Flux-Based *vs.* Topology-Based Similarity of Metabolic Genes

Oleg Rokhlenko[1], Tomer Shlomi[2,⋆], Roded Sharan[2,⋆⋆], Eytan Ruppin[2], and Ron Y. Pinter[1]

[1] Dept. of Computer Science, Technion–IIT, Haifa 32000, Israel
{olegro, pinter}@cs.technion.ac.il
[2] School of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel
{shlomito, roded, ruppin}@tau.ac.il

Abstract. We present an effectively computable measure of functional gene similarity that is based on metabolic gene activity across a variety of growth media. We applied this measure to 750 genes comprising the metabolic network of the budding yeast. Comparing the *in silico* computed functional similarities to those obtained by using experimental expression data, we show that our computational method captures similarities beyond those that are obtained by the topological analysis of metabolic networks, thus revealing—at least in part—dynamic characteristics of gene function. We also suggest that network centrality partially explains functional centrality (*i.e.* the number of functionally highly similar genes) by reporting a significant correlation between the two. Finally, we find that functional similarities between topologically distant genes occur between genes with different GO annotations.

1 Introduction

The study of biological networks has attracted considerable attention in recent years, including the construction of mathematical models to elucidate both cell activity as well as genes' function and expression. Much of the work to date has attempted to establish measures for the similarity (or distance) between genes that are based on the topological properties of metabolic networks. Even though recent analyses have provided valuable insights regarding this issue [1,2], topological characteristics alone (as devised by *e.g.* Kharchenko *et al.* [3], Chen and Vitkup [4]) offer only a static description of the properties of interest. On the other hand, accurate prediction of dynamic cell activity using kinetic models requires detailed information on the rates of enzyme activity which is rarely available; moreover, such analysis is usually limited to small-scale networks.

Fortunately, for metabolic networks, the use of stochiometry and other sources of information provides an added value over the topology of the underlying structure. Specifically, constraint-based stochiometric models have emerged as a key

⋆ Supported in part by the Tauber Fund.
⋆⋆ Supported by an Alon Fellowship.

P. Bücher and B.M.E. Moret (Eds.): WABI 2006, LNBI 4175, pp. 274–285, 2006.

method for studying such networks permitting large-scale analysis thereof. They use genome-scale networks to predict steady-state metabolic activity, regardless of specific enzyme kinetics. In these models, stoichiometric, thermodynamic, flux capacity and possibly other constraints affect the space of possible flux distributions attainable by a metabolic network.

In this paper we devise an effectively computable functional similarity measure between genes that is based on their metabolic activity. Such a measure would allow us to perform large scale *in silico* experiments and predict functional relations that can then be validated by experimental methods. Specifically, we suggest a method for determining similarities in gene activities that is based on Flux Balance Analysis (FBA). We first suggest a knockout-based measure but find it to be only moderately correlated with experimental data (of gene co-expression, see below). We then employ a measure of *metabolic genes co-activity* (MGCA), which tells how similar gene functions are in terms of the correlation between their corresponding flux activity vectors across a large variety of growth media. This latter measure, already used in a more limited scope by [5], is significantly better than the former measure in terms of correlation with experimental data.

Our evaluation of the suggested measures is based on testing their correlations with experimental data on similarity in gene expression, to assess their veracity. The basic relation between metabolic fluxes and gene expression was already studied and established previously both computationally (showing only a moderate correlation) as well as experimentally. Recall that the metabolic state of an organism is controlled via transcriptional regulation which adjusts gene expression levels according to metabolic demands [6]. Previous studies have shown that the expression patterns of enzyme coding genes are correlated with the flux patterns predicted by FBA: Schuster *et al.* [7,8] and Famili *et al.* [5] have shown that genes, associated with fluxes which are predicted to change together when shifting from one medium to another (*e.g.* in diauxic shift), are co-expressed under these conditions; Reed and Palsson [9] have shown that the genes associated with fluxes that are correlated within the solution space also exhibit moderate levels of correlation in their expression. Recently, Bilu *et al.* [10] proposed a more direct relation between expression and flux where the range of possible optimal flux values for a given reaction reflects evolutionary constraints on the expression levels of its associated enzymes; specifically, they have shown that the regulation of reactions which have an optimal fixed value is under strong selection to maintain their flux at the precise levels needed, while the regulation of reactions which may have a broad range of optimal values is under weaker selection.

