RESEARCH ARTICLE

iMTBGO: An Algorithm for Integrating Metabolic Networks with Transcriptomes Based on Gene Ontology Analysis

Zhitao Mao[1,2] and Hongwu Ma[1,*]

[1]*A Key Laboratory of Systems Microbial Biotechnology, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China;* [2]*University of Chinese Academy of Sciences, Beijing 100049, China*

Abstract: *Background*: Constraint-based metabolic network models have been widely used in phenotypic prediction and metabolic engineering design. In recent years, researchers have attempted to improve prediction accuracy by integrating regulatory information and multiple types of "omics" data into this constraint-based model. The transcriptome is the most commonly used data type in integration, and a large number of FBA (flux balance analysis)-based integrated algorithms have been developed.

Methods and Results: We mapped the K_{cat} values to the tree structure of GO terms and found that the K_{cat} values under the same GO term have a higher similarity. Based on this observation, we developed a new method, called iMTBGO, to predict metabolic flux distributions by constraining reaction boundaries based on gene expression ratios normalized by marker genes under the same GO term. We applied this method to previously published data and compared the prediction results with other metabolic flux analysis methods which also utilize gene expression data. The prediction errors of iMTBGO for both growth rates and fluxes in the central metabolic pathways were smaller than those of earlier published methods.

Conclusion: Considering the fact that reaction rates are not only determined by genes/expression levels, but also by the specific activities of enzymes, the iMTBGO method allows us to make more precise predictions of metabolic fluxes by using expression values normalized based on GO.

1. INTRODUCTION

Flux balance analysis is a constraint-based method that uses the metabolic network and an objective function to obtain the optimal flux distribution [1]. It has been successfully used to predict the metabolic phenotypes of gene knockout strains [2-4], as well as growth rates and metabolite production levels [5-7]. In recent years, many researchers have tried to integrate multiple types of "omics" data to improve the predictive power of metabolic models or to define some new restrictive boundaries to narrow the feasible solution space. In these approaches, the integration of data on regulation or "omics" data is the most common means. The transcriptome is the most commonly used data type in integration, and a large number of FBA-based integrated algorithms have been developed. Transcriptome has been integrated into constraint-based models in three ways. Firstly, this has been done by setting an apriori threshold to judge the gene expression state [8, 9]. For example, GIMME [8] directly removes the reactions for which the expression values are lower than the pre-set threshold. Secondly, one can calculate the relative expression thresholds to determine the gene expression status. This approach usually classifies the reactions into high or low expression and treats these as part of additional constraints or an objective function [10-13]. For the classification of high- and low expressions, there are generally two cases. One involves ranking the expression value. For example, EXAMO [11] ranks the expression levels of all genes from high to low, with the top 15% being defined as highly expressed genes and the last 15% being defined as the low expressing genes. Then, these high expression reactions and low expression reactions are treated as a new objective function. Another approach is to determine the state of gene expression by differential expression analysis of expression profiles, such as MADE [12] and Adam [13]. The last approach involves directly introducing transcriptome data into the objective function [14-17] or the constraints [18].

Several of these methods result in a reaction with a low expression level but a high enzyme turnover number (K_{cat}) being filtered (*e.g.* GIMME and iMAT) or strongly constrained (*e.g.* E-Flux and GX-FBA), and currently, there is no method that considers the enzymes turnover numbers. Gene Ontology [19, 20] is the most comprehensive and widely used knowledge base on gene function established by

*Address correspondence to this author at A Key Laboratory of Systems Microbial Biotechnology, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China; Tel: 022-24828756; E-mail: ma_hw@tib.cas.cn

the Gene Ontology Consortium, which covers three different aspects of gene function: molecular function (MF), cellular components (CC), and biological processes (BP). In addition, gene ontology gives the EC number corresponding to the GO term. By including the K_{cat} information corresponding to the EC number given by the Brenda database [21], we found that GO terms with similar functions have a smaller difference in K_{cat} values , and K_{cat} is also relatively close in the same GO term. Based on this observation, we developed the gene ontology-based flux analysis method (iMTBGO) that integrates transcriptome data with metabolic network analysis to make more accurate predictions of metabolic fluxes. This method first identified some marker genes with a stable expression from the published large-scale *E. coli* expression profiles. Subsequently, the expression value of the marker gene in each term was used to modify the upper bounds of each reaction. We used different experimental data (*e.g.* overexpression or knockout) to evaluate the predictive ability of iMTBGO and found that the flux predictions of growth rate and the central metabolic pathways were stable and maintained a low level of error (Fig. **1**).
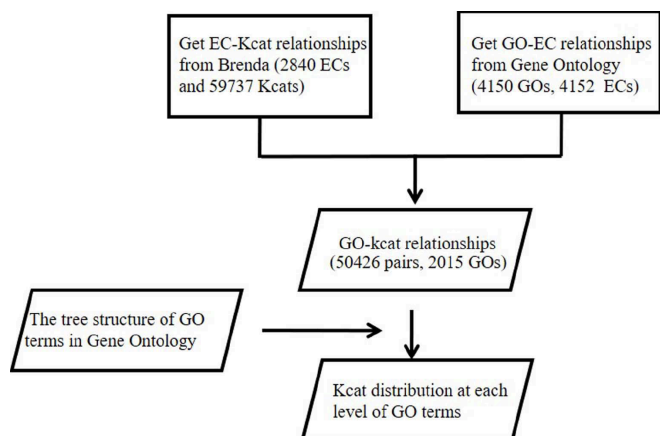

Fig. (1). Flowchart of mapping the K_{cat} values.

