dashed arrows. (b) Functional attribute network derived from the gene regulatory network in (a). In this multigraph, nodes (functional attributes) are represented by squares and ports (genes) are represented by dark circles.

T_1T_2 does not contain the edge g_1g_4. While analyzing regulatory pathways of functional attributes, however, we are interested in paths that are characterized by nodes in the functional attribute network. Clearly, such pathways may correspond to multiple paths in the functional attribute network. Therefore, we model them using *multipaths*. We use the terms *pathway* and *multipath* interchangeably, to emphasize the biological meaning of a multipath. In Figure 1(b), $\{T_1, T_2, T_3\}$ (also denoted $T_1 \rightarrow T_2 \dashv T_3$ throughout this paper) is a multipath with frequency four. On the other hand, multipath $T_2 \dashv T_4 \rightarrow T_3$ does not exist in this network, *i.e.*, it has frequency zero, although multiedges $T_2 \dashv T_4$ and $T_4 \rightarrow T_3$ both exist. A multipath with *high* frequency is likely to be biologically interesting, since it corresponds to a regulatory pathway of functional attributes that recurs in various contexts in the underlying cellular organization. In order to quantify this biological significance, it is useful to evaluate frequency from a statistical perspective.

3 Statistical Model for Pathways of Functional Attributes

We present a novel statistical model for assessing the significance of the frequency of a multipath in a functional attribute network. In this approach, the "interestingness" of a pathway is associated with its *modularity*, *i.e.*, the significance of the coupling of its building blocks. In statistical terms, this is achieved by conditioning the distribution of the frequency (modeled as a random variable) of a pathway on the frequency of its subpaths (modeled as fixed parameters).

Significance of a regulatory interaction. To quantify the significance of a pathway of shortest length (*i.e.*, a single regulatory interaction), we rely on a reference model that generates a functional attribute network. This model takes into account (i) the degree distribution of the underlying gene network, as well as (ii) the distribution of the number of genes associated with each functional attribute. This model is defined by a set of parameters that specifies the expected *multidegree* of each node in the functional attribute network. Given gene regulatory network $G(V_G, V_E)$, functional attribute set V_F, and annotation $\mathcal{A}(V_G, V_F)$, the expected in-degree $\beta(T_i)$ and out-degree $\delta(T_i)$ of a functional attribute $T_i \in V_F$ are estimated as $\hat{\beta}_i = \widehat{\beta(T_i)} = \sum_{T_j \in V_F} \phi(T_iT_j)$ and $\hat{\delta}_i = \widehat{\delta(T_i)} = \sum_{T_j \in V_F} \phi(T_jT_i)$. Given these parameters, we generate a functional attribute network as follows: there is a pool of *potential edges* that contains $\beta_i\delta_j$ potential edges between each pair of functional attributes T_i and T_j. The size of the pool is given by $m = \sum_{T_i, T_j \in V_F} \beta_i\delta_j$. A total of n edges are drawn from this pool, independently and without replacement, where n is equal to the number of edges in the observed functional attribute network, *i.e.*, $n = \sum_i \beta_i = \sum_j \delta_j$. In this model, the expected values of multidegrees in the generated network mirror the specifications.

Let $\Phi(\Pi)$ denote the random variable representing the frequency of pathway Π in the generated functional attribute network. Clearly, $\Phi_{ij} = \Phi(T_iT_j)$ is a hypergeometric random variable with parameters m (number of items), $\beta_i\delta_j$ (number of good items), n (number of selected items), and ϕ_{ij} (number of selected good items). Hence, the p-value of a regulatory interaction T_iT_j in the observed network, *i.e.*, the probability of observing at least ϕ_{ij} interactions between genes associated with T_i and genes associated with T_j, is given by

(3.1)
$$p_{ij} = P(\Phi_{ij} \geq \phi_{ij} | \mathcal{B}) = \sum_{\ell=\phi_{ij}}^{\min\{\beta_i\delta_j, n\}} \frac{\binom{\beta_i\delta_j}{\ell}\binom{m-\beta_i\delta_j}{n-\ell}}{\binom{m}{n}}.$$

Significance of a pathway. We now present a statistical model to assess the statistical significance of a pathway of functional attributes, which assumes a background distribution based on the occurrence of the building blocks of a pathway. Let $\Pi_{i,k}$ denote the path $\{T_{i_1}, T_{i_2}, ..., T_{i_k}\}$. For $1 < j < k$, we want to evaluate the significance of the coupling between pathways $\Pi_{1,j}$ and $\Pi_{j,k}$. In other words, we want to understand how strong a conclusion of the sort "If a gene $g_\ell \in \mathcal{G}(T_{i_j})$ is regulated through a chain of regulatory interactions characterized by $\Pi_{1,j}$, then this gene is likely to regulate a T_{i_k} gene through pathway $\Pi_{j,k}$" (or vice versa) can be.

