Research article

Information processing in the transcriptional regulatory network of yeast: Functional robustness

Frank Emmert-Streib*[1] and Matthias Dehmer[2]

Address: [1]Computational Biology and Machine Learning, Center for Cancer Research and Cell Biology, School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, 97 Lisburn Road, Belfast, BT9 7BL, UK and [2]Center for Mathematics, Probability and Statistics, University of Coimbra, Apartado 3008, 3001-454 Coimbra, Portugal

Email: Frank Emmert-Streib* - v@bio-complexity.com; Matthias Dehmer - dehmer@math.uc.pt

* Corresponding author

Abstract

Background: Gene networks are considered to represent various aspects of molecular biological systems meaningfully because they naturally provide a systems perspective of molecular interactions. In this respect, the functional understanding of the transcriptional regulatory network is considered as key to elucidate the functional organization of an organism.

Results: In this paper we study the functional robustness of the transcriptional regulatory network of *S. cerevisiae*. We model the information processing in the network as a first order Markov chain and study the influence of single gene perturbations on the global, asymptotic communication among genes. Modification in the communication is measured by an information theoretic measure allowing to predict genes that are 'fragile' with respect to single gene knockouts. Our results demonstrate that the predicted set of fragile genes contains a statistically significant enrichment of so called essential genes that are experimentally found to be necessary to ensure vital yeast. Further, a structural analysis of the transcriptional regulatory network reveals that there are significant differences between fragile genes, hub genes and genes with a high betweenness centrality value.

Conclusion: Our study does not only demonstrate that a combination of graph theoretical, information theoretical and statistical methods leads to meaningful biological results but also that such methods allow to study information processing in gene networks instead of just their structural properties.

Background

The advent of high-throughput technologies in molecular biology has initiated an avalanche of data that possess considerable challenges to quantitative sciences providing statistical analysis methods [1]. Due to the fundamental insight that biological processes should be studied holistically [2-4] instead of reductionistically, systems based approaches are of central importance in this respect [5].

For this reason, it is no surprise that network related studies experience an enormous interest starting with the investigation of small-world [6,7] and scale-free [8,9] networks in the mid 1990's followed by numerous studies devoted to the analysis of complex network topologies and their properties in general [8,10-14]. It is interesting to note that many apparently different networks have similar properties. Most prominent example is the degree dis-

tribution. For example, the World-Wide Web, the Internet or biological networks are found to be scale-free [8,10,11,15,16] with respect to their degree distribution. In molecular biology, metabolic, transcriptional regulatory, signaling and protein networks have been studied extensively during the last years [4,17-19] to shed light on the functional organization of these complex gene networks [20]. In this context, *functional robustness* is considered a key player for our understanding regarding the interplay of *network structure* and *network dynamics* leading to the emergence of life as omnipresent around us [9,21-23].

For general networks, one of the first studies that has thoroughly investigated *structural robustness* of systems that can be represented as networks is from ALBERT et al. [24]. ALBERT studied the error and attack tolerance of synthetic as well as real world networks and compared random and scale-free networks, e.g., the World-Wide Web or the Internet. By using purely graph theoretical measures – the diameter of the network and the size of the largest connected component – they found that scale-free networks are much more robust against random errors than random networks but more vulnerable against directed attacks. In the context of gene networks the interest shifts from the *structural robustness* of the networks to their *functional robustness* because the ultimate goal is of course to gain insights into the function of a living cell or an organism respectively. On a time scale of a living organism the question of *functional robustness* has been addressed by [25-28]. For example, in [27] the dynamics of Boolean networks [29,30] were studied serving as a simplified model for the signal processing taking place in gene networks. As major result [27] found that fluctuations occurring inevitably within the system, e.g., due to the inherent noise present on a molecular level [31,32], can be suppressed by a suitable design of the overall network topology [27]. On an evolutionary time scale the *functional robustness* of gene networks has been studied by [33-35] considering directly the role selective pressure might play during evolution leading to observable patterns of, e.g., protein structures, gene expression or network structures as present in current organisms. In this paper we tie up with previous studies aiming to analyze the *functional robustness* of networks on a time scale of living organisms. By pointing out the time scale we want to emphasize that we do not investigate the *evolutionary robustness* of an organism. Instead, the major objective of this paper is to investigate the functional robustness of the transcriptional regulatory network (TRN) of *S. cerevisiae* with respect to single gene perturbations. As quantitative measure of *functional robustness* we suggest to use an information theoretic measure [36], previously used to study synthetic networks, that does not focus directly on structural changes of the network topology due to the perturbations but on the alterations of information flow, modeled as Markov Chain [37], within the network as consequence of the structural modifications. The advantage of information theoretic measures [38-40] is that the concrete underlying dynamics does not need to be specified precisely, instead, a qualitative model is enough to gain principle insights into common working mechanisms with regard to more elaborate biological models. General entropy measures for quantifying structural information in networks have been developed in [41,42]. For our study, we use the transcriptional regulatory network of yeast [43,44] and apply our information theoretic measure to identify genes that are crucial for the functioning of the organism in the sense that disruptions of the transcriptional regulatory network are experienced strongest by these genes. For this reason we call these genes *fragile*. In this paper, we quantify our results by connecting these to the list of known so called *essential* genes of yeast [45] to demonstrate that our predictions are biologically meaningful.

Methods
In this section we present the information theoretic measure we use to analyze the transcriptional regulatory network of yeast to study its functional robustness.

Markov chains
We approximate the information flow in the network as a Markov chain. A Markov chain is a Markov process that is discrete in time and space. We define a Markov process by using a given network topology *G* and the plausible assumption that all possible interactions are equal likely. Plausible in this context does not necessarily mean that this corresponds best to the real situations, it means that it is the most simple and unbiased assumption one can make. For simplicity, we further assume the Markov process to be of first-order

$$T(X_{t+1} = j|X_t = i_t, ..X_1 = i_1) = T(X_{t+1} = j|X_t = i_t).$$

That means, the transition probability *T* depends only on the last state and not on states that are further in the past.