2. MATERIALS AND METHODS

2.1. Distribution of K_{cat} Values at Each Level of GO Terms

The K_{cat} values corresponding to all EC numbers were downloaded and extracted from the Brenda database, and the relationship between the GO and EC number was obtained from the Gene Ontology database. K_{cat} values were assigned to each GO term based on the GO-EC and EC- K_{cat} relationships. GO terms were organized hierarchically at different levels in the Gene Ontology database. The three root terms ('molecule function', 'cellular component' and 'biological process') are at level 0, their direct child GO terms are at level 1, and so on. Child terms have more specialized function descriptions than their parent terms [19]. A parent GO term contains K_{cat} values for all its child GO terms. We mainly focused on the GO terms under 'molecule function' (MF), as most EC numbers are assigned to these terms. We used Eq. 1 to calculate the standard deviation of K_{cat} values for each GO term (containing two or more K_{cat} values).

$$\sigma_i = \sqrt{\frac{1}{N} * \sum_{j=1}^{N}(x_{ij} - u_i)^2} \tag{1}$$

Where i represents the i-th GO term in a level, x_{ij} represents the j-th K_{cat} value in i-th GO term, u_i represents the mean of the K_{cat} values in i-th GO term, and N is the number of K_{cat} values.

To test if GO terms with more specific functions (higher level numbers) have similar K_{cat} values, we calculated the average standard deviations for all GO terms in the respective same levels and compared the values at different levels. To test if genes under the same GO term also have similar K_{cat} values, we also shuffled the GO-K_{cat} pairs to simulate the discrete distribution of K_{cat} values when a GO term contained K_{cat} values from other GO terms (1000 times), and ran a t-test statistic with real a K_{cat} distribution.

2.2. The iMTBGO Algorithm

iMTBGO uses the expression ratio obtained from a specific expression profile to correct the upper and lower boundary of each reaction (Fig. **2** dotted line rectangle). When we know the substrate absorption rate (such as glucose absorption rate) or the concentration of an intermediate metabolite under specific conditions, we can assess the true flux corresponding to each reaction *via* the formula $V_{j_{true}} = \frac{V_{glc_{exp}} * V_{j_{prediction}}}{V_{glc_{prediction}}}$. Finally, the optimization problem is converted into the following form, shown in Eqs. 2-8.

$$\text{Max } Z = c^T v \tag{2}$$

$$S * v = 0 \tag{3}$$

$$-ratio_j \leq V_j \leq ratio_j, \text{reversible} \tag{4}$$

$$-ratio_j \leq V_j \leq 0, \text{irreversible} \tag{5}$$

$$0 \leq V_j \leq ratio_j, \text{irreversible} \tag{6}$$

$$ratio_j = \frac{[E]_j}{[E]_{marker}} \tag{7}$$

$$V_{j_{true}} = \frac{V_{glc_{exp}} * V_{j_{prediction}}}{V_{glc_{predicion}}} \tag{8}$$

where $V_{glc_{predicion}}$ is the glucose absorption rate calculated by iMTBGO, $V_{glc_{exp}}$ is the true glucose absorption rate (*e.g.* experimentally-measured phenotype data from Holm *et al.*), $V_{j_{prediction}}$ is the flux of the j-th reaction predicted by iMTBGO that needs to be corrected, $V_{j_{true}}$ is the corrected flux of the j-th reaction, $ratio_j$ is the maximum boundary of V_j calculated from the expression data, and $[E]_{marker}$ represents the expression value of the marker genes in each GO term.

2.3. Determination of Marker Genes Using Gene Expression Profiles in EcoMAC

We directly used the expression profiles in EcoMAC, derived from the literature [22] and standardized, to determine the marker genes. We used the DAVID [23, 24] annotation information to obtain the GO term information for each gene in EcoMAC. Then, we mapped all of the gene ids in EcoMAC to Ecogene [25] and removed the unique ids from EcoMAC. If the different ids in the EcoMAC matrix corresponded to the same id in Ecogene, they were replaced
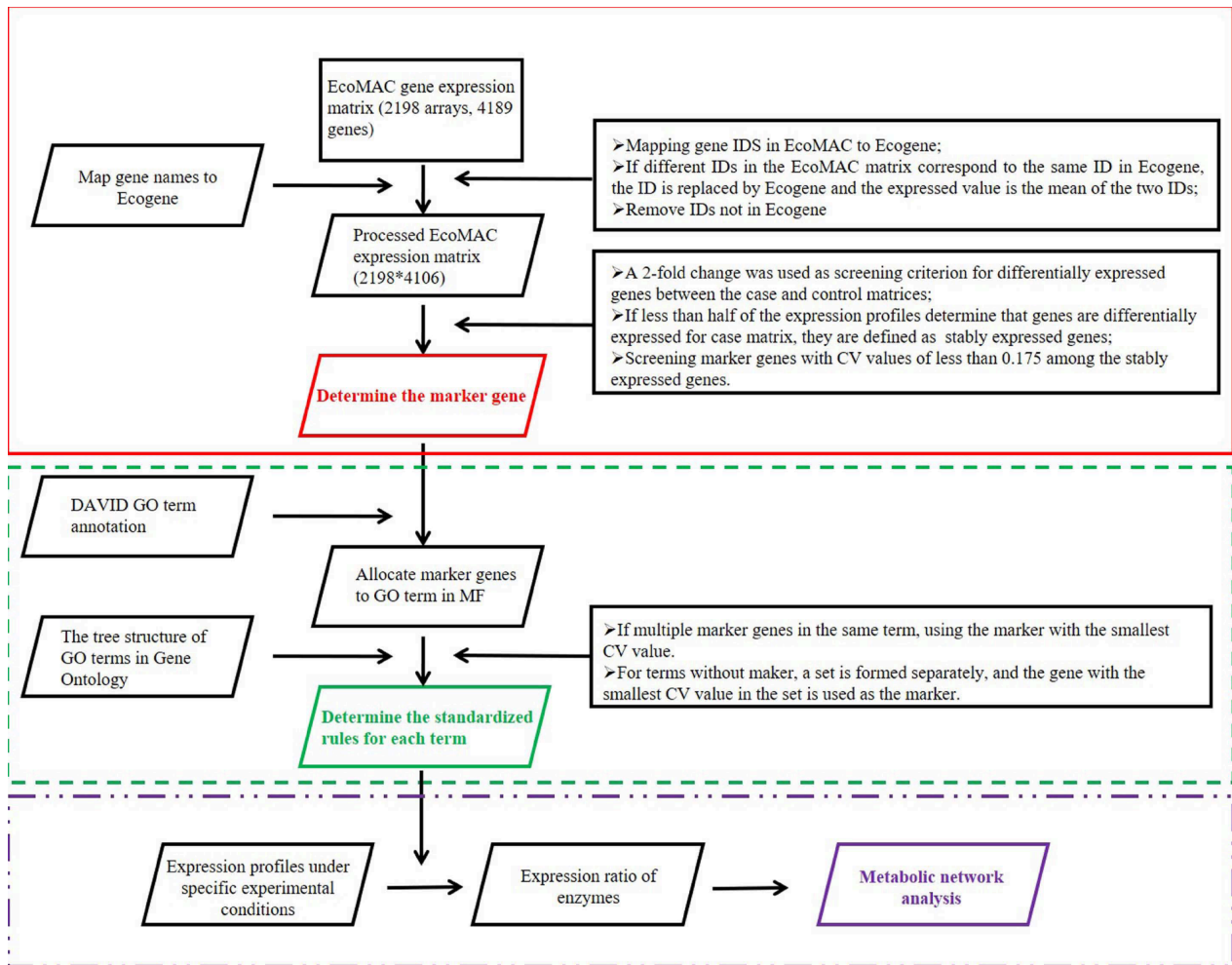
Fig. (2). Shows the flowchart of the iMTBGO algorithm. The solid rectangle represents the procedure of determining the marker genes, the dashed rectangle represents the standardized rules for each term, and the dotted line rectangle represents the calculation of fluxes using the expression profile under specific conditions. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

