RESEARCH

Open Access

CrossMark

Integrating transcriptional activity in genome-scale models of metabolism

Daniel Trejo Banos[1], Pauline Trébulle[1,2] and Mohamed Elati[1,3*]

Abstract

Background: Genome-scale metabolic models provide an opportunity for rational approaches to studies of the different reactions taking place inside the cell. The integration of these models with gene regulatory networks is a hot topic in systems biology. The methods developed to date focus mostly on resolving the metabolic elements and use fairly straightforward approaches to assess the impact of genome expression on the metabolic phenotype.

Results: We present here a method for integrating the reverse engineering of gene regulatory networks into these metabolic models. We applied our method to a high-dimensional gene expression data set to infer a background gene regulatory network. We then compared the resulting phenotype simulations with those obtained by other relevant methods.

Conclusions: Our method outperformed the other approaches tested and was more robust to noise. We also illustrate the utility of this method for studies of a complex biological phenomenon, the diauxic shift in yeast.

Keywords: Inference and interrogation of regulatory network, Metabolic modeling, *Saccharomyces cerevisiae*

Background

The modeling of biological systems has come a long way for gene regulation, signaling networks and metabolism, but even the most cutting-edge models still focus on one subsystem at the time. The integration of the many subsystems that function together, with the development of modeling paradigms, is the next step in the process, and promising results have already been obtained [1]. For example, [2], constructed a whole-cell model by connecting 28 individual models, one for each of the relevant cell functions. The resulting model included more than 1200 experimentally observed parameters. Impressive as it is, the development of this model required a huge effort for a single organism. We aimed to develop a general methodology that can be adapted to different organisms very easily through minor modifications. We aimed to retain as much

information as possible concerning external and internal effects on genotype-phenotype interactions. For example, computational techniques have recently been used to optimize the yield of substrates produced by microorganisms for industry [3] and to study gene-metabolism interactions in medicine [4].

We focus here on the integration between metabolic models and gene regulatory networks for studies of growth phenotypes. Metabolic models represent the chemical reactions required for growth and sustenance [5], whereas gene regulatory networks comprise the biological programs responsible for regulating cell function [6]. We aimed to use data analysis and mathematical modeling tools to improve both the quantity and quality of biological hypotheses relating to these two subsystems.

Related approaches include: pFBA [7] which involves two-level optimization together with post-processing and the detection of redundant fluxes, E-flux [8] in which the linear constraints on fluxes are derived from gene expression data for control and a specific conditions, GIMME [9], which uses gene expression data and a regulatory metabolic objective to detect inconsistencies

*Correspondence: mohamed.elati@univ-lille.fr
[1]UMR 8030 Génomique Métabolique / Laboratoire iSSB CEA-CNRS-UEVE, Genopole campus 1, 5 rue Henri Desbruères, 91030 Cedex Évry, France
[3]Université Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL - Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille, France
Full list of author information is available at the end of the article

Trejo Banos *et al. BMC Systems Biology* 2017, **11**(Suppl 7):134

Page 82 of 174

in fluxes, and iFBA [10], in which a kinetic model of *E. coli* catabolite repression has been integrated into a simplified metabolic model. The iFBA approach requires the setting of a number of Ordinary Differential Equations(ODE) with their kinetic parameters, which decreases the generality of the model. Another integrative approach is PROM [11] in which gene expression data sets are used to compute the conditional probability of an enzyme being expressed given that its regulators are perturbed, these probabilities then being used to constrain a flux balance analysis model. Similarly, TRFBA uses gene expression data and a piece-wise linear model to formulate an optimization program accounting for gene expression [12]. One of the main drawbacks of all these previously described methods is the need to determine which TFs regulate each gene. These approaches are therefore dependent on the quality of the curated network.

We used a statistical reverse engineering method, hLICORN [13], to infer the targets of a given set of regulators at the genome scale. We then assessed the effect of a regulator on its inferred targets in a particular data set, using the CoRegNet [14] tool, which has functions for scoring the activating or repressing effects of a regulator. The derived score, or "influence", represents the transcriptional state of the cell and forms the basis for posterior integration with metabolic models. CoRegNet allows prior knowledge from various sources to be integrated into the model, in accordance with the recommendations of the DREAM5 consortium [15]. Despite the many and varied publications on gene regulatory network inference [15], few efforts have been made to integrate these inference methods into other systems biology tools.

We based our metabolic analysis on phenotype simulations. We used a well-documented model of yeast metabolism iTO977 [16]. We assembled the inferred gene regulatory and metabolic model together in a rational manner, to simulate growth phenotype and exchange fluxes in an algorithm that we call CoRegFlux. We tested our solution against other state-of-the art methods in a rigorous experimental setting for model benchmarking and comparison [17].