Definition 1: *The transition probability T for a Markov chain of first-order for a network G with adjaceny matrix A is defined by*

$$T\left(X_{t+1} = j \mid X_t = i\right) = \begin{cases} \dfrac{1}{k_i} & : k_i > 0 \wedge A_{ji} = 1 \\ 0 & : else \end{cases}$$

for all i, j $\in$ *V.*

Here $k_i = \sum_j A_{ji}$ is the degree of node (gene) *i* in the network and A_{ij} is a component of the adjacency matrix indicating

if node i is connected with node j (A_{ij} = 1) or unconnected (A_{ij} = 0). V denotes a set comprising all genes.

Single gene perturbations
In this paper we study the effect of single gene perturbations on the information processing in the transcriptional regulatory network of yeast. Formally, we define perturbations in the following way.

Definition 2: (Single gene perturbations) *If a gene k in network G is perturbed than all outgoing and incoming edges from this gene are deleted. In addition, one self-connection is introduced.*

In Fig. 1 and 2 we visualize a single gene perturbation. One can see that the perturbed gene (shown in red) does no longer participate in the information processing in the network. However, the remainder of the network is still structurally intact and capable to transmit signals. Hence, a single knockout can be considered as a perturbation and the modified communication among the remaining genes can be studied principally provided there is a measure to quantify these alterations. This measure is given in the next subsection.

Asymptotic Communication
The information theoretic measure we use to capture the asymptotic behavior of information processing evaluates the deviation of the unperturbed (or normal (n)) state from the perturbed (p) state caused by the perturbation of gene k. We use the the relative entropy also known as Kullback-Leibler (KL) divergence D [46,47] to quantify this deviation. Our asymptotic measure is given by the following definition.

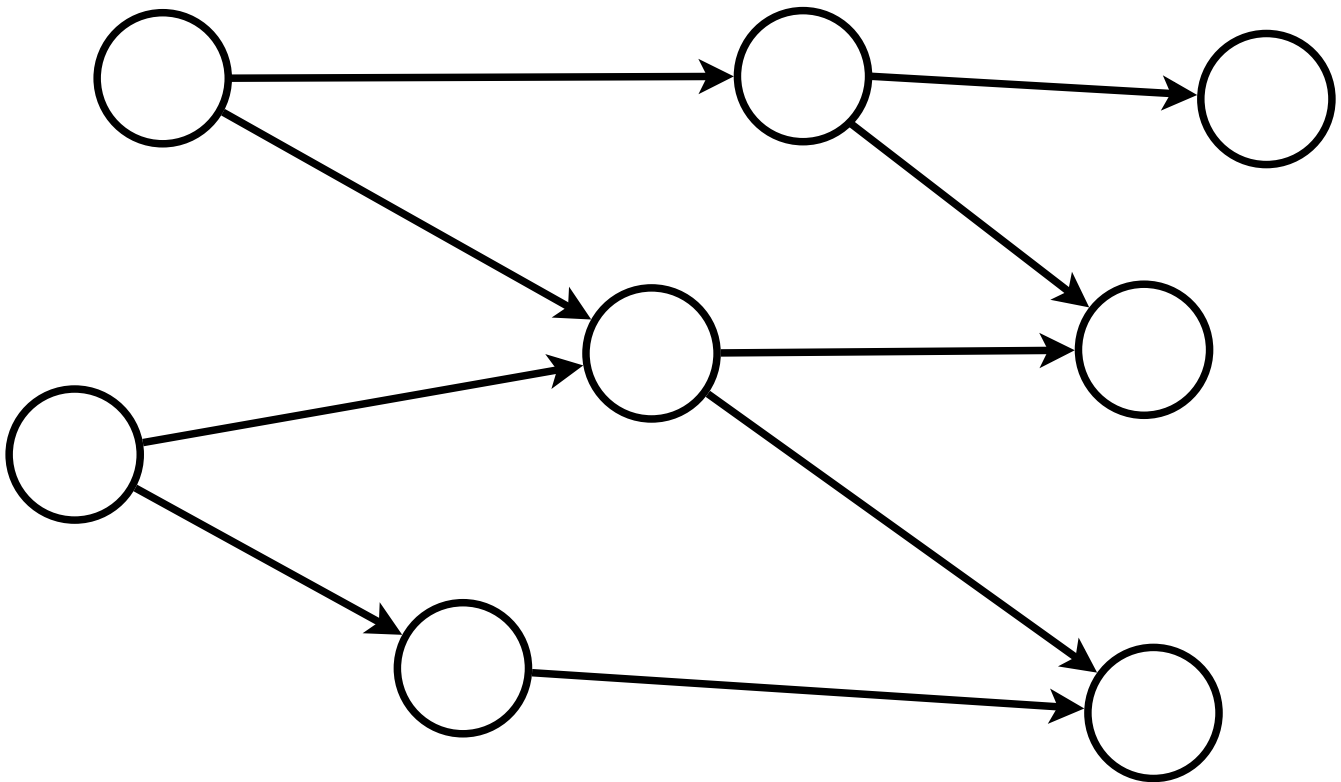**Definition 3: (Asymptotic information change)**

$$D_{ik} = D(p_{i,k}^{p,\infty} \,||\, p_i^{n,\infty}) = \sum_m p_{i,k}^{p,\infty}(m) \log \frac{p_{i,k}^{p,\infty}(m)}{p_i^{n,\infty}(m)}.$$