In this work we extend upon these previous studies to look into ways of building upon the reported correlation between fluxes and expression, to construct efficient measures of functional similarity among metabolic genes. To this end, in contrast with the previous studies, we examine the relation between fluxes and expression while concomitantly controlling for correlations caused solely by the network's topology.

Our comparison focuses on 750 metabolic genes of the yeast *Saccharomyces cerevisiae*. We find the correlation between MGCA and co-expression to be

statistically significant. Furthermore, it remains so even after cancelling the effect of the underlying (static) network topology. These results support the notion that our measure indeed captures the true functional similarity between metabolic genes.

2 Preliminaries—Modeling Metabolism and Flux Balance Analysis

Flux Balance Analysis (FBA) [11,12] is a particular constraint-based method which assumes that the network is regulated to maximize or minimize a certain cellular function, which is usually taken to be the organism's growth rate. FBA has been demonstrated to be a very useful technique for the analysis of metabolic capabilities of cellular systems [13,14]. It involves carrying out a steady state analysis, using the stoichiometric matrix (as defined below) for the system in question. The system is assumed to be optimized with respect to functions such as maximization of biomass production or minimization of nutrient utilization; it is solved accordingly to obtain a steady state flux distribution, which is then used to interpret the metabolic capabilities of the system.

In FBA, the constraints imposed by stoichiometry in a chemical network at steady state are analogous to Kirchoff's Second Law for the flow of currents in electric circuits [15], namely—for each of the M metabolites in a network the net sum of all production and consumption fluxes, weighted by their stoichiometric coefficients, is zero:

$$\sum_{j=1}^{N} S_{ij} v_j = 0, \quad i = 1, \ldots, M \tag{1}$$

Here, S_{ij} is the element of the stoichiometric matrix S corresponding to the stoichiometric coefficient of metabolite i in reaction j. The flux v_j is the rate of reaction j at steady state, and is the j-th component of an N-dimensional flux vector v, where N is the total number of fluxes. In addition to internal fluxes, which are associated with chemical reactions, v includes exchange fluxes that account for metabolite transport through the membrane. The steady-state approximation is generally valid because of the fast equilibration of metabolite concentrations (seconds) with respect to the time scale of genetic regulation (minutes) [16].

Additional constraints, including those pertaining to the availability of nutrients or to the maximal fluxes that can be supported by enzymatic pathways, can be introduced as the following inequalities:

$$\alpha_j \leq v_j \leq \beta_j \tag{2}$$

For example, for a substrate uptake flux v_j, one can set α_j and β_j to be equal to the corresponding measured or imposed values. Eq. 2 can also be used to distinguish reversible and irreversible reactions, where $\alpha_j = 0$ for the latter.

All flux vectors that satisfy the constraints mentioned above define a feasible space, Φ. For an underdetermined system, as is typically the case in FBA models of cellular metabolic networks [13], Φ is a convex set in the N-dimensional space of fluxes [17]. Due to the linear nature of Φ, it is possible to use linear programming [18] to characterize the points in Φ that maximize or minimize a given linear objective function. A natural choice for an objective function in metabolic models of prokaryotes and simple eukaryotes is biomass production [13,14], as it is reasonable to hypothesize that unicellular organisms have evolved towards maximal growth performance. This process is formalized by introducing a growth flux that transforms a linear combination of fundamental metabolic precursors into biomass.