To achieve this, we assume a reference model, in which the frequency of pathways $\Pi_{1,j}$ and $\Pi_{j,k}$ is established *a-priori*. Let Φ_{i-k} and ϕ_{i-k} denote $\Phi(\Pi_{i,k})$ and $\phi(\Pi_{i,k})$, respectively. Then, the p-value of the coupling between $\Pi_{1,j}$ and $\Pi_{j,k}$ is defined as follows:

$$p_{1,j,k} = P(\Phi_{1,k} \geq \phi_{1,k} | \Phi_{1,j} = \phi_{1,j}, \Phi_{j,k} = \phi_{j,k}). \tag{3.2}$$

Assume that a pool contains all possible occurrences of multipaths $\{T_{i_1}, T_{i_2}..., T_{i_j}\}$ and $\{T_{i_j}, T_{i_2}..., T_{i_k}\}$. Clearly, there are $m_{1,j} = \prod_{\ell=1}^{j} \phi_{i_\ell}$ and $m_{j,k} = \prod_{\ell=j}^{k} \phi_{i_\ell}$ potential occurrences of each multipath. Now consider a pair of paths, one corresponding to a potential occurrence of $\Pi_{1,j}$, the other to $\Pi_{1,k}$. Such a pair corresponds to a path, *i.e.*, an occurrence of $\Pi_{1,k}$, only if the second path originates in the port in which the first one terminates. Since there are $\phi_{1,j}$ and $\phi_{j,k}$ occurrences of $\Pi_{1,j}$ and $\Pi_{j,k}$, respectively, the problem is formulated as follows: we draw $\phi_{1,j}$ paths from $m_{1,j}$ potential occurrences of $\Pi_{1,j}$ and $\phi_{j,k}$ paths from $m_{j,k}$ potential occurrences of $\Pi_{j,k}$, forming $\phi_{1,j}\phi_{j,k}$ pairs. What is the probability that in at least $\phi_{1,k}$ of these pairs, the port on T_j will be common?

We approximate this probability using our result on the behavior of dense subgraphs [7] and Chvátal's bound on hypergeometric tail [3]. In order to apply these results, we resolve dependencies assuming that the selected path pairs are independent from each other. Then, letting $q_j = 1/\phi_j$ be the probability that a given path pair will go through the same gene and $t_{1,j,k} = \phi_{1,k}/\phi_{1,j}\phi_{j,k}$ be the fraction of observed paths among all existing pairs, we obtain the following bound:

$$p_{1,j,k} \leq \exp(\phi_{1,j}\phi_{j,k} H_{q_j}(t_{1,j,k})), \tag{3.3}$$

where $H_q(t) = t \log \frac{q}{t} + (1-t) \log \frac{1-q}{1-t}$ denotes weighted entropy. This estimate is Bonferroni-corrected for multiple testing, *i.e.*, it is adjusted by a factor of $\prod_{j=1}^{k} |\bigcup_{g_\ell \in T_{i_j}} \mathcal{F}(g_\ell)|$.