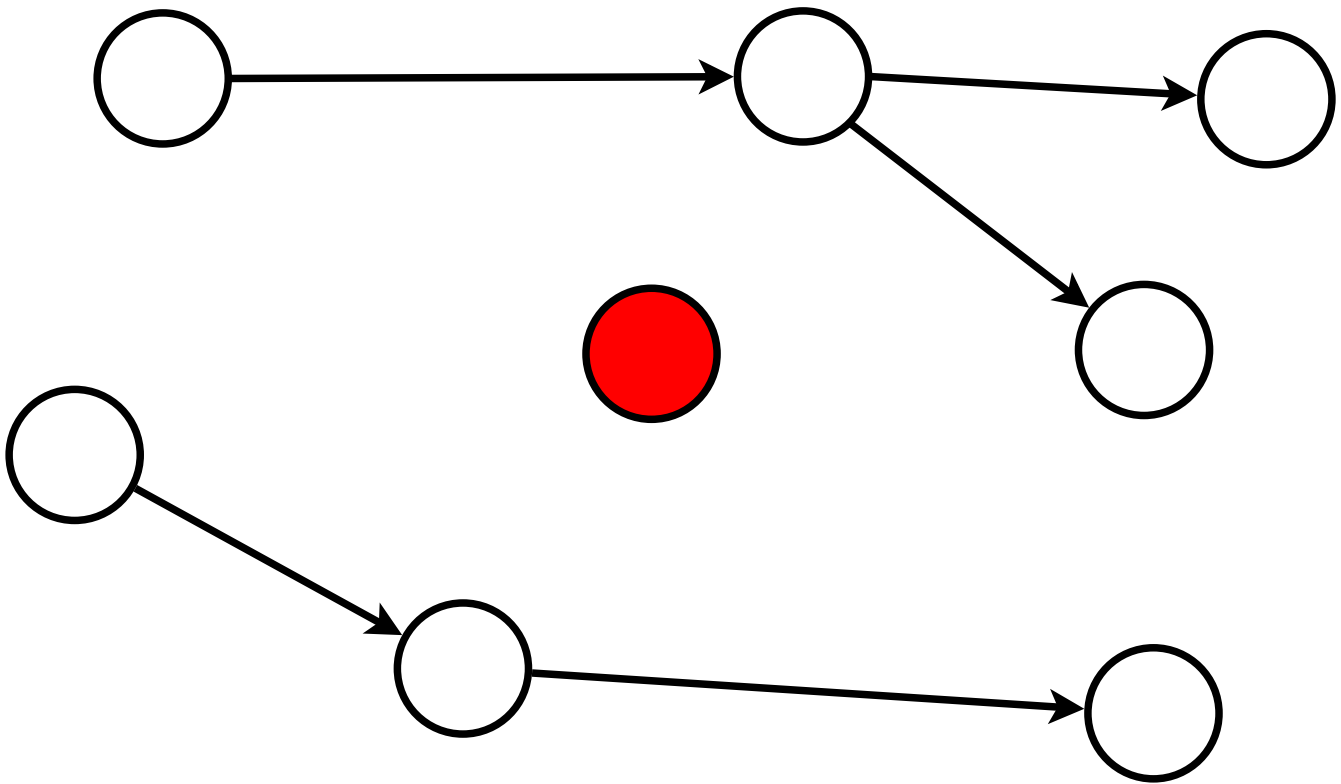Here $p_{i,k}^{p,\infty}$ and $p_i^{n,\infty}$ are stationary distributions obtained by

$$p_i^{n,\infty} = \lim_{t\to\infty} T^t p_i^0,$$

$$p_{i,k}^{p,\infty} = \lim_{t\to\infty} T_k^t p_i^0.$$

The Markov chain given by T_k corresponds to the process

Figure 1
A graph depicts the flow of information in the network.

Figure 2
Perturbing the node shown in red leads to a breakdown of communication between the red node and all other nodes. However, there is still an information flow in the remaining network.

obtained by perturbing gene k in the network. The stationary distribution $p_{i,k}^{p,\infty}$ of the perturbed (p) network obtained by perturbing gene k and starting from the initial distribution,

$$p_i^0 = \delta_{i,m} \quad \forall_m \in V,$$

depends on i because we use the Kronecker delta, which is one for $i = m$ and zero otherwise, as initial condition. The reason therefore is we consider i as starting point for the spread of information in the network. The interpretation for the unperturbed (normal (n)) distribution $p_i^{n,\infty}$ is correspondingly. We want to note that due to the directedness of the network the Markov process is not ergodic which results in a dependence of the asymptotic distributions $p_{i,k}^{p,\infty}$ and $p_i^{n,\infty}$ on the initial distribution p_i^0. For this reason it is important to use $|V|$ - 1 (starting from k is excluded because the perturbed gene has no longer outgoing edges) different initial distributions p_i^0 to evaluate D_{ik}. That means Eq. 3 defines the components of a matrix and the interpretation of D_{ik} is that the index k correspond to the deletion of gene k and index i referes to the initial

distribution $p_i^0 = \delta_{i,m}$. The diagonal elements D_{ik} ($i = k$) are not defined.

Results and discussion
Data
For our analysis we use the transcriptional regulatory network of yeast [43,44] which is a directed, unweighted network. From this network we extract the weakly connected component consisting of 3357 genes and 7230 interactions. The weakly connected component of a network is defined as the subnetwork that connects every pair of nodes by at least one directed path. That means for every pair of genes the weakly connected component ensures that communication (at least in one direction) between these genes is in principle enabled. This is an important characteristic because in our analysis we are aiming to quantify modifications of the communication among genes due to perturbations. Hence, if there would be no path between genes such an analysis would not be sensible.

On a practical note, we want to remark that our theoretical analysis described in detail in the next section is computationally expensive because we perform single gene perturbations for all genes in the network. That means, we do

not just analyze one network with our method but as many as genes in the network. Hence, the results presented in this article are obtained by analyzing 3357 networks. It is clear that this is getting more and more demanding computationally by increasing the number of genes in the network. From our simulations we found that networks with several thousand nodes can be studied within reasonable time whereas larger networks would require more algorithmic attention to reduce the computation time.