with the gene id in the Ecogene, and the expression value was the mean of these two ids. Furthermore, we selected the mean expression value of the wild-type strains during the exponential phase in the M9 medium as the control group, and other expression data in EcoMAC as the case group to run the differential expression analysis. We defined genes with fold changes of more than two times and the number of samples exceeded half of the total as unstably expressed genes and others as stably expressed genes. We chose the coefficient of variation (Eq. 9) to obtain the discrete distribution of the genes' expression and finally selected candidates among the stably expressed genes and CV less than specific values (*e.g.* 0.175 in this research) as potential marker genes. The detailed flowchart is shown in Fig. **2**, of a solid rectangle.

$$CV_i = \frac{\sqrt{\frac{1}{P}*\sum_{j=1}^{P}(e_{ij}-u_i)^2}}{u_i} \qquad (9)$$

where i represents the i-th gene in the profiles, e_{ij} represents the expression value of the i-th gene in the j-th profile, u_i represents the mean of the expression values of the i-th gene, and P is the number of expression profiles.

Based on the GO-gene relationship obtained from the DAVID database, we mapped potential marker genes onto individual GO terms. Since some GO terms did not have a marker gene, we relied on the tree structure of GO terms in gene ontology to develop the following standardized rules for different types of terms (Fig. **2** dashed rectangle). First, when a term had a number of potential marker genes, we selected the gene with the smallest CV value among these potential marker genes. Second, for terms without a maker, a single set was formed, and the gene with the smallest CV value in the set was used as the marker.

2.4. Determination of Expression Ratios Using an Expression Profile Under Specific Conditions

It is known from the Michaelis-Menten equation that $V = \frac{V_{\max}[S]}{K_m+[S]}$. When the substrate concentration is saturated, $V_{max} = K_{cat} * [E]$. In the same GO term, K_{cat} is relatively close, and we can reasonably assume that $\frac{V_{i_{max}}}{V_{marker_{max}}} \sim \frac{[E]_i}{[E]_{marker}}$, where marker represents the marker gene in GO term, i represents the i-th gene in the same GO term, $V_{i_{max}}$ represents the maximum flux of i-th gene, and

$[E]_i$ represents the expression value of i-th gene. If we split the entire metabolic network into multiple GO term combinations and assume that $V_{marker_{max}} = 1$ in each combination for computational simplicity, then $V_{i_{max}} \sim \frac{[E]_i}{[E]_{marker}}$, that is $V_{i_{max}} \sim$ ratio$_i$. Therefore, we used the standardized rules of each term to obtain the expression ratio of genes under specific experimental conditions. If an enzyme consists of multiple subunits encoded by different genes, the minimum value of these genes was used as the expression ratio of the reaction (Eq. 10). If a reaction can be independently catalyzed by multiple enzymes, the corresponding expression ratio for this enzyme is the sum of expression ratios of the subunits, as shown in Eq. 11.

$$\text{ratio}_j = \min_{i \in n} \left(\frac{g_i}{\text{marker}(\text{Term}_i)} \right) ,$$

$$\text{Enzymes} = (g_1 \text{ AND } g_2 \text{ AND } g_3 \ldots g_n) \tag{10}$$

$$\text{ratio}_j = \sum_i^n \left(\frac{g_i}{\text{marker}(\text{Term}_i)} \right) ,$$

$$\text{Enzyme} = (g_1 \text{ OR } g_2 \text{ OR } g_3 \ldots g_n) \tag{11}$$

where marker (Term$_i$) indicates the expression value of the marker gene in the term containing the i-th gene. If an enzyme consists of multiple subunits, g_i represents the expression value of the gene encoding the i-th subunit, and otherwise represents the expression value of the gene encoding the i-th enzyme.

