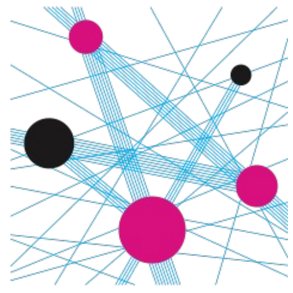
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute of Molecular Systems Biology

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



Daniela Ledezma-Tejeida, Eliane Züger & 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.

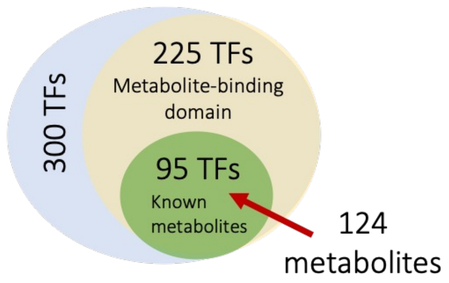
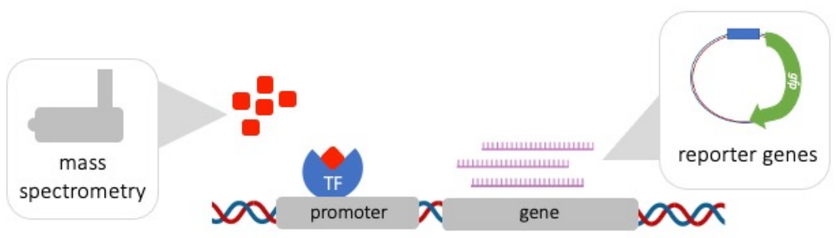


Figure 1. Known TF-metabolite interactions

Known TF-metabolite 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*

Identifying signals responsible for gene expression



When a metabolite has positive effect on a promoter's transcription via a TF, we can assume that increases in the metabolite abundance will correlate with increases in the regulated promoter's transcript abundance.

By measuring metabolite abundance – via Mass Spectrometry[4] – and transcription of genes known to be regulated by a single TF – via GFP[5] –, it is possible to identify metabolites that are interacting with the TF.

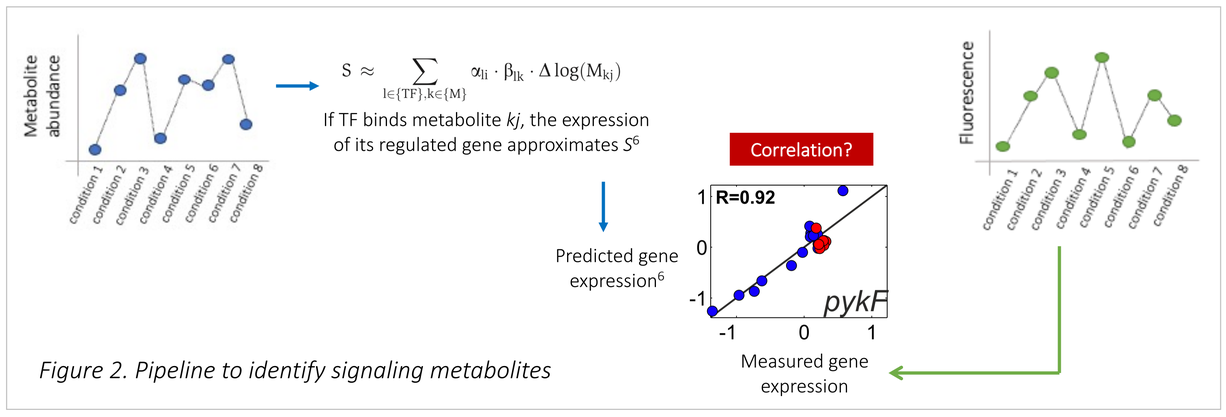


Figure 2. Pipeline to identify signaling metabolites

Proof of concept

We have evaluated our methodology by testing promoters regulated by three TFs with known binding-metabolites: ArgR-arginine, CysB-acetylserine 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).