Results

Now we study the asymptotic behavior of the transcriptional regulatory network of yeast regarding information propagation under the influence of single gene perturbation.

For the normal (unperturbed) and perturbed network topology of the transcriptional regulatory network we determine Markov chains from which we calculate the stationary distributions. The perturbations correspond to single gene perturbations and the Markov chains are obtained as described in the methods section. From the resulting stationary distributions of the Markov chains we calculate the Kullback-Leibler divergence $D_{ik} = D$ ($p_{i,k}^{p,\infty} || p_i^{n,\infty}$) for all genes $i \in V$ and perturbations $k \in V$ with $i \neq k$. We want to note that due to the directedness of the network the resulting Markov process is no longer ergodic. Hence, information sent from different genes can results in different stationary distributions. For this reason, we use all N genes consecutively as sender gene. This is reflected by the index i in Eq. 3 corresponding to the gene from which the information was sent initially. On a mathematical note we want to remark that the network does not need to be disconnected to result in a non-ergodic Markov chain. However, the need to consider different initial conditions to study the behavior of the resulting stationary distributions meaningfully remains also true in this case.

We begin our analysis by investigating if the asymptotic results summarized by D_{ik} can be connected to local, structural properties of the genes in the network. For this reason we determine all genes with

$$D_i = \sum_k D_{ik} > 0,$$

and calculate the correlation with the in- and out-degree vector of the network. More precisely, we calculate Spearman's rank-order correlation coefficient [48] between the rank ordered vectors to decide if the order in these vectors is statistically preserved. For the in-degrees we obtain a correlation of $r = -0.39$ and $p = 6 \times 10^{-9}$, for the out-degrees

$r = 0.33$ and $p = 1 \times 10^{-6}$. Using a significance level of $\alpha = 0.05$ indicates that both rank correlations are statistically significant implying that, e.g., high out-degrees correspond to high values of D_i. These results seem plausible considering the following situation: For a given gene that is connected to all other genes (outgoing edges) it is clear, that an arbitrary knockout of a single gene effects with probability one an outgoing edge of this gene. Hence, this knockout will have an influence on the information processing of this gene. The strength of this influence can not be easily predicted given just this information, however, we will have an influence with probability one. Instead, a gene having very few outgoing connections has a lower probability that a single knockout effects one of its outgoing edges ($\text{Pr} = k_{out}/N_p$ with N_p the number of genes that can be perturbed). However, it is possible that the knocked out gene destroys some communication paths (secondary- or even higher-order effect if measured as Dijkstra distance [49]) and, hence, can still have a strong impact on the information processing. It seems to be reasonable to assume that the further away the knockout gene is from the starting gene (in Dijkstra distance [49]) the less the impact will be. This is a strong indicator that information processing on a systems level depends crucially on the information processing in a local environment of the gene that sends the information. We want to remark that in our analysis the number D_i, given in Eq. 7, is a global measure, whereas the degree vector is a local measure. This result is interesting because it demonstrates that the local properties of genes, given by their local connectedness, which can be roughly summarized by their degrees, are not averaged away with respect to the stationary distribution of the Markov process. That means the local *connectivity signature* is still detectable in the asymptotic behavior. We will come back to this point in the discussions section because this is a non-trivial point.

In Fig. 3 we show the components of the asymptotic information change D_{ik} for which $D_i \geq 0.1$ holds (149 genes). Blue corresponds to low values (zero) and cyan to high values of D_{ik} (the maximal value of D_{ik} is 20.47). The vertical stripes indicate that the knockout of a few genes effects many other genes whereas most knockouts have only a minor effect on other genes. This is also the reason why we do not show D_{ik} for $N = 3357$ genes because in this case the figure would appear essentially blue.