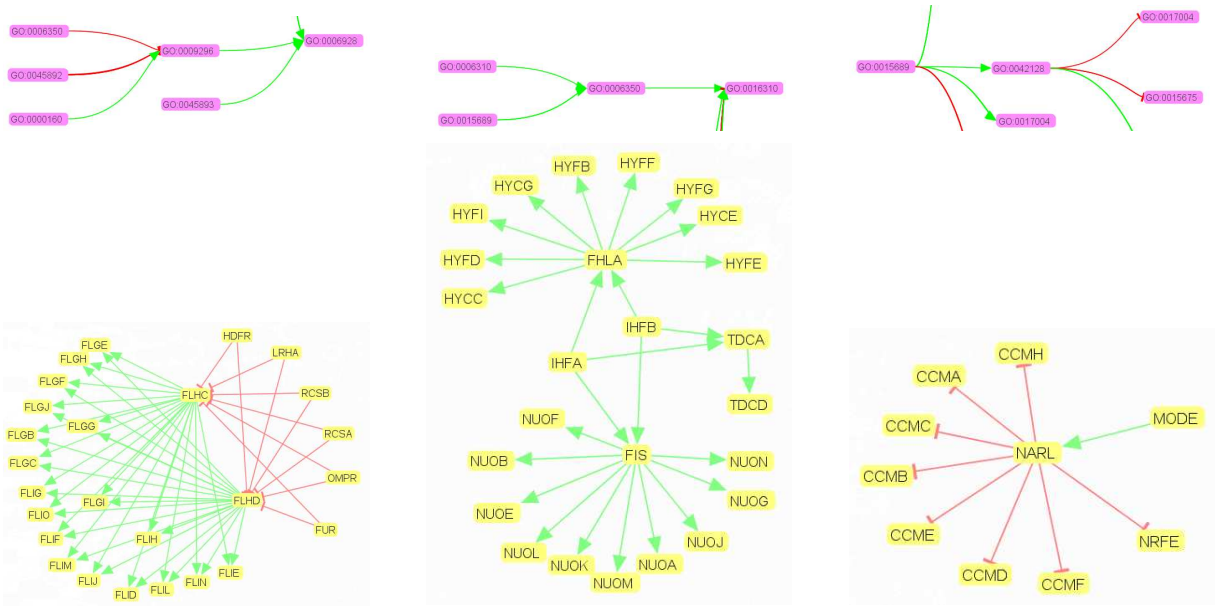
4 Experimental Results

Based on the above statistical model, we develop algorithms and a comprehensive software tool, NARADA, for projecting gene regulatory networks on the functional attribute domain. NARADA is implemented in Java and can be run as a web applet or an application. It is publicly available at `http://www.cs.purdue.edu/homes/jpandey/narada/`. A query in NARADA specifies a GO term and asks for all significantly overrepresented pathways of GO terms that originate from (or terminate at) that term. NARADA delivers near interactive query response using a novel, biologically motivated pruning technique. We call a pathway *strongly significant* if all of its subpaths are significant. In biological terms, a strongly significant pathway is likely to correspond to a significantly modular process, in which not only the building blocks of the pathway, but also the building blocks of the building blocks are tightly coupled. In the context of queries implemented in NARADA, these subpaths are limited to those that originate from (terminate at) the query term.

Table 1: Total number of significant pathways found by NARADA on *E. coli* transcription network for various path lengths.

Pathway length	2	3	4	5
All significant pathways	427	580	1401	942
Strongly significant pathways	427	208	183	142
Short-circuiting common terms	184	119	3	1

We test NARADA comprehensively on the *E. coli* transcriptional network obtained from RegulonDB [11]. The release 5.6. of this dataset contains 1364 genes with 3159 regulatory interactions. 193 of these interactions specify dual regulation. We separate these dual regulatory interactions as up and down regulatory interactions. We use Gene Ontology [1] as a library of functional attributes. The annotation of *E. coli* genes is obtained from UniProt GOA Proteome [2]. Using the mapping provided by GO, the gene network is mapped to functional attribute networks of the three name spaces in GO. Mapping to the biological process space provides maximum coverage in number of genes annotated, 881 genes mapped to one or more of 318 process terms. We collect all significant paths, using an α-value of 0.01 and varying path lengths from 2 to 5. The number of pathways obtained using combinations are shown in Table 1. On a Pentium M (1.6GHz) laptop with 1.21GB RAM the brute-force approach took on average 0.5 seconds per query for path length 2, to 12 seconds per query for paths of length 5. For strongly sig-



(a) (b) (c)

Figure 2: Pathways in gene network corresponding to (a) transcription $-|$ flagellum biogenesis $\rightarrow$ cell motility (b) DNA recombination $\rightarrow$ transcription $\rightarrow$ phosphorylation (c) molybdate ion transport $\rightarrow$ nitrate assimilation $-|$ cytochrome complex assembly. The pathways in functional attribute space are shown on the upper panel, their occurrences in the gene network are shown on the lower panel.

nificant paths, it took less than 2 seconds per query for paths of length 5, while for shortcutting terms it was 8 seconds per query for paths of length 4. Strongly significant pathways, *i.e.*, those obtained by extending only significant pathways, compose a significant portion of the highly significant pathways. This observation suggests that significantly modular pathways are also likely to be composed of significantly modular building blocks.