Methods

The CoRegFlux workflow can be summarized as follows: inference of the gene regulatory network from transcriptomic data, network interrogation to predict enzyme activity in a given context and, finally, adjustment of the metabolic model for phenotype simulation. The complete workflow is presented in Fig. 1 along with a step-by-step description for a case study in *S. cerevisae* in the sections below, data and source code can be downloaded from http://github.com/i3bionet/CoRegFlux.
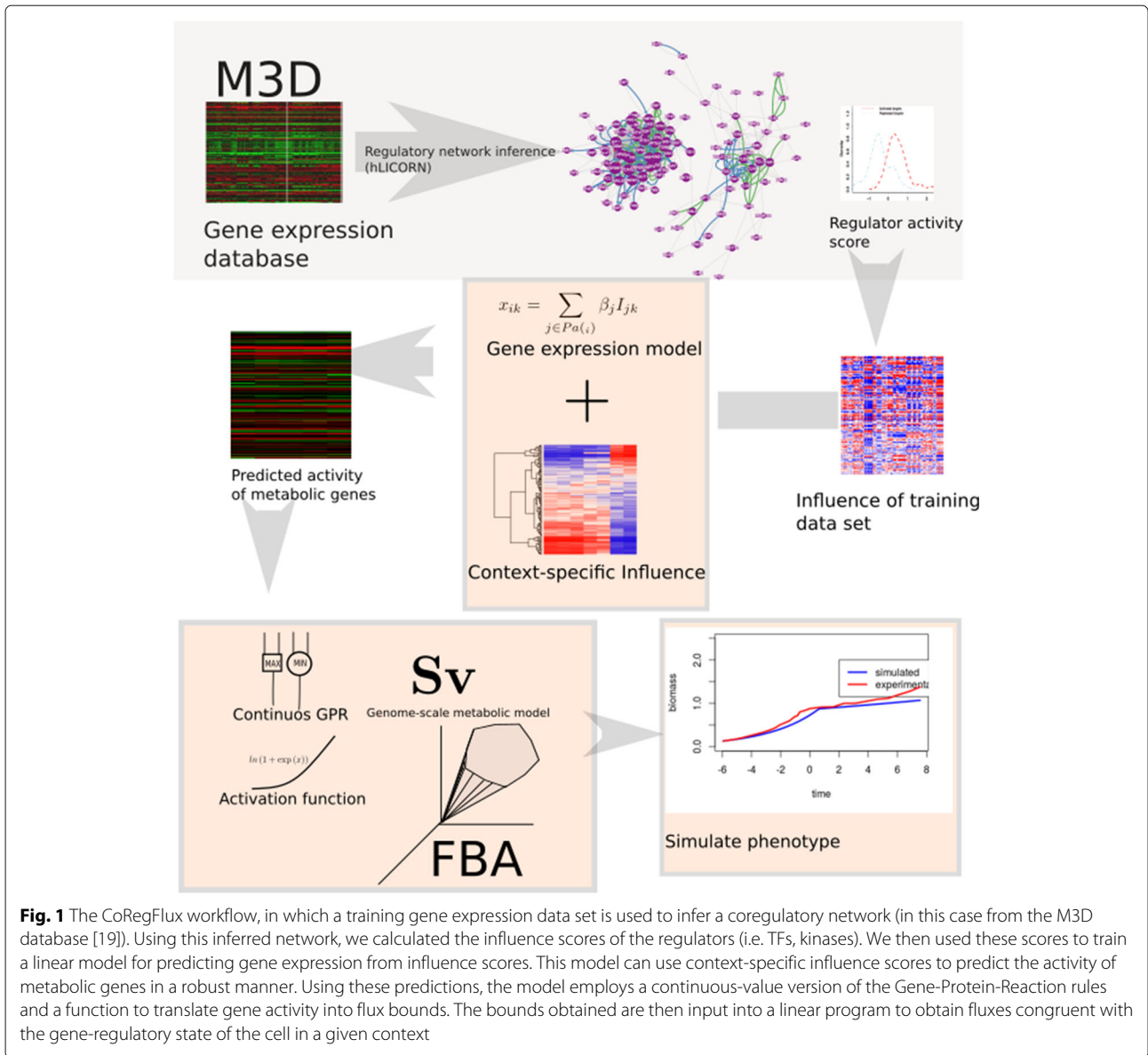
Genetic regulatory network inference

The first step of our algorithm is the inference of a genome-scale regulatory network. This network captures the interactions between regulators (TF and/or kinases) and target genes, which, in our case, encode metabolic enzymes.

For this purpose, we use CoRegNet [14], a Bioconductor package suitable for reverse-engineering and analyisis of large biological networks. Briefly, CoRegNet is a workflow that use the algorithm [13, 18] to mine candidates GRNs set of co-activators and co-inhibitors for each genes. h-LICORN splits genes into regulator and target sets, then discretizes gene expression on the basis of a specified threshold and uses a frequent itemset mining algorithm to find the regulatory elements for each target. In a second step, it determines for each gene the best sets among those candidates by running a regression model. The continuous data can be used alone to refine the original network by selecting for each target the gene regulatory network (GRN) with the best R^2 score based on the linear model used to estimate the expression. However, CoRegNet can also refine GRNs by incorporating evidence into the network using an integrative selection algorithm and applies it to the selection of local GRN models. Each GRN is scored by the inference method h-LICORN and by each of the integrated dataset. Following this, to each GRN is associated as many scores as they are integrated regulatory and cooperative datasets in addition to the network inference R^2 score, all which range from 0 to 1. Finally, for each gene, the GRN with the maximum merged score is selected. The refined network obtained is then transformed into a cooperativity network, based on the common targets of regulators.

We began by selecting a data set containing enough gene expression samples to obtain a representative network of gene regulation in yeast: Many Microbe Microarray Database (M3D) [19]. This database contains data from 247 experiments measuring gene expression under different conditions in microarray assays. The data were collated, normalized and averaged (in the case of replicates) for 5520 probes mapping onto ORF.

We used CoRegNet to infer a representative regulatory network for yeast. This network should provide insight into the regulators that work together in the performance of a particular biological function. We enriched the network by searching the Yeastract database [20] for known TF-target interactions, and the Biogrid database [21] for known protein-protein interactions. The inferred CoRegNet network has a data structure extending beyond information about regulator-target and regulator-regulator cooperativity [14], see Fig. 2. It also represents the regulatory state of the cell for a given gene expression sample, as explained below.

Trejo Banos *et al. BMC Systems Biology* 2017, **11**(Suppl 7):134

Page 83 of 174

Fig. 1 The CoRegFlux workflow, in which a training gene expression data set is used to infer a coregulatory network (in this case from the M3D database [19]). Using this inferred network, we calculated the influence scores of the regulators (i.e. TFs, kinases). We then used these scores to train a linear model for predicting gene expression from influence scores. This model can use context-specific influence scores to predict the activity of metabolic genes in a robust manner. Using these predictions, the model employs a continuous-value version of the Gene-Protein-Reaction rules and a function to translate gene activity into flux bounds. The bounds obtained are then input into a linear program to obtain fluxes congruent with the gene-regulatory state of the cell in a given context