In Fig. 4 we show the histogram of D_i. From this figure one can see that the distribution of D_i has a heavy tail and that most values are around zero. This indicates that our measure has the desirable property to be very selective by evaluating most perturbations as minor. This corresponds to experimental results showing that only about 10% of all genes in yeast are categorized as *essential* [50] which
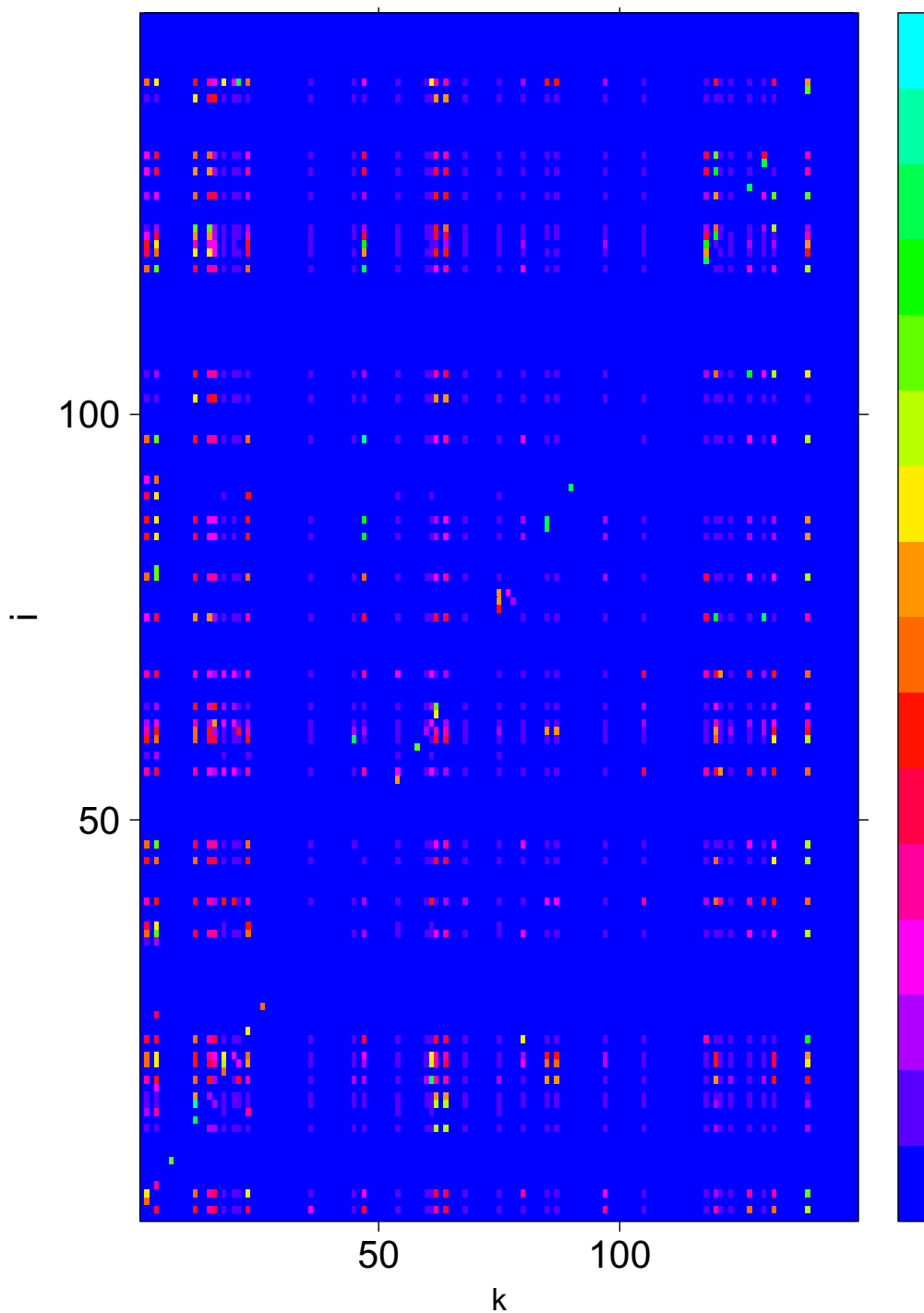
Figure 3
Asymptotic information change D_{ik} for all genes with $D_i \geq 0.1$.
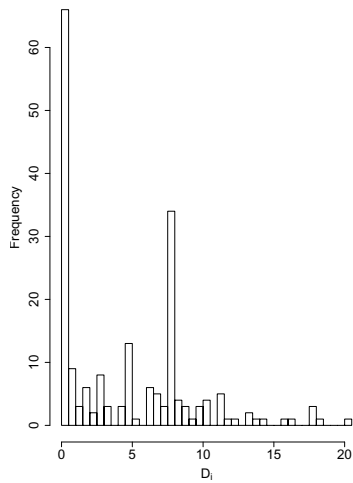
Figure 4
Histogram of D_i for all genes with $D_i > 0$.

means that their knockout has a catastrophic influence on the organism.

To point out the properties of our measure we show in Fig. 5 results connecting genes with high D_i values quantitatively to the appearance of essential genes from gene-deletion experiments in yeast [45]. Figure 5 shows N_e/N_c in dependence on Θ_D. N_c is the number of genes found for which $D_i > \Theta_D$ holds and N_e is the number of essential genes found in this set,

$$S_c = \{i | D_i > \Theta_D\},$$

$$N_c = \#S_c,$$

$$S_e = \{i | i \in S_c \text{ and } i \text{ is essential}\},$$

$$N_e = \#S_e.$$

The highest values found by this comparison are over 40%. A natural question arising now is if this occurred just by chance or is this high coverage unlikely to happened accidentally. Figure 6 provides information regarding this question. There we show p_D in dependence on Θ_D. The probability p_D is the sum of a hypergeometric distribution $p(k; N, N_E, n)$ giving the probability to observe exactly k essential genes in a set of size $n = N_c(\Theta_D)$ when the total number of genes is N containing N_E essential genes.

Hence,

$$p_D = \sum_{k=N_e}^{N_c} p(k; N, N_E, n),$$
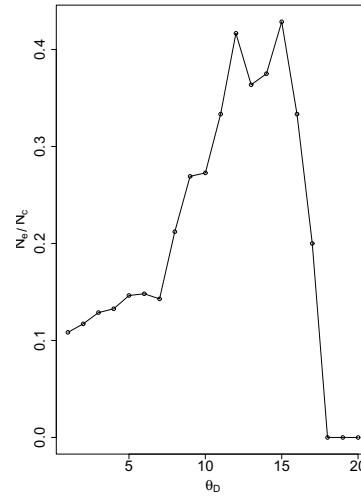
Figure 5
N_e/N_c **in dependence on Θ_D**.

is the probability to observe k or more essential genes in the set N_c. From Fig. 6 one can see that for $\Theta_D \in [8,15]$ the probability $p_D < 0.05$ (the dotted line corresponds to 0.05). This result suggests that the peaks observed in Fig. 5 do not appear by chance. Further, we obtain possible cut-off values for our gene set to be considered which corresponds to the interval $\Theta_D \in [8,15]$. From these results we decided to choose $\Theta_D^c = 12$ as cut-off value because for this value N_e/N_c assumes a maximum value. We call all genes for which $D_i > \Theta_D^c$ holds *fragile* genes.

In table 1 we show a list of 12 genes found by setting $\Theta_D^c = 12$ for which $p_D = 0.0039$. These genes are ordered according to their out-degree d_{out} in descending order. The first column gives the name of the gene. The second and third, the out- correspondingly in-degree of the gene in the TRN. The forth column gives the value of D_i and the fifth column indicates if the gene is found to be essential (Y) or non-essential (N) according to [45]. As one can see, the first gene (YNL216W) is a hub because $d_{out} = 240$. However, all other genes are not. Interestingly, YNL216W is not an essential gene according to [45]. The results in table 1 demonstrate that our measure does not prefer to select hub genes because only one hub was selected.