The maximization of biomass production is implemented by defining an additional flux v_{gro} associated with cell growth. For this flux, the stoichiometric factors of the reactants are the experimentally known proportions c_i of metabolite precursors X_i contributing to biomass production [13]:

$$c_1X_1 + c_2X_2 + \ldots + c_MX_M \stackrel{v_{gro}}{\rightarrow} Biomass \tag{3}$$

The search for the flux vector maximizing v_{gro} under the constraints of Eqs. 1 and 2 is solved using the Simplex algorithm.

The theoretical basis of FBA is supported by several experiments. These include empirical validation of growth yield and flux predictions [13,14], measurements of uptake rates around the optimum under various conditions [19], and results from large-scale gene deletion experiments [20].

For the stoichiometric analysis of the metabolic network of *S. cerevisiae*, we have used the reconstruction by Duarte, Herrgard, and Palsson [21]. The nodes of this network correspond to metabolic genes, and the edges correspond to the connections established by metabolic reactions. Two metabolic genes are connected if the corresponding enzymes share a common metabolite among their substrates or products. The list of metabolic reactions, and the 1060 (metabolites) by 1149 (fluxes) stoichiometric matrix (available at `http://gcrg.ucsd.edu`) were compiled using data from public databases and the literature. The 1149 reactions are associated with 750 genes. As in previous FBA formulations, we use inequalities (Eq. 2) to limit nutrient uptake and to implement reactions' irreversibility. In addition to the 1149 internal reactions, we added to the model 116 uptake/excretion reactions, for each of the metabolites listed as "extracellular" in the basic model.

3 Similarity Measures for Metabolic Genes

In the context of the aforementioned motivation, we suggest two techniques for obtaining the distance between metabolic genes: a knockout-functional (KF) scheme and a growth-functional (GF) scheme. The biological plausibility of the obtained distance measures is validated by correlating them with the corresponding similarity measure obtained by expression data.

3.1 Knockout-Functional Scheme

Cellular response to a gene knockout involves rerouting of metabolic flux through alternative pathways and the utilization of isoenzymes [22,23]. We hypothesize that similar metabolic responses to gene knockouts may provide evidence for similar metabolic functionality between genes. Based on this hypothesis, we define the KF similarity measure between gene pairs as the similarity in the metabolic response following their knockout.

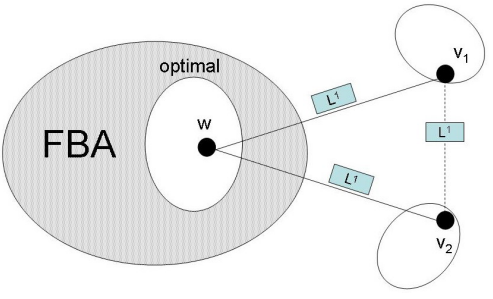


Fig. 1. Schematic illustration of the proposed flux similarity model. w stands for the optimal flux distributions on the wild-type metabolic network, v_1 stands for the optimal flux distribution on the metabolic networks with the first flux knocked-out, and v_2 stands for the optimal flux distribution on the metabolic networks with the second flux knocked-out.

Predicting the metabolic response for gene knockouts is a more difficult task than predicting the metabolic state of wild-type strains. Gene deletion is commonly modeled by constraining the flux through the reactions associated with a given gene to zero, and applying FBA [13]. However, it turns out that the metabolic state of the knocked-out strain is not necessarily optimal in terms of growth rate, and thus in many cases FBA's predictions are inaccurate. Instead, it was hypothesized that the cell adapts to gene knockouts by minimizing the change in its metabolic state. Specifically, the Minimization of Metabolic Adjustment (MOMA) approach searches for a metabolic state for a knocked-out strain with minimal distance, under the L2 norm, from the flux distribution of the wild-type strain [22]. Recently, a new method called Regulatory On-Off Minimization (ROOM) was suggested to predict metabolic states following gene knockouts, and was shown to provide better predictions of knockout phenotypes [23]. ROOM aims to minimize the number of regulatory changes required for the adaptation by minimizing the number of significant flux changes between the metabolic states of the wild-type and knocked-out states (*i.e.* using the norm L0).