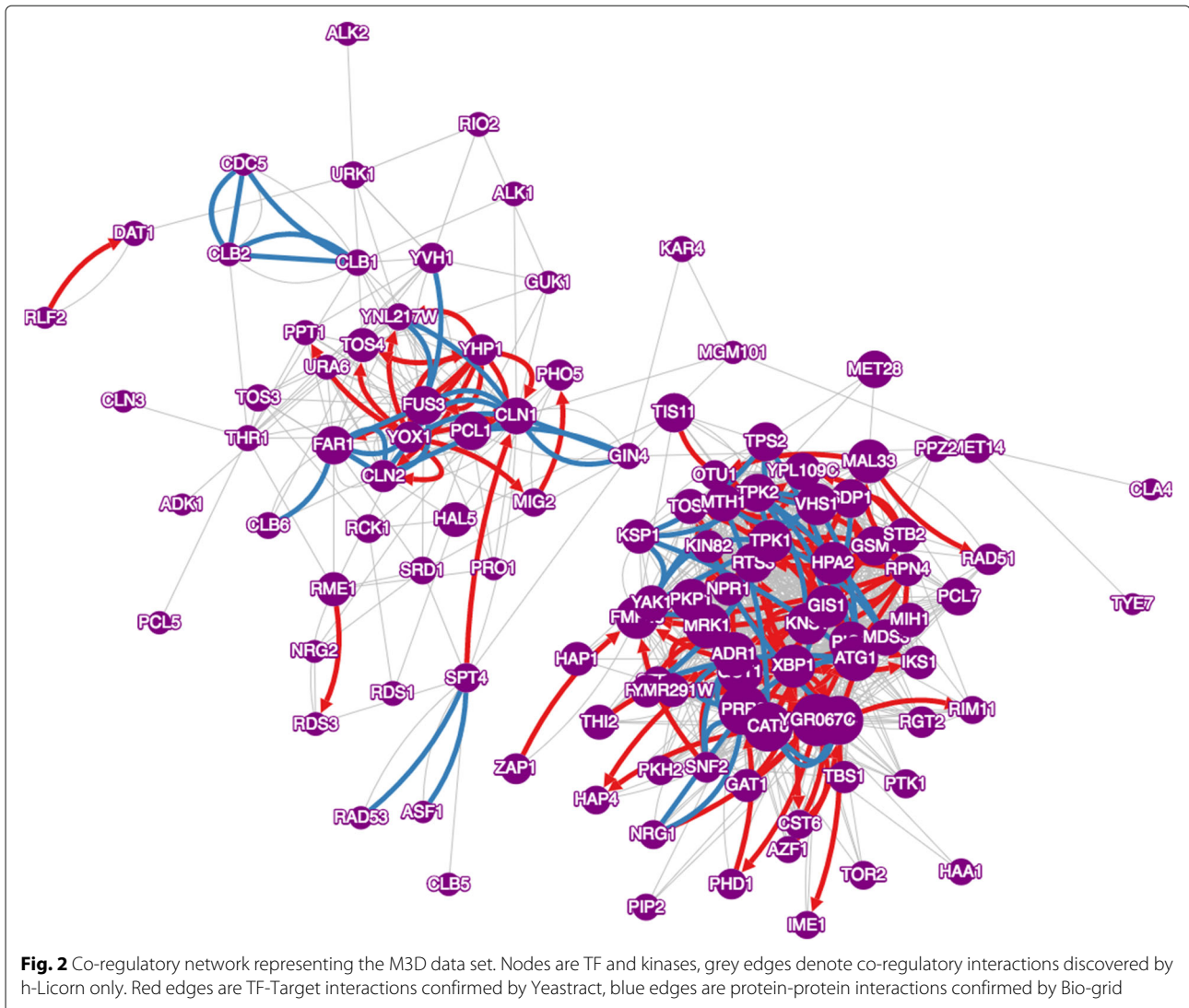**Network interrogation and influence score**

The bipartite graph generated makes it possible to generate a low-dimensional representation of the transcriptomic data. Nicolle et al. [22] introduced the notion of regulator influence. Here, the impact of a regulator on its targets is represented by the scaled difference of the mean expression levels of its activated and repressed targets. This score is given by the expression

$$I_j = \frac{\hat{X}_A - \hat{X}_R}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_R^2}{n_R}}} \qquad (1)$$

where $\hat{X}_A$ and $\hat{X}_R$ are the means of the activated and repressed targets of regulator j respectively. The variables s_A, s_R represent the respective standard deviations and n_A, n_R represent the number of genes contained in the respective set. The influence score accounts for the effect of a regulator on its targets according to the regulator-target relationships inferred using hLICORN and additional data-sources integrated in the network as evidences using the CoRegNet Bioconductor package. Briefly, this measure is based on a Welch't-test between the expression of the activated and repressed targets genes in a given samples.

Thus, a data set of thousands of gene expression measurements is reduced to just dozens or hundreds of activity scores (one score for each regulator with a significant influence). In the case of the M3D data set, the influence score was computed for the set of TFs and

Trejo Banos *et al. BMC Systems Biology* 2017, **11**(Suppl 7):134

Page 84 of 174

Fig. 2 Co-regulatory network representing the M3D data set. Nodes are TF and kinases, grey edges denote co-regulatory interactions discovered by h-Licorn only. Red edges are TF-Target interactions confirmed by Yeastract, blue edges are protein-protein interactions confirmed by Bio-grid

kinases as given by [20] (200304 possible TF-target interactions) and the *S. cerevisiae* Kinase and Phosphatase Interactome resource [23] (262354 possible P-P interactions), with a total of 567 potential regulators.

Using this background knowledge of the network and scores, over a wide range of conditions, we aim to predict gene expression and, by extension, the enzymatic activity of proteins encoded, in a metabolic model. We argue that, unlike gene expression alone, influence provides a robust portrait of the transcriptomic state of the cell, improving predictions of the behaviour of targets [22]. We used the inferred network and calculated regulator influences to train a linear regression model over a set of training samples K. In this training set the gene expression level of a target in a given sample is a function of the influence of its regulators:

$$x_{ik} = \sum_{j \in Pa(i)} \beta_j I_{jk} \qquad (2)$$

where x_{ik} is the expression level for enzyme i in sample k, $Pa(x_i)$ is the set of regulators of i in the network and I_{jk} is the influence of regulators j in sample k. The objective of the linear regression is to calculate the regression coefficients β_j. For our purposes, we trained the linear model on the M3D data set. Thus, the *beta* coefficients capture the general relationship between gene expression and influence over a wide range of conditions. The linear regression model is then used to predict the level of expression of a gene encoding a given enzyme in the set of context-specific samples of interest. For this, we calculate influence for the samples of interest and predict the gene expression of the metabolic genes with 2. In this study, we used the data set of [24], from a study in which a yeast strain was subjected to various oxygen concentrations. Using the inferred network, we calculated influence scores for this data set. According to the CoRegFlux work flow, we used influence to predict enzyme activity for each sample, based on a linear regression model. We limited

predictions to the enzymes present in the genome-scale metabolic model of yeast [16]. These predictions sum up all the available regulatory information for yeast, as given by the inferred network, along with the context specificity of TF influences.

Metabolic model adjustment

We then translated the predicted enzymatic activities into bounds, corresponding to the extent to which the enzyme encoded by the gene can catalyse a given reaction. These bounds can be used to constrain a linear program representing the metabolic fluxes of a stoichiometric model under the assumption of steady state, a method known as flux balance analysis [25]. The algorithm is as follows:

1. We transform the gene-protein-reaction (gpr) rules, which relate enzyme and enzyme complexes to a given reaction in the model. The original rules are in boolean form and our substitution follows a continuous approximation similar to [26, 27]. Thus the conversion is:

    (a) OR sentences, which represent isoenzymes regulating the same reaction are substituted by a function max(), which returns the maximum of the expression of the corresponding enzymes.
    (b) AND sentences, which represent the formation of enzyme complexes are substituted by a function min() which returns the minimum of the expression of the corresponding enzymes.
    (c) If the enzyme expression is not available, the enzyme is discarded from the rules.