2.5. Evaluation of the Predictive Accuracy of iMTBGO

Microarray and experimentally-measured phenotype data for *E. coli* from Holm *et al.* [26] (wild-type, overexpressed ATPase, and overexpressed NADH oxidase) and Ishii *et al.* (two wild-type strains: WT0.5, WT0.7, and five mutant strains: Δpgm, Δpgi, $\Delta gapC$, Δzwf and Δrpe) [27] were used to solve the metabolic fluxes under specific conditions. For the datasets of Holm *et al.*, the raw data (GSE20374) was first read into R using the affy package. Then, the RMA algorithm was applied for a set of replicates for background correction and normalization. Since the data of Holm *et al.* involved three replicates, we used the mean of the three sets of experiments as the expression value. Because the datasets provided by Ishii *et al.* did not give the corresponding GEO id, so we directly downloaded the chip signal file from the literature, and took the logarithm of the "normalized intensity of sample" column as the expression value.

Since the datasets of Ishii and Holm [26, 27] provided experimentally-measured phenotype data of the central metabolic pathways, we used the growth rate error [28] (Eq. 12) and flux error measurement [29] (Eq. 13) to evaluate the prediction.

$$\text{Normalized growth rate error} = \frac{|v_{growth,sim} - v_{growth,exp}|}{v_{growth,exp}} \tag{12}$$

$$\text{Normalized fluxes error} = \frac{\sqrt{\sum_i^n (v_{sim,i} - v_{exp,i})^2}}{\sqrt{\sum_i^n v_{exp,i}^2}} \tag{13}$$

where $v_{growth,sim}$ and $v_{growth,exp}$ represent the predicted growth rate and the experimentally measured growth rate, respectively. n is the number of experiments, $v_{sim,i}$ is the flux of the i-th reaction in the central metabolic pathways and $v_{exp,i}$ is the measured flux of i-th reaction. We also carried out sensitivity analysis to determine the optimal CV value by changing the range of CV values according to the literature [29]. We then performed robustness analysis by using the optimal CV value to determine the sensitivity of fluxes to gene expression levels. In the robustness analysis, we shuffled each column of the expression matrix (x) to determine a random matrix (r), and then we obtained the noisy expression matrix (y) according to Eq. 14, as published previously [29]. The value of λ was consistent with the literature [28], which set the values of 0, 0.25, 0.5, 0.75, and 1 for λ.

$$y = x + \lambda(r - x), \lambda \in [0, 1] \tag{14}$$

We obtained the complete codes of E-Flux, RELATCH [15], iMAT [10], GIMME, and GX-FBA [16] from the literature [29]. In all simulation experiments, the genome-scale metabolic model iJO1366 [30] was used in conjunction with the COBRA toolbox 2.0 in MATLAB software to solve the optimal solution by glpk/qpng, in which the objective function is the maximization of biomass.

For the Ishii datasets, the oxygen uptake rate was also used as a constraint, which was set as the measured value when running the simulation.

3. RESULTS

3.1. The Distribution of K_{cat} in the GO Terms

We obtained 4178 pairs of GO-EC numbers from the Gene Ontology database, including 4150 GO terms and 3950 EC numbers (Supplementary file). In addition, we obtained 59615 pairs of EC-K_{cat} information with 2840 EC numbers from the Brenda database (Supplementary file). We mapped the corresponding K_{cat} to each GO term through a common EC number, and finally obtained 44118 pairs of GO-K_{cat} data including 2043 GO terms (Supplementary file). We also used child nodes to assign K_{cat} values to the terms without K_{cat} values level by level, and re-updated 463 GO terms. Only 413 pairs of GO-K_{cat} records belonged to CC (4 terms) and BP (120 terms), so these two types of terms were not considered in the subsequent analysis. We then used the standard deviation to evaluate the distribution of K_{cat} at each level of MF and found that except for levels 10 and 6, the larger the level, the smaller the mean value of the standard deviation (Table **1**). It can be inferred that GO terms with similar functions have a smaller difference in K_{cat} values. Furthermore, we also found that the standard deviation of the GO term of real K_{cat} was significantly smaller than that of the simulation (nearly 90% of the simulations when $p < 0.001$). When the *p*-value was set to 0.05, the standard deviation of the real K_{cat} was smaller than that of all simulations, which indicated that K_{cat} values in the same GO term were relatively similar.

3.2. Determining Marker Genes and Standardization Rules

We found that GO annotation information for 126 genes of 4190 genes in EcoMAC could not be obtained using the

Table 1.   The distribution of K$_{cat}$ values according to GO levels.