A naive method for measuring the distance between the metabolic responses of two gene knockouts would be to simulate the knockout of each of them individually using ROOM, and then compute the distance between the obtained flux distributions. However, in many cases ROOM (like FBA and MOMA) provides multiple possible metabolic states for the knocked-out strain rather than a single solution. In these cases, it is not clear how to define the similarity measure between two genes.

To overcome this problem we define the KF similarity measure as the minimal distance between the optimal ROOM solutions for the two genes[1]. This is achieved by formulating a single optimization problem to find two ROOM solutions with minimal distance between them. The schematic illustration of our model is presented in Figure 1.

Notably this formulation depends on the choice of a wild-type and thus we repeat our analysis for several different wild-types. Furthermore, since ROOM requires Mixed Integer Linear Programming (MILP) optimization which is NP-hard, we use a relaxed version of ROOM and, in addition, we use the L1 norm instead of L0. The use of the L1 norm is similar to a variant of ROOM, called ROOM-LP, that was shown to provide similar predictions to ROOM [23]. The L1 norm was also used by Kuepfer *et. al.* [24] for a similar purpose of knockout prediction. The distance between the two flux distributions of the knocked-out strains is also minimized using the L1 norm.

The optimization problem is formulated as a LP problem as follows:

$$\begin{aligned}
&\min \|v_1 - v_2\|_{L1} \\
&s.t. \\
&\quad S \cdot v_1 = 0; \quad v_{min} \le v_1 \le v_{max}; \quad v_1[ko1] = 0,\ ko1 \in A_1; \\
&\quad S \cdot v_2 = 0; \quad v_{min} \le v_2 \le v_{max}; \quad v_2[ko2] = 0,\ ko2 \in A_2; \\
&\quad \|w - v_1\|_{L1} = l_1; \quad \|w - v_2\|_{L1} = l_2;
\end{aligned}$$

where w is the wild-type flux distribution, A_1 and A_2 are sets of reactions associated with the deleted genes, and l_i $(i = 1, 2)$ are the optimal solutions of a single optimization problem:

$$\begin{aligned}
&\min \|v - w\|_{L1} \\
&s.t. \\
&\quad S \cdot v = 0; \quad v_{min} \le v \le v_{max}; \quad v_{ko1} = 0,\ ko1 \in A;
\end{aligned}$$

Solving the above optimization problem we receive a measure of similarity between fluxes.

3.2 Growth-Functional Scheme

We hypothesize that the regulation of reactions that are active (different than zero) together across certain media and passive (equal to zero) together across others should be similar. In order to evaluate our hypothesis, we follow and extend the approach of [10], computing genes' activities across 100 randomly generated growth media.

To pursue this possibility we used flux variability analysis [9,25]: for each reaction we computed the maximal and minimal flux values attainable in the space of optimal flux distributions for growth conditions simulating 100 different

[1] We use the distance notion instead of the similarity one both in the KF and GF schemes for sake of clarity and for being consistent with commonly used network topology distances.

growth media. Random growth media were generated by setting limiting values to the uptake reactions independently at random. With probability 0.5, the maximal uptake rate was set to 0, *i.e.* only excretion was allowed. Otherwise, uptake rate was limited to a value chosen uniformly at random in the range [0.01, 5], at a resolution of 0.01. A similar sampling method was used in [26]. In addition, in order to ensure enough variability between media, we switched between aerobic and anaerobic growth media with probability 0.5.

For each generated medium we simulated growth conditions similar to [5] and for each reaction checked if it is active across the current growth media. A reaction is considered active in a given flux distribution if its associated flux is non-zero, namely either its maximum or minimum are different than zero. Active genes were denoted by '0' and nonactive ones by '1'. This way we created for each gene a binary vector of its activity across a series of generated media.