This underlines the non-trivial characteristics of our measure. For reasons of completeness we show in table 2 the top four knockout genes that cause the largest influence, as measured by

$$D_k = \sum_i D_{ik},$$
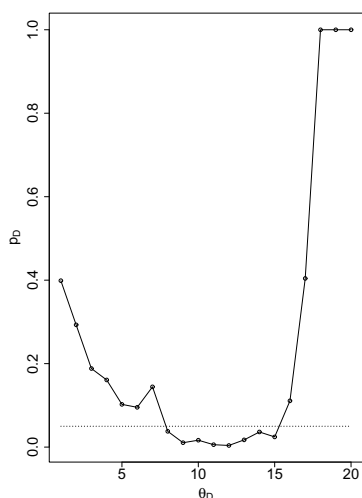
Figure 6
p_D **in dependence on** Θ_D. The dotted line corresponds to 0.05.

on other genes. The genes are again ranked according to their out-degrees. All of these genes are hubs. Considering the top 50 genes reveals that in this set 20 genes have an out-degree below 25 and even genes with an out-degree one and two are among these. Again, this demonstrates that hubness is no sufficient property to characterize these genes.

Next, we analyze the structure of the TRN containing our 12 genes shown in table 1. We find that each gene pair is connected (both ways) via a directed path. This implies that the subgraph formed from these 12 genes is part of the strongly connected component of the TRN. As side note we remark that for $\Theta_D \leq 9$ the resulting set of genes is no longer strongly connected and that also for $\Theta_D = 10$ this gene set does not correspond to the entire strongly connected component of the whole TRN comprising $N_{sc} = 36$ genes. Analysis of the strongly connected component of the whole TRN shows that it contains only 8 essential genes. From this we calculate the probability to find 8 or more essential genes by chance in the strongly connected component. By summing up the probabilities from a hypergeometric distribution we find $p_{sc} = 0.021$. This shows that essential genes are enriched in the strongly connected component, however, due to $p_D(\Theta_D = 12) << p_{sc}$ the strongly connected component represents a less favorable set to identify essential genes than the set found by our method. Fig. 7 shows D_{ik} for the strongly connected component. In contrast to Fig. 3 the influence of the perturbations is now much more severe as can be seen by the many non-blue dots. As a remark we want to remind that the diagonal of D_{ik} is not defined as explained in the methods section.

Table 1: Genes obtained for $\Theta_D^c = 12$.

gene	d_{out}	d_{in}	D_i	essential
YNL216W	240	2	17.5	N
YKL043W	92	1	17.9	N
YML007W	89	2	13.0	N
YFR034C	73	2	14.0	N
YER040W	44	1	17.8	Y
YBR112C	26	1	13.6	Y
YPL177C	23	2	13.3	N
YOL148C	13	1	20.4	N
YGL207W	12	2	18.1	N
YLR399C	9	3	12.2	Y
YDR138W	5	1	15.6	Y
YPR072W	4	1	16.1	Y

The first column gives the name of the gene.
The second and third, the out-correspondingly in-degree of the gene in the TRN. The fourth column gives the value of D_i and the fifth column indicates if the gene is found to be essential (Y) or non-essential (N) according to [45]. For this set $p_D = 0.0039$.

Finally, we test for the transcriptional regulatory network if our measure and betweenness centrality are similar by calculating Spearman's rank sum correlation coefficient. For the genes in table 1 we obtain a correlation coefficient $r = 0.0139$ and a p-value of 0.965 indicating that the results of both measures are not correlated. Further, we find that among the top 100 ranked genes of both measures only two genes are selected by both measures.

Discussion
In table 3 we provide some information about the biological processes the genes in table 1 are involved in. All genes found by our measure belong to the category 'regulation of transcription, DNA-dependent'. Further, some additional categories are listed for each gene in table 3. It is apparent that involvement in transcription regulation is the dominating category. This is interesting because the genes listed in the bottom of the table have a fairly low out-degree (see table 1). From our results we hypothesize that the genes found by our method, who have not been declared 'essential' by [45], should be 'fragile' in the sense

Table 2: Top four knockout genes that have the largest impact on other genes.

gene	d_{out}	d_{in}	D_k	essential
YML027W	314	2	26.8	N
YGL096W	248	0	97.4	N
YDL056W	129	0	149.8	Y
YHR206W	128	0	27.8	N

First column: gene name. Second column: out-degree. Third column: in-degree. Fourth column: D_k. Fifth column: essential genes (yes (Y) or no (N)).
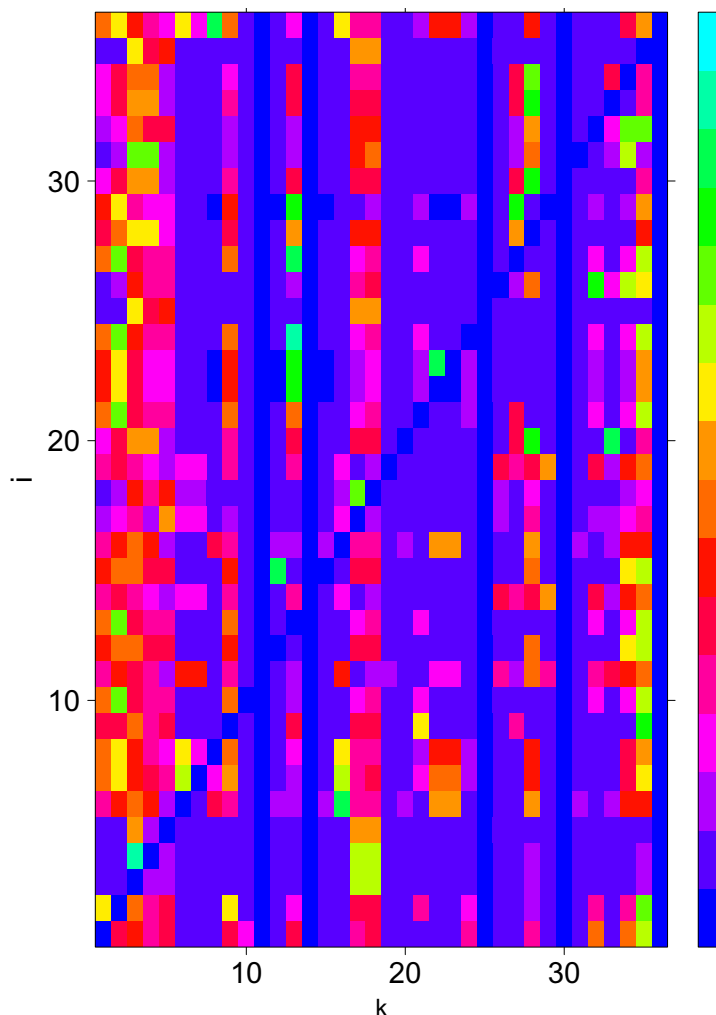
Figure 7
Asymptotic information change D_{ik} for all genes in the strongly connected component of the transcriptional regulatory network.