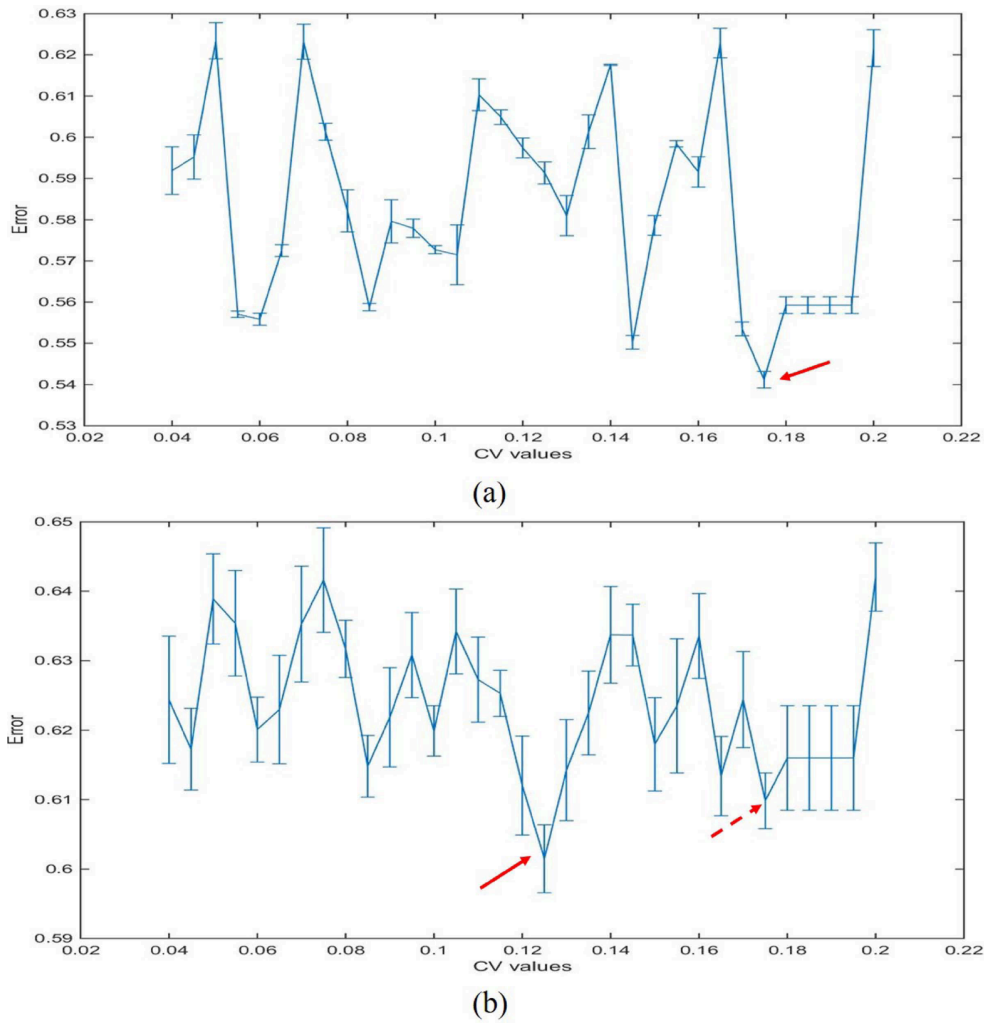| GO Level	Mean of the Standard Deviation of K$_{cat}$	Numbers of GO Terms with K$_{cat}$
0	959774.44	1
1	242600.75	4
2	147651.83	24
3	43151.06	129
4	16857.15	369
5	12258.59	1181
6	19380.66	364
7	11525.28	120
8	387.64	24
9	6.25	5
10	20.78	2

Fig. (3). Shows sensitivity analysis of iMTBGO for the CV parameter using the datasets of Holm *et al.* (**a**) and Ishii *et al.* (**b**). Each average error and its variation (the error bars represent the standard error of the mean) represent the error distribution across all conditions in one dataset. The solid red arrow indicates the optimal CV and the dashed red arrow indicates the suboptimal CV for the datasets of Ishii *et al.* (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).
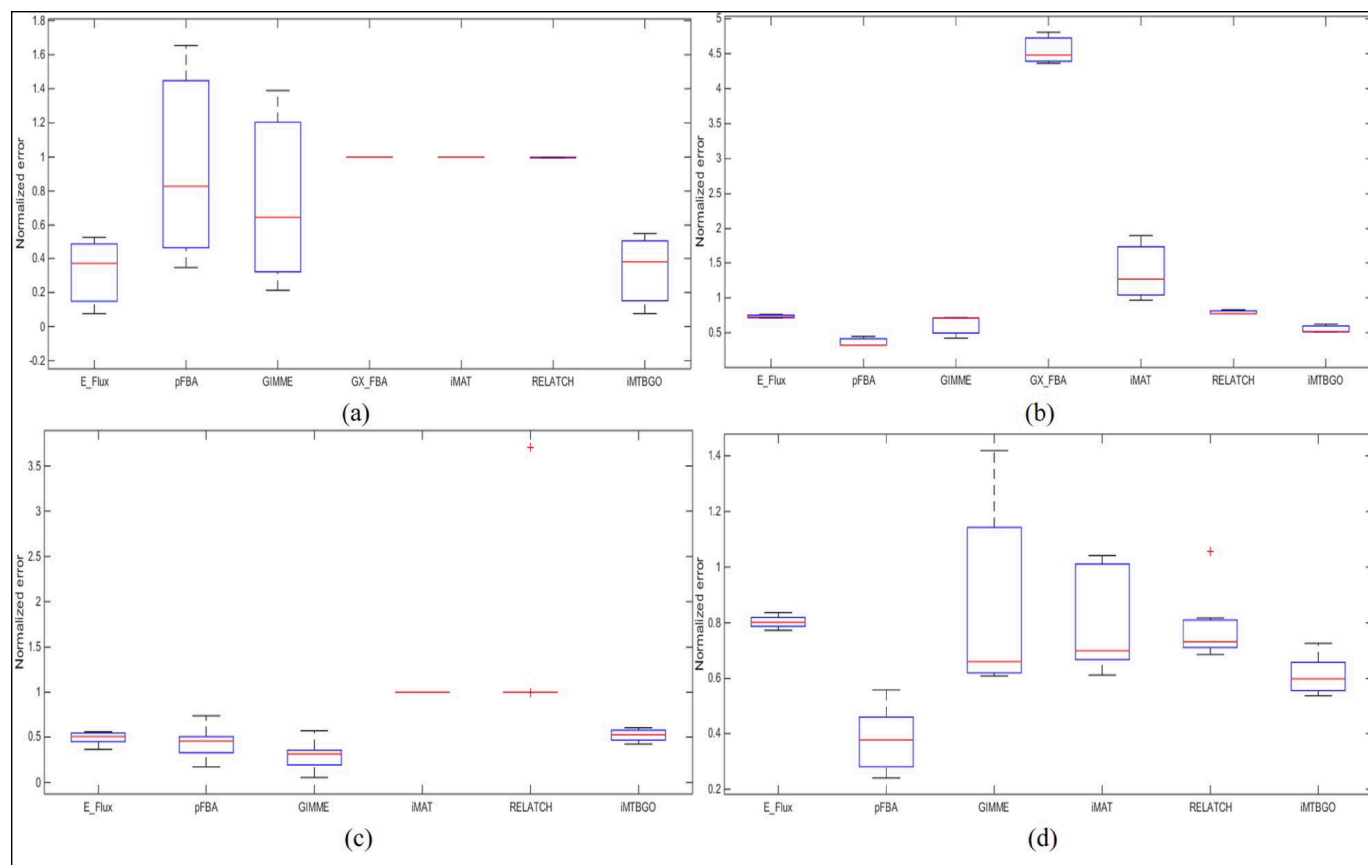
Fig. (4). Shows the distribution of the relative error in the prediction of growth rates (**a**, **c**) or fluxes in the central metabolic pathways (**b**, **d**) for each method across multiple perturbations for the datasets of Holm *et al.* (**a**, **b**) and Ishii *et al.* (**c**, **d**). Each box plot represents the distribution of errors for all perturbations in one dataset. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