We define a measure of metabolic genes co-activity (MGCA) as the Jaccard coefficient [27] between two binary vectors reflecting metabolic genes' activity. The binary Jaccard coefficient measures the degree of overlap between two sets of values, x_a and x_b, and is computed as the ratio between the number of shared attributes of $\mathbf{x}_a$ **and** $\mathbf{x}_b$ and the number possessed by $\mathbf{x}_a$ **or** $\mathbf{x}_b$:

$$J(x_a, x_b) = \frac{x_a \cap x_b}{x_a \cup x_b} \tag{4}$$

The pseudo-code of the entire procedure is presented in Figure 2.

Algorithm 1: *FindGenesDist*(N)

Input: N: the number of required media.
Output: *results*: matrix *num_genes* × *num_genes* containing the distance between metabolic genes.

```
for k=1..N do
    for each external flux f do
        with probability 0.5, set f = 0;
        otherwise f receives a random value chosen uniformly in [0.01, 5];
    Run FBA to maximize biomass(growth rate)
        and obtain objective value (wild_growth_rate);
    Add constraint: biomass ≥ 0.9 * wild_growth_rate;
    for i=1..num_fluxes do
        Run FBA to maximize flux i, obtain i_max;
        Run FBA to minimize flux i, obtain i_min;
    for each gene g do
        if for one of its related fluxes i_max = i_min = 0 then
            MT[g][k] = 1;
        else
            MT[g][k] = 0;
    for each gene g1 do
        for each gene g2 ≠ g1 do
            results[g1][g2] = Jaccard_coefficient(MT[g1],MT[g2]);
```

Fig. 2. The process for computing the GF-based measure

4 Results

Recall that the metabolic state of an organism is controlled by transcriptional regulation which adjusts gene expression levels according to metabolic demands [6]. Thus the experimental pairwise correlations serve as the true benchmark rod to which we compare the computational measures we compute to find out which is the best, *i.e.* closest to reality.

The first computational similarity measure proposed [3] was based on topological properties of the metabolic network . We start by repeating these experiments and then show how our measure can go beyond topological measures.

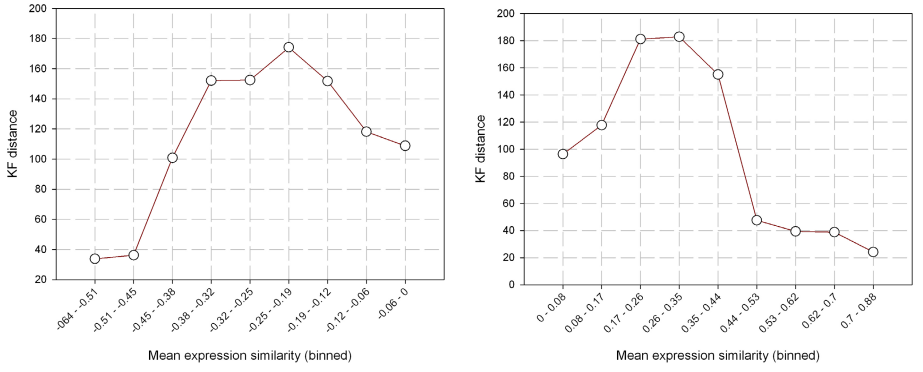
The obtained metabolic network is used to calculate network distance between genes. We define a pair of directly connected metabolic genes as separated by distance 1, and the network distance between genes X and Y is the length of the shortest path from X to Y in the metabolic network. While any metabolite can be used to establish connections between metabolic genes, the relationships established by the common metabolites and cofactors—such as ATP, water or hydrogen—are not likely to connect genes with similar metabolic functions.

In compiling a metabolic network, we consider a subset of metabolites which excludes the most highly connected metabolic species. An exclusion threshold was determined based on the connectivity of the resulting network. A total of the 10 most highly connected metabolites (ATP, ADP, AMP, CO_2, H, H_2O, NADP, NADPH, phosphate and diphosphate), which compose 1% of all metabolites, and their mitochondrial and external analogs were excluded. We also tried to exclude up to the top 3% of all metabolites, however we found out that the general trends described in this paper are not sensitive to the precise choice of the excluded set of metabolites.