We denote the evaluation of a gpr rule for an enzyme-associated reaction by $gpr_r(X_{pred})$, with X_{pred} being the set of predicted gene expressions as a function of the influence scores.
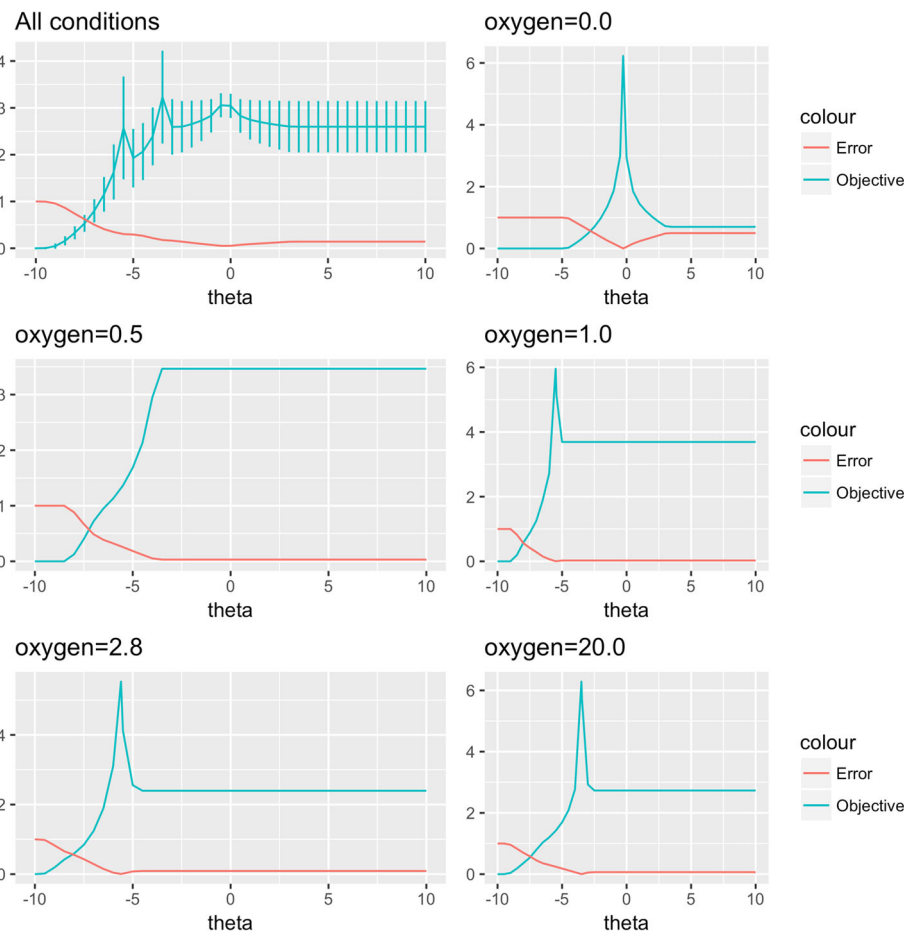

Fig. 3 Absolute relative error and objective function for CoRegFlux for different values of theta over all conditions (top left). Values of error and objective function for each condition/oxygenation level (top right, middle left and right, bottom left and right). Bayesian optimization yielded a single optimum for all conditions except an oxygen concentration of 0.5 , for which no solution better than the FBA solution was found

Trejo Banos *et al. BMC Systems Biology* 2017, **11**(Suppl 7):134

Page 86 of 174

2. Using the continuos gpr rules we adjust the fluxes bounds for each gpr-associated flux, denoted v^r by the following relation

$$v^r \leq \ln\left(1 + \exp\left(gpr_r\left(X_{pred}\right) + \theta\right)\right) \qquad (3)$$

where we introduce the parameter θ to account for enzymatic action over the reaction. We assume this parameter is condition-specific. We chose the activation function, known as softplus [28], to convey the non-linear relationship between gene expression and protein concentrations. Unlike other non-linear activation functions, like sigmoids, the softplus has a range of $(0, +\infty)$ making it straightforward to use as flux bounds.

With these new constraints, the flux values and biomass yield can be calculated by solving the linear program associated with the model. We used the R package sybil [29] to find the flux distribution optimizing growth under the new bounds.

Bayesian optimisation of the parameters

If we wish to adjust parameter θ such that the observed phenotype matches the simulated fluxes, a Bayesian optimization algorithm can be used. Bayesian optimization provides an effective out-of-the box solution for a non-linear optimization problem [30]. For the data set of Rintala et al., we ran CoRegFlux, varying the value of parameter θ over a uniform grid of 10 points in the interval (-10,10). We then applied the R optimization package [31], to maximize the objective:

$$\max_{\theta}\left[-\log\left(\frac{\left|v_{observed}^B - v_{simulated}^B\right|}{v_{observed}^B}\right)\right]$$

This function reaches its maximum when the simulated biomass yield $v_{simulated}^B$ is closest to the observed growth rate $v_{observed}^B$. We chose to use this function, to improve appreciation of the effect of the parameter value on approximation error. We used the default settings of the Bayesian optimization package [31] to estimate optimal values of θ for each condition. The results are shown in Fig. 3, in which, for each condition, the value of the parameter increases as it approaches the optimum value (reducing the relative error). If we continue to increase the parameter, the relative error of the solution settles at the value for the flux balance analysis model. A special case occurs when oxygen concentration is 0.5 . In this case, the flux balance analysis solution underestimates the growth yield, and the method is unable to find an optimum value for the parameter. For all other cases, a clear optimum value is identified.