DAVID database, 35 of the 126 genes were only found in EcoMAC, and the remaining 91 genes used outdated gene ids. After excluding these, we obtained a final set of 4106 genes. We carried out sensitivity analysis to find the optimal value of CV to determine the marker genes and standardization rules. The results showed that the average value of the growth error changes with the change of CV, as well as that the variation is monotonic with respect to the CV value, and the average error changes little with the change of CV (Fig. **S1**). When considering the fluxes in the central metabolic pathways, we found that the optimal values of CV were 0.175 and 0.125 in the datasets of the Holm *et al.* and Ishii *et al.* respectively. Furthermore, 0.175 was also a suboptimal value for the dataset of Ishii *et al*. (Fig. **3**). When considering the ratio of the number of genes in a GO term to the total number of genes (Fig. **S2 a**) and the proportion of terms containing potential marker genes to the total numbers of terms in the different CVs (Fig. **S2 b**), we found that the proportion occupied by CV at 0.175 was basically saturated. Based on these findings, we finally selected the genes with CV less than 0.175, which resulted in 1083 potential marker genes. We obtained 21411 pairs of GO-gene information from the DAVID database, including 3675 GO terms and 3815 genes (Supplementary file). Using the GO-gene relationship, we mapped these potential marker genes onto individual GO terms of MF, encompassing 1705 GO terms and 1235 genes are annotated into these GO terms (Supplementary file). We used the gene with the smallest CV value as the final marker

gene with multiple potential markers, and finally determined 1209 unique marker genes. For terms without a marker, we used b3033 (yqiB) as the marker gene, since it had the lowest CV value among these terms.

3.3. Evaluation of the Predictive Accuracy of iMTBGO

We found that the accuracy of iMTBGO for growth rate prediction was the highest under different experimental conditions among several methods for the datasets of Holm *et al.* (Fig. **4a**), but did not perform the same for the datasets of Ishii *et al.* (Fig. **4c**), which may have a certain relationship to the use and omission of the optimal CV. For fluxes in the central metabolic pathways, iMTBGO still maintained lower flux errors than other methods (Fig. **4b, d**). It was found to be superior to several other methods for integrating the transcriptome data, but not superior to pFBA, especially for the datasets of Ishii *et al.* (Fig. **4d**). In addition, iMAT and RELACTH always remained unchanged for growth rate prediction (Fig. **4a**, **c**) and GX-FBA was not able to calculate the fluxes for partial expression profiles (Fig. **4c**, **d**). Compared with these methods, except for the way of integrating transcriptome data, the biggest difference is that the iMTBGO considers the enzymes turnover numbers. By analyzing GO terms, the genes stably expressed in the same GO term are used to standardize the genes in the term, eliminating the constraint error caused by low enzyme expression in conjunction with a high turnover number, and the flux is only related to the enzyme expression level. This is why the iMTBGO method is superior to several other methods that

integrate the transcriptome data. As the central metabolic pathways are more heavily regulated at post-transcriptional levels [31-33], transcript levels are not suitable to represent the final enzyme activity levels, which leads to problems in the upper constraints of fluxes in the central metabolic pathways, as observed in our method and finally results in incorrect prediction compared with pFBA. Subsequently, robustness analysis was carried out to evaluate the effects of noise in the expression data, and we found that iMTBGO to be robust due to a relatively smooth increase in the error with an increase in the noise level for the datasets of Holm *et al.* and Ishii *et al.* (Fig. **S3**). However, it was also evident that the error of several noise points was reduced, which indirectly indicated that iMTBGO was dependent on the expression profiles, and it was necessary to provide more accurate transcriptomic data or directly provide proteomic data if possible. In conclusion, this GO-based metabolic flux analysis method is relatively reliable and can be used to estimate flux changes in metabolic networks with comparatively good accuracy.