In Figure 2(a), significant pathways that regulate cell motility are shown. The *flhD* operon that encodes *flhC* and *flhD* has been shown to act as positive regulator of flagellar regulons(*fli, flg*) [10]. The flagellar master operon *flhDC*, in turn, is tightly regulated at the transcriptional level [6, 9, 4]. The output of NARADA captures this indirect regulation of flagellar expression perfectly. Parts of the significant pathways that regulate phosphorylation via genes involved in transcription and DNA recombination are shown in Figure 2(b). In Figure 2(c), indirect regulation of cytochrome complex assembly by molybdate ion transport is shown.

5 Conclusion

In this paper, we introduce the notion of statistically significant regulatory pathways of functional attributes. We provide a formal framework for projecting regulatory networks from gene space to functional attribute space. We propose a statistical model for functional attribute networks that emphasizes the modularity of pathways by conditioning on its building blocks. We present a comprehensive software tool, NARADA, which is based on the proposed models and methods. Finally, we present results obtained by testing NARADA on the *E. coli* transcription network.

References

[1] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. Gene ontology: Tool for the unification of biology. the gene ontology consortium. *Nat Genet*, 25(1):25–29, 2000.

[2] E. Camon, M. Magrane, D. Barrell, V. Lee, E. Dimmer, J. Maslen, D. Binns, N. Harte, R. Lopez, and R. Apweiler. The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. *Nucleic Acids Res*, 32(Database issue):262–266, Jan 2004.

[3] V. Chvátal. The tail of the hypergeometric distribution. *Discrete Mathematics*, 25:285–287, 1979.

[4] A. Francez-Charlot, B. Laugel, A. Van Gemert, N. Dubarry, F. Wiorowski, M.-P. Castanie-Cornet, C. Gutierrez, and K. Cam. RcsCDB His-Asp phosphorelay system negatively regulates the flhDC operon in Escherichia coli. *Mol Microbiol*, 49(3):823–832, Aug 2003.

[5] N. Friedman, M. Linial, I. Nachman, and D. Pe'er. Using bayesian networks to analyze expression data. *J Comput Biol*, 7(3-4):601–620, 2000.

[6] M. Ko and C. Park. H-NS-Dependent regulation of flagellar synthesis is mediated by a LysR family protein. *J Bacteriol*, 182(16):4670–4672, Aug 2000.

[7] M. Koyutürk, A. Grama, and W. Szpankowski. Assessing significance of connectivity and conservation in protein interaction networks. In *RECOMB'06*, volume LNBI 3909, pages 45–59, 2006.

[8] T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J. B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, and R. A. Young. Transcriptional regulatory networks in saccharomyces cerevisiae. *Science*, 298(5594):799–804, October 2002.

[9] D. Lehnen, C. Blumer, T. Polen, B. Wackwitz, V. F. Wendisch, and G. Unden. LrhA as a new transcriptional key regulator of flagella, motility and chemotaxis genes in Escherichia coli. *Mol Microbiol*, 45(2):521–532, Jul 2002. Comparative Study.

[10] X. Liu and P. Matsumura. The FlhD/FlhC complex, a transcriptional activator of the Escherichia coli flagellar class II operons. *J Bacteriol*, 176(23):7345–7351, Dec 1994.

[11] H. Salgado, S. Gama-Castro, M. Peralta-Gil, E. Daz-Peredo, F. Snchez-Solano, A. Santos-Zavaleta, I. Martnez-Flores, V. Jimnez-Jacinto, C. Bonavides-Martnez, J. Segura-Salazar, A. Martnez-Antonio, and J. Collado-Vides. RegulonDB (version 5.0): Escherichia coli K-12 transcriptional regulatory network, operon organization, and growth conditions. *Nucleic Acids Res*, 34(Database issue), January 2006.

[12] A. H. Y. Tong, G. Lesage, G. D. Bader, H. Ding, H. Xu, X. Xin, J. Young, G. F. Berriz, R. L. Brost, M. Chang, Y. Chen, X. Cheng, G. Chua, H. Friesen, D. S. Goldberg, J. Haynes, C. Humphries, G. He, S. Hussein, L. Ke, N. Krogan, Z. Li, J. N. Levinson, H. Lu, P. Menard, C. Munyana, A. B. Parsons, O. Ryan, R. Tonikian, T. Roberts, A.-M. Sdicu, J. Shapiro, B. Sheikh, B. Suter, S. L. Wong, L. V. Zhang, H. Zhu, C. G. Burd, S. Munro, C. Sander, J. Rine, J. Greenblatt, M. Peter, A. Bretscher, G. Bell, F. P. Roth, G. W. Brown, B. Andrews, H. Bussey, and C. Boone. Global mapping of the yeast genetic interaction network. *Science*, 303(5659):808–813, February 2004.