ETH Eidgenössische Technische Hochschule Zürich Swiss Federal Institute of Technology Zurich

Institute of Molecular Systems Biology

Identification of transcription factor signaling molecules by coupling gene expression and metabolomics

Daniela Ledezma-Tejeida & Uwe Sauer
Institute of Molecular Systems Biology, ETH Zürich, Zurich, Switzerland

Transcription Factors – the unknown

Escherichia coli has the best characterized bacterial transcriptional regulatory network, comprising ~300 transcription factors[1] (TF), of which 75% have a predicted metabolite-binding domain[2]. Nowadays, only 93 TFs[3] have had one or more binding metabolites identified, suggesting that there are many metabolite-TFs left to identify.

Known metabolite-TF interactions have been identified through *in vitro* assays that do not provide evidence of the functional relevance of the interaction.

Here, we combine metabolomics and gene expression data to identify new, functionally-relevant metabolite-TF interactions *in vivo*

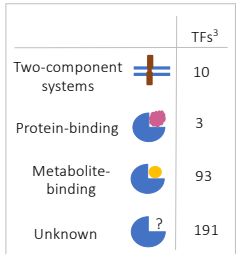


Figure 1. Mechanisms for conformational change of TFs and their abundance in the regulatory network of E. coli

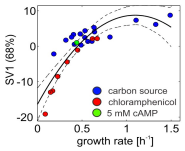
Identifying signals responsible for gene expression

Promoter activity (PA), a proxy of gene expression, is determined by[4]:

Global PA

Influenced by the global physiology of the cell, proportional to growth rate

Explains most of the variance in PA, corresponds to the first principal component in PCA[5]



Specific PA

Influenced by TFs

Specific PA = total PA – global PA
Can be obtained by substracting global PA from total PA

Alternatively, it can be approximated from the abundance of metabolites binding the TF regulating the promoter[5]:

$$S \approx \sum_{l\in\{TF\},k\in\{M\}} \alpha_{li} \cdot \beta_{lk} \cdot \Delta \log(M_{kj}) \quad \text{Eq 1}$$

binding parameters

TFs regulating the promoter

metabolites binding the TF

Total PA can be obtained through reporter strains (GFP)[6], while metabolite abundance can be measured using mass spectrometry. Theoretically, by introducing abundances of different metabolites into Eq 1 and comparing the obtained specific PA to the measured specific PA through GFP, it is possible to identify true signaling metabolites (Figure 2).

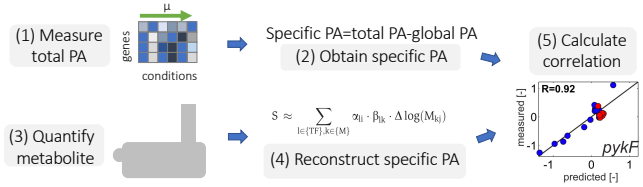


Figure 2. Pipeline to identify signaling metabolites

Proof of concept

We have evaluated our methodology by testing promoters regulated by three TFs with known binding-metabolites: arginine-ArgR, acetylserine-CysB and TyrR known to bind aromatic amino acids. We compared four metrics to identify true positives and the best-scoring was Pearson correlation (Figure 3a).

A true binding metabolite was statistically significant in 8/9 promoters tested and (Figure 3b) its correlation coefficient was ranked among the top 4%, supporting the applicability of our approach (Figure 3c).