CONCLUSION

We developed the new method iMTBGO that integrates metabolic network and transcriptome data. This method first introduces Gene Ontology information into metabolic network analysis by obtaining the expression ratios of the enzymes by normalizing the marker genes in the same GO term. iMTBGO maintained higher predictive accuracy in some experiments and always maintained a lower level of flux errors in the central metabolic pathways than other tested methods. The iMTBGO method provides a feasible process for integrating metabolic networks with omics datasets, which offers new broader ideas for using expression profiles. Although there are still some problems when the method integrates metabolic networks with transcriptome data, it cannot be denied that this integrated analysis can improve the prediction process under certain experimental conditions.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No Animals/Humans were used for studies that are the basis of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The data supporting the findings of the article is available in the iMTBGO database at https://github.com/maozhitao/iMTBGO.

FUNDING

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's website along with the published article.

REFERENCES

[1]   Llaneras, F.; Picó, J. Stoichiometric modelling of cell metabolism. *J. Biosci. Bioeng.,* **2008**, *105*(1), 1-11.
[http://dx.doi.org/10.1263/jbb.105.1] [PMID: 18295713]

[2]   Fong, S.S.; Palsson, B.O. Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nat. Genet.,* **2004**, *36*(10), 1056-1058.
[http://dx.doi.org/10.1038/ng1432] [PMID: 15448692]

[3]   Famili, I.; Forster, J.; Nielsen, J.; Palsson, B.O. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc. Natl. Acad. Sci. USA,* **2003**, *100*(23), 13134-13139.
[http://dx.doi.org/10.1073/pnas.2235812100] [PMID: 14578455]

[4]   Edwards, J.S.; Palsson, B.O. Metabolic flux balance analysis and the *in silico* analysis of *Escherichia coli* K-12 gene deletions. *BMC Bioinformatics,* **2000**, *1*, 1.
[http://dx.doi.org/10.1186/1471-2105-1-1] [PMID: 11001586]

[5]   Beste, D.J.; Hooper, T.; Stewart, G.; Bonde, B.; Avignone-Rossa, C.; Bushell, M.E.; Wheeler, P.; Klamt, S.; Kierzek, A.M.; McFadden, J. GSMN-TB: A web-based genome-scale network model of *Mycobacterium tuberculosis* metabolism. *Genome Biol.,* **2007**, *8*(5), R89.
[http://dx.doi.org/10.1186/gb-2007-8-5-r89] [PMID: 17521419]

[6]   Varma, A.; Palsson, B.O. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.,* **1994**, *60*(10), 3724-3731.
[PMID: 7986045]

[7]   Edwards, J.S.; Ibarra, R.U.; Palsson, B.O. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.,* **2001**, *19*(2), 125-130.
[http://dx.doi.org/10.1038/84379] [PMID: 11175725]

[8]   Becker, S.A.; Palsson, B.O. Context-specific metabolic networks are consistent with experiments. *PLOS Comput. Biol.,* **2008**, *4*(5), e1000082.
[http://dx.doi.org/10.1371/journal.pcbi.1000082]   [PMID: 18483554]

[9]   Åkesson, M.; Förster, J.; Nielsen, J. Integration of gene expression data into genome-scale metabolic models. *Metab. Eng.,* **2004**, *6*(4), 285-293.
[http://dx.doi.org/10.1016/j.ymben.2003.12.002]   [PMID: 15491858]

[10]   Zur, H.; Ruppin, E.; Shlomi, T. iMAT: An integrative metabolic analysis tool. *Bioinformatics,* **2010**, *26*(24), 3140-3142.
[http://dx.doi.org/10.1093/bioinformatics/btq602]   [PMID: 21081510]

[11]   Rossell, S.; Huynen, M.A.; Notebaart, R.A. Inferring metabolic states in uncharacterized environments using gene-expression measurements. *PLOS Comput. Biol.,* **2013**, *9*(3), e1002988.
[http://dx.doi.org/10.1371/journal.pcbi.1002988]   [PMID: 23555222]

[12]   Jensen, P.A.; Papin, J.A. Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics,* **2011**, *27*(4), 541-547.
[http://dx.doi.org/10.1093/bioinformatics/btq702]   [PMID: 21172910]

[13]   Töpfer, N.; Jozefczuk, S.; Nikoloski, Z. Integration of time-

resolved transcriptomics data with flux-based methods reveals stress-induced metabolic adaptation in *Escherichia coli. BMC Syst. Biol.,* **2012**, *6*, 148.
[http://dx.doi.org/10.1186/1752-0509-6-148] [PMID: 23194026]

[14]   Lee, D.; Smallbone, K.; Dunn, W.B.; Murabito, E.; Winder, C.L.; Kell, D.B.; Mendes, P.; Swainston, N. Improving metabolic flux predictions using absolute gene expression data. *BMC Syst. Biol.,* **2012**, *6*, 73.
[http://dx.doi.org/10.1186/1752-0509-6-73] [PMID: 22713172]

[15]   Kim, J.; Reed, J.L. RELATCH: Relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations. *Genome Biol.,* **2012**, *13*(9), R78.
[http://dx.doi.org/10.1186/gb-2012-13-9-r78] [PMID: 23013597]

[16]   Navid, A.; Almaas, E. Genome-level transcription data of *Yersinia pestis* analyzed with a new metabolic constraint-based approach. *BMC Syst. Biol.,* **2012**, *6*, 150.
[http://dx.doi.org/10.1186/1752-0509-6-150] [PMID: 23216785]

[17]   Collins, S.B.; Reznik, E.; Segrè, D. Temporal expression-based analysis of metabolism. *PLOS Comput. Biol.,* **2012**, *8*(11), e1002781.
[http://dx.doi.org/10.1371/journal.pcbi.1002781]   [PMID: 23209390]

[18]   Colijn, C.; Brandes, A.; Zucker, J.; Lun, D.S.; Weiner, B.; Farhat, M.R.; Cheng, T.Y.; Moody, D.B.; Murray, M.; Galagan, J.E. Interpreting expression data with metabolic flux models: Predicting *Mycobacterium tuberculosis* mycolic acid production. *PLOS Comput. Biol.,* **2009**, *5*(8), e1000489.
[http://dx.doi.org/10.1371/journal.pcbi.1000489]   [PMID: 19714220]