Results

We evaluated the results generated by our method in terms of the accuracy with which they predicted exchange fluxes and to illustrate the use of this approach in a relevant case study. We performed robustness tests to determine whether influence gave a more reliable picture of the regulatory state of the cell. As our case study we choose the diauxic shift, a complex biological process involving
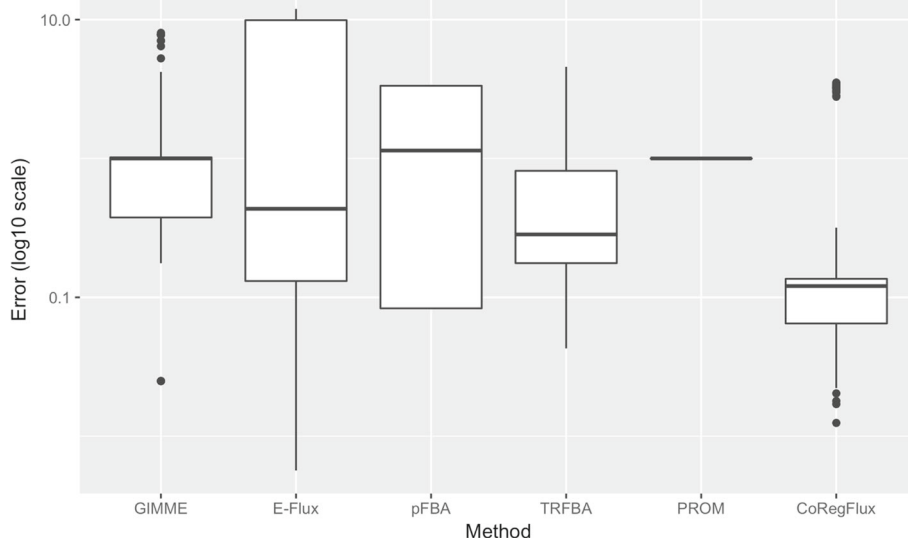

Fig. 4 Results for robustness analysis in all conditions and for five different noise levels. The normalized error, corresponding to the difference between observed and simulated exchange fluxes, is shown on the y axis. A log10 transformation was applied to the data to improve readability. CoRegFlux had a lower median error than two other state-of-the-art methods, TRFBA and pFBA

Trejo Banos *et al. BMC Systems Biology* 2017, **11**(Suppl 7):134

Page 87 of 174

major changes in transcriptional and metabolic elements in yeast.

Robustness tests

We evaluated the robustness of our method to random permutations of gene expression, as recommended by [17]. We set our θ parameter to its optimum value for each condition, and we then tested five different noise levels for each condition. The mean squared error for exchange fluxes was calculated as described by [17]; Fig. 4 shows a boxplot for the base 10 logarithm of the error. We also considered the results generated by competing methods: GIMME, E-Flux, pFBA, TRFBA and PROM [12]. Our method had a better median performance and a smaller variance than the other methods. As all tests were performed in the wild-type strain, PROM [11] displayed no

variation, as this method was designed for prediction for knockout strains and does not seem to take regulatory information into account for the wild type.

Case study

We used diauxic shift as a case study, with the gene expression measurements of [32] corresponding to 12 time points (9 before and 3 after glucose depletion). We compared this data set with that of [33], for seven samples during the diauxic shift. We plotted influence score heatmaps for both data sets. We used canonical correlations analysis to compare the correlation between sample points in the two different data sets. We alternated between gene expression and regulatory influence, and the corresponding correlation plot is shown in Fig. 5. Plots based on regulatory influence separated two distinct clusters of
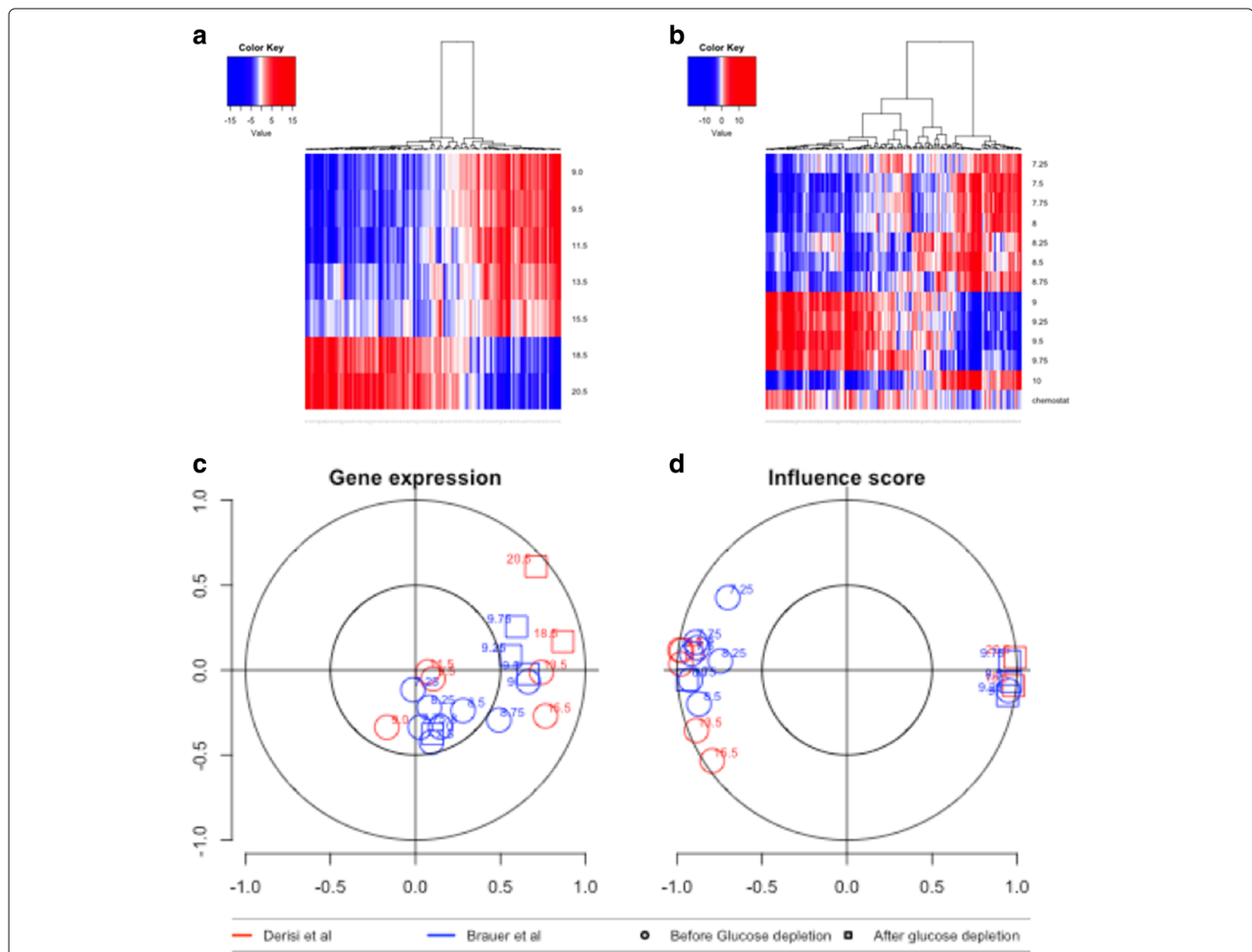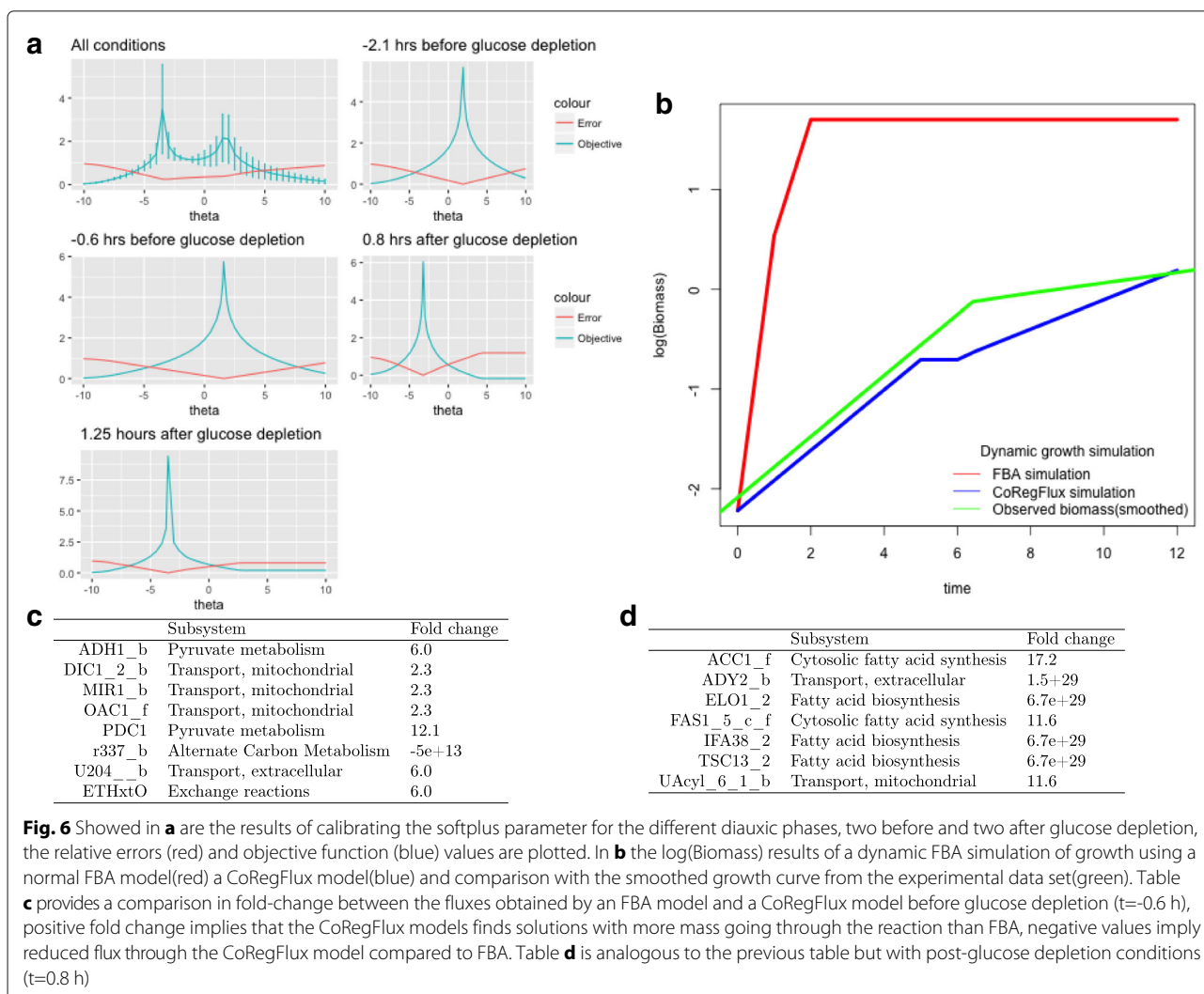