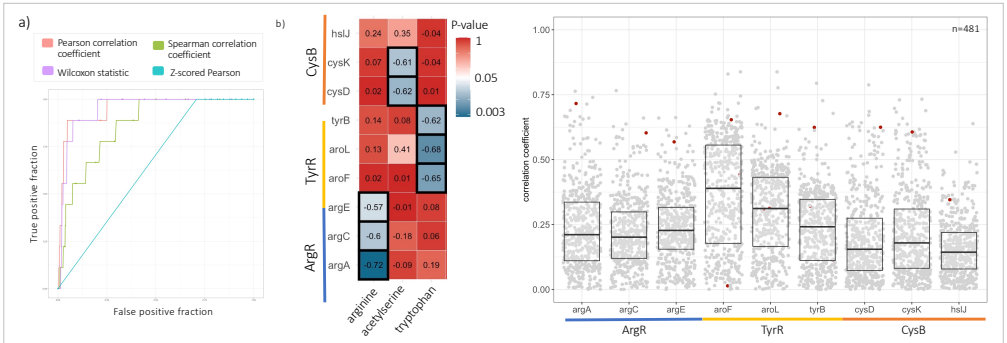


Figure 3. Correlation coefficients of measured and predicted specific PA. (a) ROC curves comparing (i) Pearson and (ii) Spearman correlation of PA, (iii) z-scored Pearson correlation and (iv) Wilcoxon test statistic of each metabolite compared to z-scored distribution. (b) Correlation coefficients of metabolites known to bind the TFs regulating the promoters tested. Colour indicates p-value of correlation, p-values < 0.05 are indicated by a black border. (c) Distribution of absolute correlation coefficients of 481 metabolites tested per promoter.

Prediction of signaling metabolites

We applied our pipeline to 4 TFs without known signaling metabolites: CdaR, CsgD, GadX and FlhDC. The number of predictions varied between promoters, from three (ycgR, nlpA) to over a hundred (pepD) (Figure 4). The median number of predictions was 20. In order to filter false positives, we identified the metabolites that were statistically significant across all the promoters regulated by a TF (Table 1).

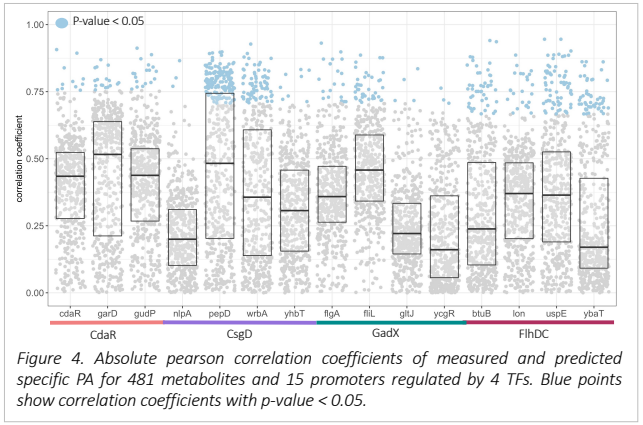


Figure 4. Absolute pearson correlation coefficients of measured and predicted specific PA for 481 metabolites and 15 promoters regulated by 4 TFs. Blue points show correlation coefficients with p-value < 0.05.

TF	metabolites
CdaR	xylonate
CsgD	Dioxobutanoic acid AMP Hexose Lactate
GadX	cysteinylglycine 2-keto-glutaramate Urocanic acid
FlhDC	47 metabolites

Table 1. Singalling molecule predictions

CdaR stands for "carbohydrate diacid regulator" and regulates genes involved in carbon uptake[7], which is consistent with the prediction of xylonate, a carbon source, as signalling molecule. CsgD is involved in curli assembly and biofilm formation[8], it is known to respond to starvation and high cell density, among other perturbations that affect E. coli growth rate. It is likely that its signal is involved in central carbon metabolism, such as the predicted hexoses, lactate or AMP. GadX regulates genes involved in pH regulation and is glutamate dependent[9], making 2-keto-glutaramate a highly likely signal. Additionally, cysteinylglicine is a product of the hydrolyzation of gluthatione into glutamate, making also a good signal candidate. FlhDC is involved in flagellar synthesis[10], which might be signalled by several inputs and could explain the high number of predictions. Further analyses are needed to rank the most likely signalling metabolite from the predicted pool.

Conclusions

We have designed a method to predict signaling metabolites of TFs that yielded correct results in 8/9 promoters tested. We used it to predict signaling metabolites for 4 TFs. The next steps are *in vitro* validations of predictions.

More information: ledezma@imsb.biol.ethz.ch

References

1. Pérez-Rueda et al. *Comput Biol Chem*, 2015.
2. Babu & Teichmann. *NAR*, 2003.
3. Santos-Zavaleta et al. *NAR*, 2018.
4. Berthoumieux et al. *Mol Syst Biol*, 2013.
5. Kochanowski et al. *Mol Syst Biol*, 2017.
6. Zaslaver et al. *Nature Methods*, 2006.
7. Monterrubio et al. *J Bacteriol*, 2000.
8. Loferer et al. *Mol Microbiol*, 1997.
9. Ma et al. *J Bacteriol*, 2002.
10. Claret & Hughes. *J Mol Biol*, 2002.