that they are influenceable quite severely by the malfunctioning of many other genes. Here, 'fragile' should not necessarily be equalized with 'essential' but the organism may be viable, however, it's overall function substantially impaired. It is also important to bear in mind that the list of essential genes used in our analysis is not necessarily complete. After intense literature research we found that YGL207W (also known as SPT16 – subunit of the heterodimeric FACT complex (Spt16p-Pob3p), facilitates RNA Polymerase II transcription elongation) is reported to be essential by [51] confirming our findings.

On a mathematical note we want to remark that the fact that $rank(D_i)$ and $rank(d_{out})$ respectively $rank(d_{in})$ are correlated, as shown in the beginning of the results section,

does not imply that our measure approximates or is even identical to the ranking of the degrees. This can be seen in table 1 because, e.g., the five bottom genes have $d_{out} < 20$, however, in the whole transcriptional regulatory network are 79 genes that have an out-degree larger than 20. But only seven of them appear in the list.

From a perspective of information processing the connection between asymptotic information change and local network structure represented by their degrees is interesting because it indicates that a local subgraph may be sufficient to study information processing in the overall network. This dissection is interesting because it would allow to reduce the computational complexity considerably that arises studying genomes like yeast or even organ-

Table 3: Biological processes the genes provided in table 1 participate.

gene	biological process
YNL216W	chromatin silencing at telomere, ribosome biogenesis
YKL043W	pos. regulation of trans. from RNA polymerase II promoter
YML007W	regulation of trans. from RNA polymerase II promoter in response to oxidative stress
YFR034C	response to starvation,
YER040W	positive regulation of transcription
YBR112C	negative regulation of transcription, chromatin remodeling
YPL177C	cellular copper ion homeostasis
YOL148C	chromatin modification, mitotic cell cycle
YGL207W	DNA repair, chromatin modification
YLR399C	DNA repair, chromatin remodeling
YDR138W	response to DNA damage stimulus
YPR072W	protein ubiquitination

In addition to the listed biological processes, all genes are involved in the category 'regulation of transcription, DNA-dependent'.

isms with more genes. In a former study [27], a similar idea has been proposed in a different methodological framework.

Finally, we want to remark that we repeated the analysis using $D_k = \sum_i D_{ik}$ as fragility measure of genes. However, for D_k we did not obtain meaningful results regarding the enrichment of essential genes. That means that the information captured by D_{ik} is asymmetric, as one would expect from it's construction.

Conclusion
In this paper we analyzed the influence that single gene perturbations have on the asymptotic communication abilities of the transcriptional regulatory network of yeast [43,44] to learn about the functional robustness of this network. To study the influence of the perturbations we used an information theoretic measure [36] and approximated the information propagation as a first order Markov chain directly defined for a given network topology. Our numerical studies obtained three major results. First, the asymptotic distributions for the perturbed and unperturbed network states carry implicitly information about their local origin from which the initial signal was transmitted. This confirms results previously found for synthetic networks [36]. Second, using our measure of asymptotic information change we could demonstrate that the predicted set of fragile genes contains a statistically significant enrichment of so called essential genes that are experimentally found to be necessary to ensure vital yeast. Third, a structural analysis of the transcriptional regulatory network revealed that there are significant differences between fragile genes, hub genes and genes with a high betweenness centrality value.

In addition to these findings we consider it to be important to emphasize that we employed graph theoretical, information theoretical as well as statistical methods [52]

because the biological information processing in gene networks is unlikely to be treated correctly in a deterministic framework. This demonstrates the power of interdisciplinary approaches and is at the heart of computational systems biology.

Authors' contributions
All authors contributed to all aspects of the article.

Acknowledgements

References
1. Emmert-Streib F, Dehmer M, Eds: *Analysis of Microarray Data: A Network Based Approach* Wiley VCH; 2008.
2. Alon U: *An Introduction to Systems Biology: Design Principles of Biological Circuits* Chapman & Hall/CRC; 2006.
3. Kitano H: *Foundations of Systems Biology* MIT Press; 2001.
4. Palsson BO: *Systems Biology* Cambridge University Press; 2006.
5. von Bertalanffy L: *General System Theory: Foundation, Development, Application* New York, George Braziller; 1968.
6. Watts D, Strogatz S: **Collective dynamics of 'small-world' networks.** *Nature* 1998, **393:**440-442.
7. Watts D: *Small Worlds: The Dynamics of Networks between Order and Randomness* Princeton University Press; 1999.
8. Albert R, Jeong H, Barabasi AL: **Diameter of the world wide web.** *Nature* 1999, **401:**130-131.
9. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL: **The large-scale organization of metabolic networks.** *Nature* 2000, **407:**651-654.
10. Albert R, Barabasi A: **Statistical Mechanics of Complex Networks.** *Rev of Modern Physics* 2002, **74:**47-97.
11. Newman MEJ: **The Structure and Function of Complex Networks.** *SIAM Review* 2003, **45:**167-256.
12. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL: **Hierarchical Organization of Modularity in Metabolic Networks.** *Science* 2002, **297:**1551-1555.
13. Schwikowski B, Uetz P, Fields S: **A network of protein-protein interactions in yeast.** *Nat Biotechnol* 2000, **18:**1257-1261.
14. Soinov LA, Krestyaninova MA, Brazma A: **Towards reconstruction of gene networks from expression data by supervised learning.** *Genome Biology* 2003, **4:**R6.
15. Bornholdt S, Schuster H, Eds: *Handbook of Graphs and Networks: From the Genome to the Internet* Wiley-VCH; 2003.
16. van Noort V, Snel B, Huymen MA: **The yeast coexpression network has a small-world, scale-free architecture and can be explained by a simple model.** *EMBO reports* 2004, **5(3):**280-284.