Fig. 5 In **a**, heatmap for influences for the data set of Derisi et al., in **b**, heatmap for the dataset of Brauer et al. Positive influences are shown in red and correspond to an upregulation of targets, whereas negative influences are shown in blue and correspond to the downregulation of targets. At **c**, the correlation between samples based on gene expression is shown (calculated by canonical correlation analysis). At **d**, the equivalent correlation plot based on influence scores is shown. Samples obtained before and after glucose depletion can be clearly differentiated on the basis of influence scores, whereas the relationship is less clear for gene expression

Trejo Banos *et al. BMC Systems Biology* 2017, **11**(Suppl 7):134

Page 88 of 174

samples, corresponding to the samples taken before and after glucose depletion, except for the sample taken at t=9 h [32] appearing closer to a post-depletion state (it should be noted that the authors of the paper reported regulatory changes beginning a few minutes before glucose depletion). Another interesting point is that of t=10 h in [32], for which the influence profile seems more similar to those obtained before glucose depletion, which may point to a growth-state. These interesting patterns were not evident in analyses of regulator gene expression, in which the separation between pre- and post-depletion samples was less clear.

We matched the sample points to the different regulatory phases identified by [34], with metabolic states attributed to the phases -2.1 and -0.6 h before glucose depletion and 0.8 and 4 h after glucose depletion. In our case, the last gene expression sample from [32] was taken at t=1.25 h after glucose depletion. With these phases in mind, we adjusted the glucose and ethanol exchange bounds to those for the metabolomic data set

of [34] and parametrized our CoRegFlux models with the growth rates reported for each phase in [34]. As shown in Fig. 6 top-left, the pre- and post-glucose depletion models had different optimal values for the parameter θ, again reflecting the information about the regulatory state of the cell provided by the influence score. We further investigated the differences between CoRegFlux model output and flux balance solutions. The differences are shown in tables bottom-right and bottom left of Fig. 6, in which FBA fluxes and CoRegFlux fluxes are compared in fold change for the pre- exhaustion and post exhaustion respectively, we chose to present only those fluxes that experience more than a two-fold change. The tables show increased ethanol excretion (reaction ETHxt0), in fact ethanol excretion is predicted as 0 by FBA, along with increase transport of metabolites to the mitochondria for the pre-glucose exhaustion phase at -0.6 h which matches the observations of [34] . In the post-glucose depletion state at 0.8 h, CoRegFlux predicts an increase in fluxes regulated by ACC1 and FAS1 genes, which are important in

a All conditions / -2.1 hrs before glucose depletion / -0.6 hrs before glucose depletion / 0.8 hrs after glucose depletion / 1.25 hours after glucose depletion

b Dynamic growth simulation

c

	Subsystem	Fold change
ADH1_b	Pyruvate metabolism	6.0
DIC1_2_b	Transport, mitochondrial	2.3
MIR1_b	Transport, mitochondrial	2.3
OAC1_f	Transport, mitochondrial	2.3
PDC1	Pyruvate metabolism	12.1
r337_b	Alternate Carbon Metabolism	-5e+13
U204__b	Transport, extracellular	6.0
ETHxtO	Exchange reactions	6.0

d

	Subsystem	Fold change
ACC1_f	Cytosolic fatty acid synthesis	17.2
ADY2_b	Transport, extracellular	1.5+29
ELO1_2	Fatty acid biosynthesis	6.7e+29
FAS1_5_c_f	Cytosolic fatty acid synthesis	11.6
IFA38_2	Fatty acid biosynthesis	6.7e+29
TSC13_2	Fatty acid biosynthesis	6.7e+29
UAcyl_6_1_b	Transport, mitochondrial	11.6