We compared the correlation between the gene functional similarity measure and their expression similarity. To this end, we used Rosetta's "compendium" dataset [28] which measures expression profiles of over 6200 *S. cerevisae* ORFs across 287 deletion strains and 13 chemical conditions. In addition, the dataset contains 63 negative control measurements comparing two independent cultures of the same strain. These were used to establish individual error models for each ORF, providing not only the raw intensity and the ratio measurement values for each experimental data point, but also a *p-value* evaluating the significance of change in expression level. The expression similarity measure between ORFs X and Y was computed according to $1 - Spearman\_rank(p_x, p_y)$ where p_x and p_y are expression profile vectors of X and Y, respectively, and the Spearman rank was calculated as in [29].

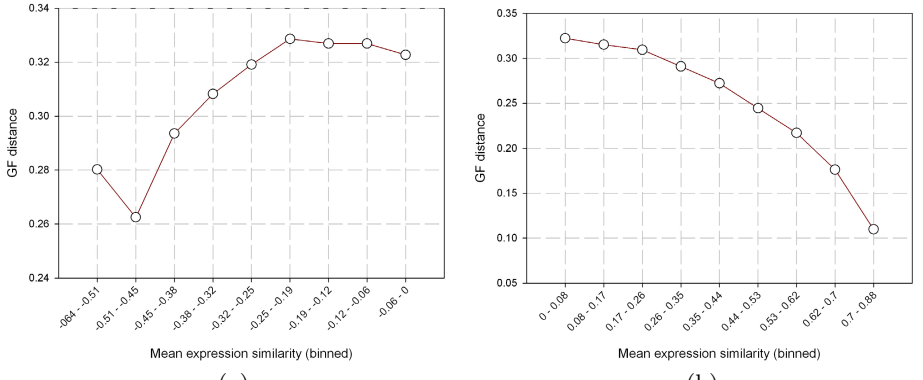
As in [3], we observed that the expression distance increases monotonically with network distance ($R^2 = 0.78$, p-value $= 1.2 \cdot 10^{-8}$), demonstrating that genes closer to each other in the metabolic network tend to have, on average, higher level of coexpression.

Measuring the correlation between the KF-based distance and those based on the expression data we observed (see Figure 3) a moderate correlation (R^2=0.36 in the negatively correlated expressed profiles with a p-value of $8.6 \cdot 10^{-2}$, and $R^2 = 0.45$ in the positively ones with a p-value of $\leq 4.6 \cdot 10^{-2}$). Note that the



(a) (b)

Fig. 3. Correlation between expression levels and genes activities under the KF measure. (a) Negatively expressed pairs. (b) Positively expressed pairs.



(a) (b)

Fig. 4. Correlation between expression levels and genes' activities under the GF measure. (a) Negatively expressed pairs. (b) Positively expressed pairs.

obtained correlation is robust to the initial wild-type metabolic state, as similar correlation levels were observed when starting from different wild-types.

As for the GF-based measure, we observe (see Figure 4) that it exhibits a significant correlation with the expression similarity ($R^2 = 0.78$ in the negatively correlated expressed profiles with a p-value of $5.15 \cdot 10^{-8}$, and $R^2 = 0.94$ in the positively ones with p-value $\leq 1 \cdot 10^{-20}$).

Finally, we observe a significant enhancement of the GF-based measure over the static (topological) metabolic distance indicating that this static distance can explain only partially the demand for common regulation. We use a partial correlation method that describes the relationship between two variables whilst eliminating the effects of another variable on this relationship, namely network distance in our case. Our results show significant partial correlation (R^2=0.65,

with a p-value of $3.8 \cdot 10^{-6}$) between expression levels and our MGCA measure given a metabolic network distance. This higher correlation for our measure supports the fact that the FBA model captures the dynamic metabolic activity of the cell, and that the regulation system indeed works to maximize the growth rate. Moreover, the results stay significant with every thresholds for excluding "currency metabolites" from the metabolic network in the range from 1% to 3%.