17. Friedman N: **Inferring cellular networks using probabilistic graphical models.** *Science* 2004, **303(5659):**799-805.
18. Margolin A, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, Califano A: **ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context.** *BMC Bioinformatics* 2006, **7:**S7.
19. Werhli A, Grzegorczyk M, Husmeier D: **Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical gaussian models and bayesian networks.** *Bioinformatics* 2006, **22(20):**2523-31.
20. Barabasi AL, Oltvai ZN: **Network Biology: Understanding the Cell's Functional Organization.** *Nature Reviews* 2004, **5:**101-113.
21. Kitano H: **Biological robustness.** *Nat Rev Genet* 2004, **5:**826-37.
22. Stelling J, Sauer U, Szallasi Z, Doyle F III, Doyle J: **Robustness of Cellular Functions.** *Cell* 2004, **118:**675-685.
23. Kitano H: **Towards a theory of biological robustness.** *Mol Syst Biol* 2007, **3:**137.
24. Albert R, Jeong H, Barabasi AL: **Error and attack tolerance of complex networks.** *Nature* 2000, **406:**378-482.
25. Li F, Long T, Lu Y, Ouyang Q, Tang C: **The yeast cell-cycle network is robustly designed.** *Proc Natl Acad Sci USA* 2004, **101(14):**4781-6.
26. Ingolia N: **Topology and robustness in the Drosophila segment polarity network.** *PLoS Biol* 2004, **2(6):**e123.
27. Klemm K, Bornholdt S: **Topology of biological networks and reliability of information processing.** *PNAS* 2005, **102(51):**18414-18419.
28. Wilhelm T, Behre J, Schuster S: **Analysis of structural robustness of metabolic networks.** *Syst Biol* 2004, **1:**114-120.
29. Kauffman S: **Metabolic stability and epigenesis in randomly constructed genetic nets.** *Journal of Theoretical Biology* 1969, **22:**437-467.
30. Kauffman S: *Origins of Order: Self-Organization and Selection in Evolution* Oxford University Press; 1993.
31. McAdams HH, Arkin A: **Stochastic Mechanisms in Gene Expression.** *Proc Natl Acad Sci USA* 1997, **94:**814-819.
32. Tu Y, Stolovitzky G, Klein U: **Quantitative noise analysis for gene expression microarray experiments.** *Proc Natl Acad Sci USA* 2002, **99:**14031-6.
33. Wagner A: **Robustness against mutations in genetic networks of yeast.** *Nature Genetics* 2000, **24:**355-61.
34. Wagner A: **Robustness, Neutrality, and Evolvability.** *FEBS Letters* 2005, **579:**1772-1778.
35. Wagner A: *Robustness and Evolvability in Living Systems* Princeton University Press; 2007.
36. Emmert-Streib F, Dehmer M: **Fault Tolerance of Information Processing in Gene Networks.** *Physica A* 2009, **388:**541-548.
37. Norris J: *Markov Chains* Cambridge University Press; 1998.
38. Gallager R: *Information Theory and Reliable Communication* Wiley, New York; 1968.
39. MacKay D: *Information theory, inference and learning algorithms* Cambridge University Press; 2003.
40. Shannon C, Weaver W: *The Mathematical Theory of Communication* University of Illinois Press; 1949.
41. Dehmer M, Emmert-Streib F: **The Structural Information Content of Chemical Networks.** *Zeitschrift für Naturforschung A* 2008, **63a:**155-158.
42. Dehmer M, Borgert S, Emmert-Streib F: **Entropy Bounds for Hierarchical Molecular Networks.** *PLoS ONE* 2008, **3(8):**e3079.
43. Luscombe N, Badu M, Yu H, Snyder M, Teichmann S, Gerstein M: **Genomic analysis of regulatory network dynamics reveals large topological changes.** *Nature* 2004, **431:**308-312.
44. Yu H, Kim P, Sprecher E, Trifonov V, Gerstein M: **The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics.** *PLoS Computational Biology* 2007, **3(4):**e59.
45. Giaever G, *et al.*: **Functional profiling of the Saccharomyces cerevisiae genome.** *Nature* 2002, **418:**387-391.
46. Kullback S, Leibler R: **On information and sufficiency.** *Ann Math Stat* 1951, **22:**79-86.
47. Kullback S: *Information theory and statistics* Wiley, New York; 1959.
48. Sheskin DJ: *Handbook of Parametric and Nonparametric Statistical Procedures* 3rd edition. RC Press, Boca Raton, FL; 2004.
49. Dijkstra E: **A note on two problems in connection with graphs.** *Numerische Math* 1959, **1:**269-271.
50. Deutscher D, Meilijson I, Schuster S, Ruppin E: **Can single knockouts accurately single out gene functions.** *BMC Systems Biology* 2008, **2:**50.
51. Orphanides G, *et al.*: **The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins.** *Nature* 1999, **400:**284-288.
52. Emmert-Streib F, Dehmer M, Eds: *Information Theory and Statistical Learning* Springer, New York; 2008.