Fig. 6 Showed in **a** are the results of calibrating the softplus parameter for the different diauxic phases, two before and two after glucose depletion, the relative errors (red) and objective function (blue) values are plotted. In **b** the log(Biomass) results of a dynamic FBA simulation of growth using a normal FBA model(red) a CoRegFlux model(blue) and comparison with the smoothed growth curve from the experimental data set(green). Table **c** provides a comparison in fold-change between the fluxes obtained by an FBA model and a CoRegFlux model before glucose depletion (t=-0.6 h), positive fold change implies that the CoRegFlux models finds solutions with more mass going through the reaction than FBA, negative values imply reduced flux through the CoRegFlux model compared to FBA. Table **d** is analogous to the previous table but with post-glucose depletion conditions (t=0.8 h)

Trejo Banos *et al. BMC Systems Biology* 2017, **11**(Suppl 7):134

Page 89 of 174

the production of Acyl CoA, which oxidizes and becomes Acetyl-CoA, primary precursor of ATP production by the TCA cycle post shift [35]. Finally we assessed the utility of our model for dynamic growth simulations using a dynamic FBA formulation as in [36]. We used the initial biomass, glucose and ethanol concentrations from [34] and computed the metabolite consumption and growth rate at each time step. We compared the results of using a normal FBA model constrained only to the initial glucose uptake rate (but allowing ethanol consumption), to the results using CoRegFlux models. We proceed as follows: one of the previously constrained models was assigned to the corresponding time points, then at the switch points between diauxic phases, the current biomass and metabolite concentrations was used as initial conditions for the next model. The derived growth curves are presented in top right of Fig. 6, where we see that the FBA model both over-estimates growth and does not initiate the second phase of diauxic growth. The CoRegFlux model on the other hand, follows more closely the smoothed growth curve provided by the authors of said study.

Discussion and conclusions

We propose CoRegFlux, a new algorithm for integrating gene regulatory network inference with constraint-based metabolic models. Our method provided better median predictions with a lower variance prediction than other state-of-the-art methods for predicting exchange fluxes under different levels of perturbation of gene expression data. One of the limitations of this method is that it cannot determine the optimal parameter value for systems in which the normal FBA solution underestimates biomass yield, although it should be pointed out that FBA overestimates the growth rate in most cases [37]. From the robustness tests and the case study, we can conclude that influence score calculation is a reliable way to assess the overall effects of gene regulation. This advantage of the influence score places it among other approaches to dimensionality reduction for gene expression such as network component analysis [38]. The importance of having a clear idea of the transcriptomic state of the cell has been demonstrated in studies of metabolism and responses to particular conditions. For example, recent results have suggested that at least 70% of the total variance in promoter activity across conditions can be accounted for by global transcriptional regulation in *E. coli* [39].

As mentioned above, this method has potential applications in research, industry and medicine, and its improvement would therefore be worthwhile. For example, it would be interesting to include different models of gene regulation as additional predictors of enzyme activity. Future research studies could also include metabolic network learning, with a view to the development of a data-driven integrated algorithm. Finally, this method is designed to serve as a basis for the *in-silico* optimization of biological objectives, of potential value for experimental design in systems and synthetic biology.

Availability of data and materials
Data and source code can be downloaded from http://github.com/i3bionet/CoRegFlux.

About this supplement
This article has been published as part of *BMC Systems Biology* Volume 11 Supplement 7, 2017: 16th International Conference on Bioinformatics (InCoB 2017): Systems Biology. The full contents of the supplement are available online at https://bmcsystbiol.biomedcentral.com/articles/supplements/volume-11-supplement-6.

Authors' contributions
ME and DT conceived the study. DT and ME designed it and wrote the manuscript. All authors provided valuable advises in developing the proposed method and modifying the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]UMR 8030 Génomique Métabolique / Laboratoire iSSB CEA-CNRS-UEVE, Genopole campus 1, 5 rue Henri Desbruères, 91030 Cedex Évry, France. [2]Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France. [3]Université Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL - Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille, France.

Published: 21 December 2017

References
1. Macklin DN, Ruggero NA, Covert MW. The future of whole-cell modeling. Curr Opin Biotechnol. 2014;28:111–5.
2. Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival Jr B, Assad-Garcia N, Glass JI, Covert MW. A whole-cell computational model predicts phenotype from genotype. Cell. 2012;150(2):389–401.
3. Michael DG, Maier EJ, Brown H, Gish SR, Fiore C, Brown RH, Brent MR. Model-based transcriptome engineering promotes a fermentative transcriptional state in yeast. Proc Natl Acad Sci. 2016;113(47):E7428–37.
4. Gatto F, Miess H, Schulze A, Nielsen J. Flux balance analysis predicts essential genes in clear cell renal cell carcinoma metabolism. Sci Rep. 2015;5:10738.
5. Steuer R. Computational approaches to the topology, stability and dynamics of metabolic networks. Phytochemistry. 2007;68(16):2139–51.
6. Davidson EH, (ed). The Regulatory Genome. Burlington: Academic Press; 2006. p. 1–29.

Trejo Banos *et al. BMC Systems Biology* 2017, **11**(Suppl 7):134

Page 90 of 174

7. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, Adkins JN, Schramm G, Purvine SO, Lopez-Ferrer D, et al. Omic data from evolved e. coli are consistent with computed optimal growth from genome-scale models. Mol Syst Biol. 2010;6(1):390.

8. Colijn C, Brandes A, Zucker J, Lun DS, Weiner B, Farhat MR, Cheng TY, Moody DB, Murray M, Galagan JE. Interpreting expression data with metabolic flux models: predicting mycobacterium tuberculosis mycolic acid production. PLoS Comput Biol. 2009;5(8):e1000489.

9. Becker SA, Palsson BO. Context-specific metabolic networks are consistent with experiments. PLoS Comput Biol. 2008;4(5):e1000082.