In order to evaluate the difference between the MGCA measure and the metabolic network distance measure we analyzed two sets of pairs of genes: one containing pairs of genes that are close under the network distance and distant under the MGCA measure, and vice versa. We observed that the first set is significantly enriched with the GO term *protein biosynthesis* (GO:006412)—25 annotated genes out of 104 resulting in a p-value ≤ 0.001, as well as with the GO term *nucleobase, nucleoside, nucleotide and nucleic acid metabolism* (GO:006139)—40 annotated genes out of 104 also resulting in a p-value ≤ 0.001. An engrossing result was that the complementary set (genes that are close under the MGCA measure but are distant under the network topology measure) showed no significant enrichment, possibly testifying that such functional similarities occur across a broad and homogeneous span of functional annotations.

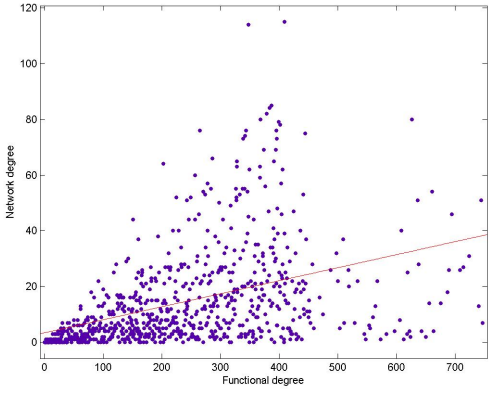


Fig. 5. The correlation between functional centrality and network centrality

Functional enrichments were computed based on the GO-SLIM process annotations [30] for genes. Yeast GO-SLIM annotations were obtained from SGD [31]. For a given set S and a given term t, the functional enrichment score was computed as follows: suppose S has $n(t)$ genes that are annotated with term t (or with a more specific term). Let $p(t)$ be the hypergeometric probability for observing $n(t)$ or more proteins annotated with the term t in a protein subset of size $|S|$. Having found a term t_0 with minimal probability $p(t_0)$, the score was set to the p-value of the enrichment under term t_0, computed by comparing $p(t_0)$ with the analogous probabilities for 10,000 random sets of proteins of size $|S|$.

In addition we looked at the correlation between the network degree of each gene and its functional degree, *i.e.* the number of functionally highly similar (Jaccard coefficient ≤ 0.3) genes (see Figure 5). As we received a significant correlation of $R^2 = 0.4$ with a p-value ≤ 0.001, it seems that network centrality explains (at least in part) functional centrality. Namely, the more alternative

pathways go through a given gene, the more functionally significant it is. We also observed that the correlation is robust to the functionally similarity threshold in the range form 0.01 to 0.3.

5 Conclusions

This paper is the first to show that functional flux-based similarity measures between genes can go beyond previous computational measures based on network topology. We applied two schemes to compute this distance: the knockout-functional (KF) scheme and the growth-functional (GF) scheme. While the former shows a fairly moderate correlation with the experimental results, the latter provides a strong, statistically-significant measure. One possible explanation of this behavior may be that the GF studies probe the natural wild type across a variety of media, whereas the KF method does it in less natural strains and in a sole media. The other reason is the more cumbersome computational method used in the KF case, which is likely to add significant noise to the results obtained.