[19]   Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; Harris, M.A.; Hill, D.P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J.C.; Richardson, J.E.; Ringwald, M.; Rubin, G.M.; Sherlock, G. The Gene Ontology Consortium. Gene ontology: Tool for the unification of biology. *Nat. Genet.,* **2000**, *25*(1), 25-29.
[http://dx.doi.org/10.1038/75556] [PMID: 10802651]

[20]   The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still Going strong. *Nucleic Acids Res.,* **2019**, *47*(D1), D330-D338.
[http://dx.doi.org/10.1093/nar/gky1055] [PMID: 30395331]

[21]   Placzek, S.; Schomburg, I.; Chang, A.; Jeske, L.; Ulbrich, M.; Tillack, J.; Schomburg, D. BRENDA in 2017: New perspectives and new tools in BRENDA. *Nucleic Acids Res.,* **2017**, *45*(D1), D380-D388.
[http://dx.doi.org/10.1093/nar/gkw952] [PMID: 27924025]

[22]   Carrera, J.; Estrela, R.; Luo, J.; Rai, N.; Tsoukalas, A.; Tagkopoulos, I. An integrative, multi-scale, genome-wide model reveals the phenotypic landscape of Escherichia coli. *Mol. Syst. Biol.,* **2014**, *10*, 735.
[http://dx.doi.org/10.15252/msb.20145108] [PMID: 24987114]

[23]   Huang, W.; Sherman, B.T.; Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.,* **2009**, *37*(1), 1-13.
[http://dx.doi.org/10.1093/nar/gkn923] [PMID: 19033363]

[24]   Huang, W.; Sherman, B.T.; Lempicki, R.A. Systematic and integra-

tive analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.,* **2009**, *4*(1), 44-57.
[http://dx.doi.org/10.1038/nprot.2008.211] [PMID: 19131956]

[25]   Zhou, J.; Rudd, K.E. EcoGene 3.0. *Nucleic Acids Res.,* **2013**, *41*(Database issue), D613-D624.
[PMID: 23197660]

[26]   Holm, A.K.; Blank, L.M.; Oldiges, M.; Schmid, A.; Solem, C.; Jensen, P.R.; Vemuri, G.N. Metabolic and transcriptional response to cofactor perturbations in *Escherichia coli. J. Biol. Chem.,* **2010**, *285*(23), 17498-17506.
[http://dx.doi.org/10.1074/jbc.M109.095570] [PMID: 20299454]

[27]   Ishii, N.; Nakahigashi, K.; Baba, T.; Robert, M.; Soga, T.; Kanai, A.; Hirasawa, T.; Naba, M.; Hirai, K.; Hoque, A.; Ho, P.Y.; Kakazu, Y.; Sugawara, K.; Igarashi, S.; Harada, S.; Masuda, T.; Sugiyama, N.; Togashi, T.; Hasegawa, M.; Takai, Y.; Yugi, K.; Arakawa, K.; Iwata, N.; Toya, Y.; Nakayama, Y.; Nishioka, T.; Shimizu, K.; Mori, H.; Tomita, M. Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science,* **2007**, *316*(5824), 593-597.
[http://dx.doi.org/10.1126/science.1132067] [PMID: 17379776]

[28]   Motamedian, E.; Mohammadi, M.; Shojaosadati, S.A.; Heydari, M. TRFBA: An algorithm to integrate genome-scale metabolic and transcriptional regulatory networks with incorporation of expression data. *Bioinformatics,* **2017**, *33*(7), 1057-1063.
[http://dx.doi.org/10.1093/bioinformatics/btw772]   [PMID: 28065897]

[29]   Machado, D.; Herrgård, M. Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLOS Comput. Biol.,* **2014**, *10*(4), e1003580.
[http://dx.doi.org/10.1371/journal.pcbi.1003580]   [PMID: 24762745]

[30]   Orth, J.D.; Conrad, T.M.; Na, J.; Lerman, J.A.; Nam, H.; Feist, A.M.; Palsson, B.O. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism-2011. *Mol. Syst. Biol.,* **2011**, *7*, 535.
[http://dx.doi.org/10.1038/msb.2011.65] [PMID: 21988831]

[31]   Daran-Lapujade, P.; Jansen, M.L.A.; Daran, J.M.; van Gulik, W.; de Winde, J.H.; Pronk, J.T. Role of transcriptional regulation in controlling fluxes in central carbon metabolism of *Saccharomyces cerevisiae.* A chemostat culture study. *J. Biol. Chem.,* **2004**, *279*(10), 9125-9138.
[http://dx.doi.org/10.1074/jbc.M309578200] [PMID: 14630934]

[32]   Daran-Lapujade, P.; Rossell, S.; van Gulik, W.M.; Luttik, M.A.H.; de Groot, M.J.L.; Slijper, M.; Heck, A.J.R.; Daran, J.M.; de Winde, J.H.; Westerhoff, H.V.; Pronk, J.T.; Bakker, B.M. The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels. *Proc. Natl. Acad. Sci. USA,* **2007**, *104*(40), 15753-15758.
[http://dx.doi.org/10.1073/pnas.0707476104] [PMID: 17898166]

[33]   Kochanowski, K.; Sauer, U.; Chubukov, V. Somewhat in control-the role of transcription in regulating microbial metabolic fluxes. *Curr. Opin. Biotechnol.,* **2013**, *24*(6), 987-993.
[http://dx.doi.org/10.1016/j.copbio.2013.03.014]   [PMID: 23571096]