10. Covert MW, Xiao N, Chen TJ, Karr JR. Integrating metabolic, transcriptional regulatory and signal transduction models in escherichia coli. Bioinformatics. 2008;24(18):2044–50.

11. Chandrasekaran S, Price ND. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in escherichia coli and mycobacterium tuberculosis. Proc Natl Acad Sci. 2010;107(41):17845–50.

12. Motamedian E, Mohammadi M, Shojaosadati SA, Heydari M. Trfba: an algorithm to integrate genome-scale metabolic and transcriptional regulatory networks with incorporation of expression data. Bioinformatics. 2017;33(7):1057–63. doi:10.1093/bioinformatics/btw772.

13. Elati M, Neuvial P, Bolotin-Fukuhara M, Barillot E, Radvanyi F, Rouveirol C. Licorn: learning cooperative regulation networks from gene expression data. Bioinformatics. 2007;23(18):2407–14.

14. Nicolle R, Radvanyi F, Elati M. Coregnet: reconstruction and integrated analysis of co-regulatory networks. Bioinformatics. 2015;31(18):3066–8. doi:10.1093/bioinformatics/btv305.

15. Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, Allison KR, Kellis M, Collins JJ, Stolovitzky G, et al. Wisdom of crowds for robust gene network inference. Nat Methods. 2012;9(8):796–804.

16. Österlund T, Nookaew I, Bordel S, Nielsen J. Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. BMC Syst Biol. 2013;7(1):36.

17. 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.

18. Chebil I, Nicolle R, Santini G, Rouveirol C, Elati M. Hybrid method inference for the construction of cooperative regulatory network in human. NanoBioscience IEEE Trans. 2014;13(2):97–103.

19. Faith JJ, Driscoll ME, Fusaro VA, Cosgrove EJ, Hayete B, Juhn FS, Schneider SJ, Gardner TS. Many microbe microarrays database: uniformly normalized affymetrix compendia with structured experimental metadata. Nucleic Acids Res. 2008;36(suppl 1):D866–70.

20. Teixeira MC, Monteiro P, Jain P, Tenreiro S, Fernandes AR, Mira NP, Alenquer M, Freitas AT, Oliveira AL, Sá-Correia I. The yeastract database: a tool for the analysis of transcription regulatory associations in saccharomyces cerevisiae. Nucleic Acids Res. 2006;34(suppl 1):D446–51.

21. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. Biogrid: a general repository for interaction datasets. Nucleic Acids Res. 2006;34(suppl 1):D535–9.

22. Nicolle R, Elati M, Radvanyi F. Network transformation of gene expression for feature extraction. In: 2012 11th International Conference on Machine Learning and Applications. Piscataway: IEEE. 2012. p. 108–113. doi:10.1109/ICMLA.2012.27.

23. Breitkreutz A, Choi H, Sharom JR, Boucher L, Neduva V, Larsen B, Lin ZY, Breitkreutz BJ, Stark C, Liu G, et al. A global protein kinase and phosphatase interaction network in yeast. Science. 2010;328(5981):1043–6.

24. Rintala E, Toivari M, Pitkänen JP, Wiebe MG, Ruohonen L, Penttilä M. Low oxygen levels as a trigger for enhancement of respiratory metabolism in saccharomyces cerevisiae. BMC Genomics. 2009;10(1):461.

25. Orth JD, Thiele I, Palsson B. What is flux balance analysis? Nat Biotechnol. 2010;28(3):245–8.

26. Jensen PA, Lutz KA, Papin JA. Tiger: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. BMC Syst Biol. 2011;5(1):147.

27. Osorio D, Botero K, Gonzalez J, Pinzon A. exp2flux: Convert Gene EXPression Data to FBA FLUXes. 2016. http://CRAN.R-project.org/package=exp2flux. R package version 0.1. Accessed Feb 2017.

28. Dugas C, Bengio Y, Bélisle F, Nadeau C, Garcia R. Incorporating second-order functional knowledge for better option pricing. Adv Neural Inf Process Syst. 2001;472–8.

29. Gelius-Dietrich G, Fritzemeier CJ, Desouki AA, Lercher MJ. sybil – efficient constraint-based modelling in r. BMC Syst Biol. 2013;7(1):125. doi:10.1186/1752-0509-7-125. http://www.biomedcentral.com/1752-0509/7/125.

30. Močkus J. On bayesian methods for seeking the extremum. In: Marchuk GI, editor. Optimization Techniques IFIP Technical Conference Novosibirsk, July 1-7, 1974. Berlin: Springer Berlin Heidelberg. 1975. p. 400–4. doi:10.1007/3-540-07165-2_55.

31. Yan Y. rBayesianOptimization: Bayesian Optimization of Hyperparameters. http://github.com/yanyachen/rBayesianOptimization. R package version 1.1.0. Accessed Feb 2017.

32. Brauer MJ, Saldanha AJ, Dolinski K, Botstein D. Homeostatic adjustment and metabolic remodeling in glucose-limited yeast cultures. Mol Biol Cell. 2005;16(5):2503–17.

33. DeRisi JL, Iyer VR, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science. 1997;278(5338):680–6.

34. Zampar GG, Kümmel A, Ewald J, Jol S, Niebel B, Picotti P, Aebersold R, Sauer U, Zamboni N, Heinemann M. Temporal system-level organization of the switch from glycolytic to gluconeogenic operation in yeast. Mol Syst Biol. 2013;9(1):651.

35. de Jong BW, Siewers V, Nielsen J. Physiological and transcriptional characterization of saccharomyces cerevisiae engineered for production of fatty acid ethyl esters. FEMS Yeast Res. 2016;16(1):fov105.

36. Varma A, Palsson BO. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type escherichia coli w3110. Appl Environ Microbiol. 1994;60(10):3724–31.

37. Segrè D, Zucker J, Katz J, Lin X, D'haeseleer P, Rindone WP, Kharchenko P, Nguyen DH, Wright MA, Church GM. From annotated genomes to metabolic flux models and kinetic parameter fitting. OMICS A J Integrative Biol. 2003;7(3):301–16.

38. Liao JC, Boscolo R, Yang YL, Tran LM, Sabatti C, Roychowdhury VP. Network component analysis: reconstruction of regulatory signals in biological systems. Proc Natl Acad Sci. 2003;100(26):15522–7.

39. Kochanowski K, Gerosa L, Brunner SF, Christodoulou D, Nikolaev YV, Sauer U. Few regulatory metabolites coordinate expression of central metabolic genes in escherichia coli. Mol Syst Biol. 2017;13(1):903.