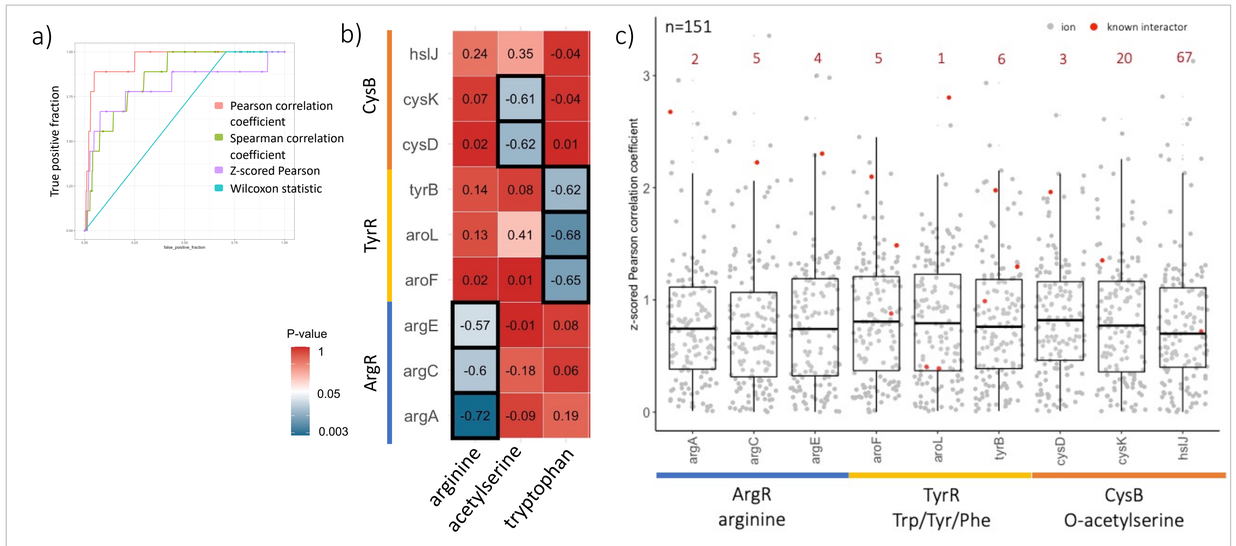


Figure 3. Correlation coefficients of measured and predicted specific PA. (a) ROC curves comparing (i) Pearson and (ii) Spearman correlation of PA, (iii) z-scored Pearson correlation and (iv) Wilcoxon test statistic of each metabolite compared to z-scored distribution. (b) Correlation coefficients of 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 151 metabolites tested per promoter.

A true binding metabolite was statistically significant in 8/9 promoters tested (Figure 3b) and in 7/9, its correlation coefficient was ranked among the top 4%, supporting the applicability of our approach (Figure 3c).

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

Identification of new signaling metabolites

We observed that among ArgR's highest-scoring metabolites were ornithine and lysine. Lysine has been previously shown to bind ArgR[7], but it's effect on the TF's affinity to DNA has never been shown. Electrophoretic Mobility Shift Assays (EMSA) showed that lysine and ornithine increase the binding affinity of ArgR to the *argA* promoter (Figure 4), further supporting that our method is capable of identifying true, functionally-relevant, TF-metabolite interactions.

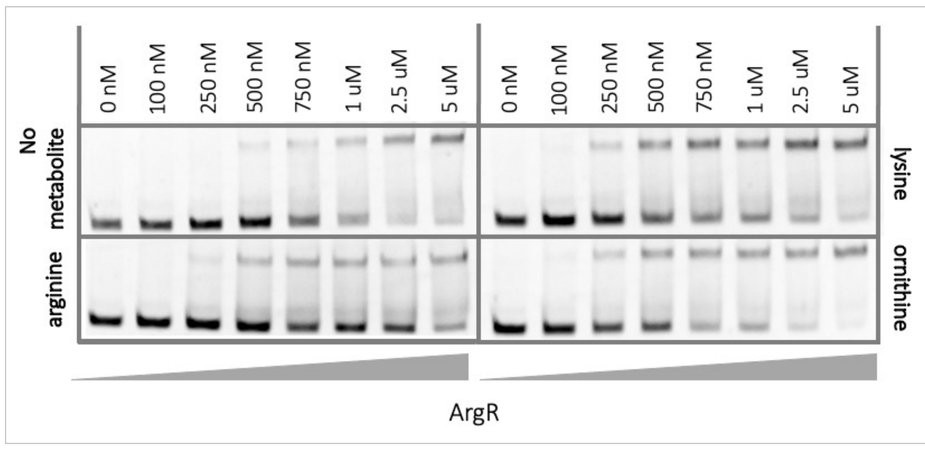


Figure 4. EMSA using argA promoter region as DNA probe and purified ArgR. Free DNA shows in lower bands, ArgR-DNA complexes show in upper bands. TF and DNA probes were incubated in the presence of 16 mM metabolite.

We applied our pipeline to 8 TFs without known signaling metabolites: CdaR, CsgD, FlhDC, GadX, CecR, PgrR, MarA and MngR (Figure 5).