References

1. Jeong, H., Tombor, B., Albert, R., Oltavi, Z., Barabasi, A.: The large-scale organization of metabolic networks. Nature **407** (2000) 651–654
2. Ravasz, E., Somera, A., Mongru, D., Oltvai, Z., Barabasi, A.: Hierarchical organization of modularity in metabolic networks. Science **297** (2002) 1551–1555
3. Kharchenko, P., Church, G.M., Vitkup, D.: Expression dynamics of a cellular metabolic network. Molecular Systems Biology **1** (2005) E1–E6
4. Chen, L., Vitkup, D.: Predicting genes for orphan metabolic activities using phylogenetic profiles. Genome Biol. **7** (2006) R17
5. Famili, I., Forster, J., Nielsen, J., Palsson, B.Ø.: Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. Proc Natl Acad Sci U S A **100** (2003) 13134–13139
6. Zaslaver, A., Mayo, A., Rosenberg, R., Bashkin, P., Sberro, H., et al: Just-in-time transcription program in metabolic pathways. Nat Genet **36** (2004) 486–491
7. Schuster, S., Dandekar, T., Fell, D.: Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. Trends Biotechnol **17** (1999) 53–60
8. Schuster, S., Klamt, S., Weckwerth, W., Moldenhauer, F., Pfeiffer, T.: Use of network analysis of metabolic systems in bioengineering. Bioprocess and Biosystems Engineering **24** (2002) 363–372
9. Reed, J., Palsson, B.: Genome-scale in silico models of e. coli have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. Genome Res **14** (2004) 1797–1805
10. Bilu, Y., Shlomi, T., Barkai, N., Ruppin, E.: Conservation of expression and sequence of metabolic genes is reflected by activity across metabolic states. PLoS Comp. Bio. (*in press*) (2006)

11. Fell, D., Small, J.: Fat synthesis in adipose tissue. An examination of stoichiometric constraints. Biochem J **238** (1986) 781–786
12. Kauffman, K., Prakash, P., Edwards, J.: Advances in flux balance analysis. Curr Opin Biotechnol **14** (2003) 491–496
13. Price, N.D., Reed, J.L., Palsson, B.Ø.: Genome-scale Models of Microbial Cells: Evaluating the consequences of constraints. Nature Reviews Microbiology **2** (2004) 886–897
14. Varma, A., Palsson, B.: Metabolic capabilities of Escherichia coli: II. Optimal growth patterns. J. Theor. Biol. **165** (1993) 503–522
15. Schilling, C.H., Edwards, J.S., Palsson, B.: Toward metabolic phenomics: analysis of genomic data using flux balances. Biotechnol. Prog **15** (1999) 288–295
16. Fell, D.: Understanding the Control of Metabolism. Portland Press, London (1996)
17. Schilling, C.H., Edwards, J.S., Letscher, D., Palsson, B.Ø.: Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. Biotechnol. Bioeng. **71** (2000) 286–306
18. Vanderbei, R.J.: Linear Programming: Foundations and Extensions. Kluwer Academic Publishers, Boston (1996)
19. Edwards, J., Ibarra, R., Palsson, B.: In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nat Biotechnol **19** (2001) 125–130
20. Badarinarayana, V., Estep, P.W., Shendure, J., Edwards, J., Tavazoie, S., Lam, F., Church, G.M.: Selection analyses of insertional mutants using subgenic-resolution arrays. Nat. Biotechnol. **19** (2001) 1060–1065
21. Duarte, N., Herrgard, M., Palsson, B.Ø.: Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. Genome Res **14** (2004) 1298–1309
22. Segre, D., Vitkup, D., Church, G.: Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. U. S. A. **99** (2002) 15112–15117
23. Shlomi, T., Berkman, O., Ruppin, E.: Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proc. Natl. Acad. Sci. U. S. A. **102** (2005) 7695–7700
24. Kuepfer, L., Sauer, U., Blank, L.M.: Metabolic functions of duplicate genes in Saccharomyces cerevisiae. Genome Res. **15** (2005) 1421–1430
25. Mahadevan, R., Schilling, C.: The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng **5** (2003) 264–276
26. Almaas, E., Oltvai, Z., Barabasi, A.: The activity reaction core and plasticity of metabolic networks. PLoS Comput Biol **1** (2005) e68
27. Salton, G., McGill, M.J.: Introduction to Modern Information Retrieval. McGraw-Hill, New-York (1983)
28. Hughes, T., et. al.: Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. Cell **102** (2000) 109–126
29. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes in C++: the art of scientific computing. Cambridge University Press, Cambridge (2002)
30. Ashburner, M., et. al.: Gene Ontology: tool for the unification of biology. Nat. Genet. **25** (2000) 25–29
31. Issel-Tarver, L., et. al.: Saccharomyces Genome Database. Methods Enzymol **350** (2002) 329–346