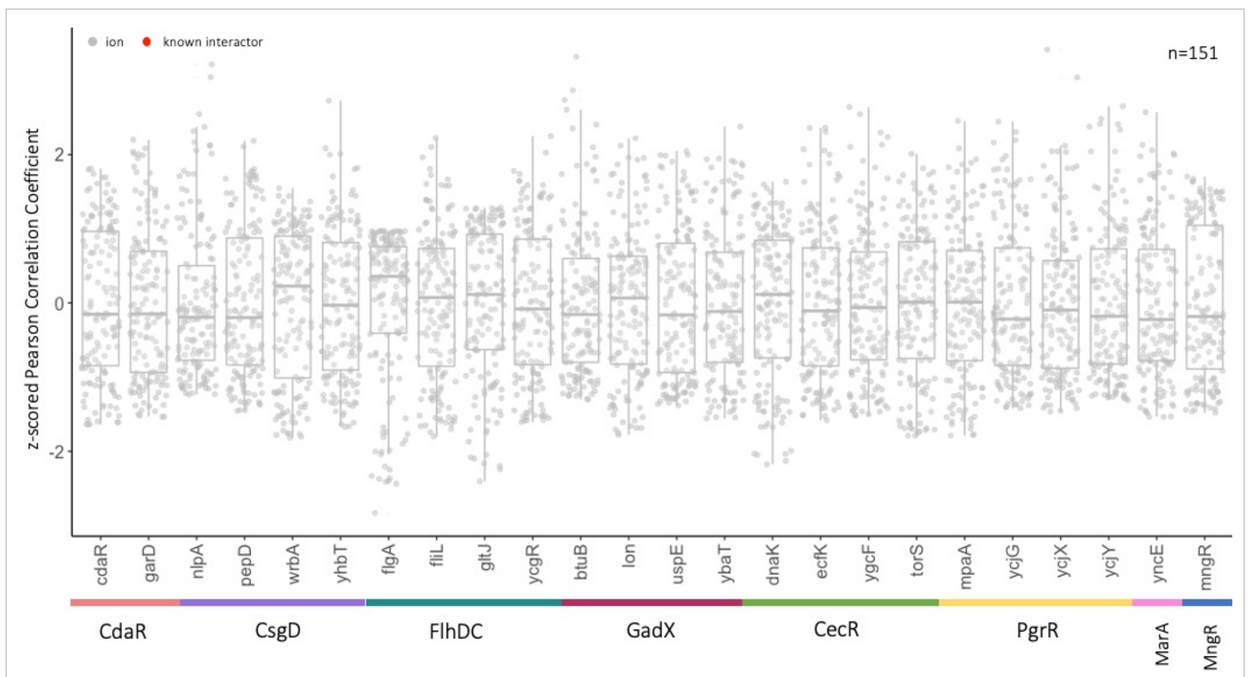


Figure 5. Correlation coefficients among 151 metabolites and 24 promoters regulated by 8 understudied TFs.

To identify the metabolites with the most biological and statistical relevance, we used 3 filtering approaches: (1) highest-scoring metabolites across all the promoters tested, (2) metabolites that belonged to a metabolic pathway enriched among the top 10% scoring metabolites, and (3) metabolites that participate in a reaction catalyzed by an enzyme whose gene is directly regulated by the TF of interest, a criteria that holds true for ~85% of TFs with known signaling metabolites[8].
Results were classified depending on how many criteria a metabolite fit for a specific TF (Figure 6).

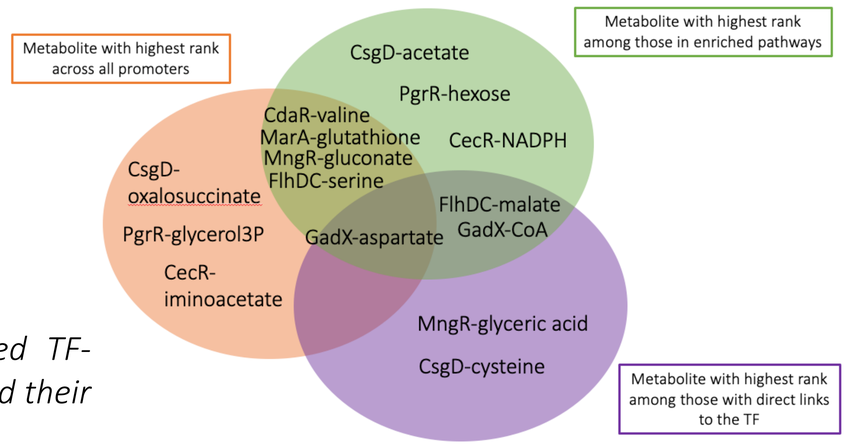


Figure 6. Fifteen identified TF-metabolite interactions and their confidence level.

Conclusions

We have designed a method to predict signaling metabolites of TFs that yielded correct results in 8/9 promoters tested. We identified ornithine and lysine as previously unknown metabolites that interact with ArgR and have an effect on ArgR affinity to DNA. Additionally, we predicted 15 interacting partners for 8 understudied TFs.

References

1. Pérez-Rueda et al. *Comput Biol Chem*, 2015.
2. Babu & Teichmann. *NAR*, 2003.
3. Santos-Zavaleta et al. *NAR*, 2018.
4. Fuhrer et al. *Analytical Chemistry*, 2011.
5. Zaslaver et al. *Nature Methods*, 2006.
6. Kochanowski et al. *Mol Syst Biol*, 2017.
7. Lempp et al. *Nat Comm*, 2019.
8. Ledezma-Tejeida et al. *Front Microbiol*, 2017.