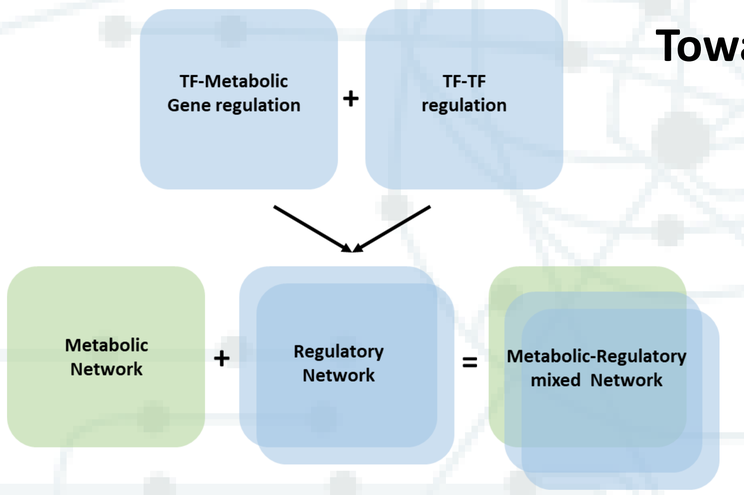
Aiming at the Automation of Genome-wide Regulatory Network Inference in Yeast

Diogo Couceiro [1,2,3(*)], Maria Pereira [1,2(*)], Carolina Parada [1,2], Romeu Viana [2,3], Miguel Cacho Teixeira [2,3], Pedro Tiago Monteiro [1,4]

[1] INESC-ID, Lisboa ; [2] iBB - Institute for Bioengineering and Biosciences and Department of Bioengineering, Instituto Superior Técnico, Universidade de Lisboa – Lisboa; [3] I4HB – Institute for Health and Bioeconomy - Laboratório Associado, Portugal; [4] Department of Computer Science and Engineering, Instituto Superior Técnico, Universidade de Lisboa – Lisboa

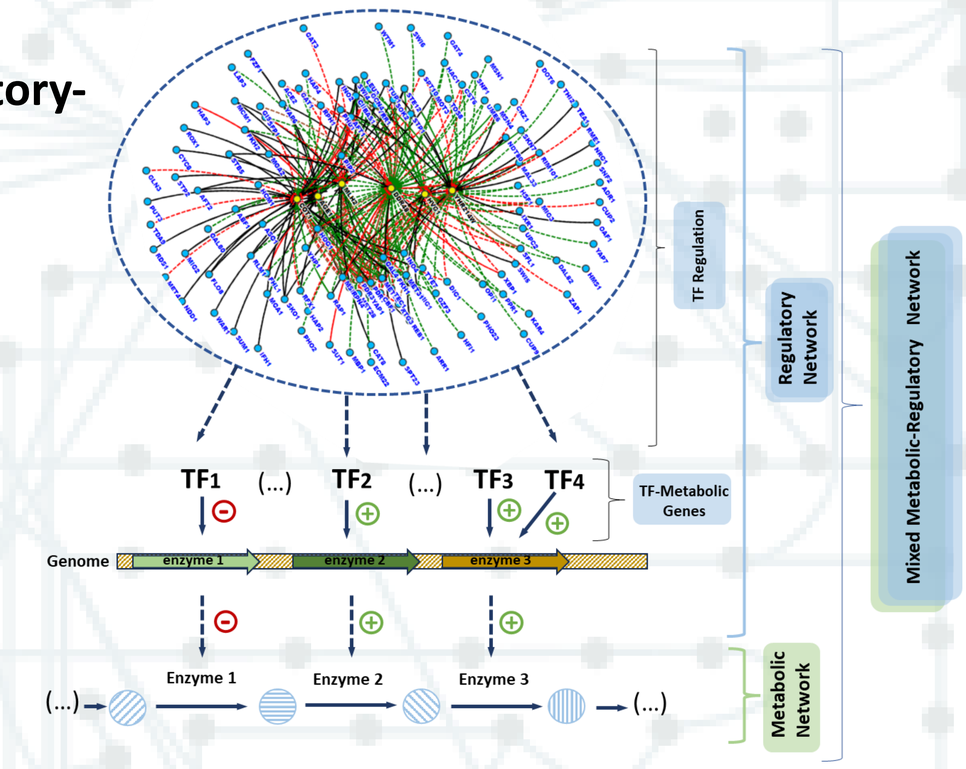
Metabolism is much more than a set of connected reactant-catalyst-product combinations, an assumption too often made by standard genome-scale metabolic models. Metabolism is complex and dynamic, and in close liaison with environmental conditions that control metabolism through signaling/regulatory pathways. Altogether, it's vital to consider all these players when wishing to represent cell behavior using mathematical models, in a way that they can be useful to increase our ability to optimize cell factories through transcription factor manipulation.

This work aims at developing tools for the automated inference of Boolean regulatory networks from a set of regulatory information, such as that present in Yeastract+ (containing around 200K TF-target associations for *Saccharomyces cerevisiae*) [Teixeira et al, NAR, 51: D785-D791, 2023], in several different environment conditions. Regulatory network topology provides interesting static information on the structure and hierarchy of genome-scale regulatory networks. However, further information on the impact of combinations of transcription factors (TF) in target gene (TG) expression is required to obtain a model that describes the dynamics of regulatory-metabolic systems. With that objective in mind, a set of Boolean forms was selected to be tested on global yeast regulatory data, to assess the adequacy of the chosen Boolean forms to be used as general rules of TFs-TG interactions. Further tools for the manual curation of the regulation of individual genes are being built to enable the fine-tuning of the developed models. Finally, these models will enable the prediction of the impact of environmental cues on metabolic outcome and the optimization of the use of genome-scale metabolic models in cell factory optimization through regulatory engineering.



Towards Integrated Mixed Regulatory-Metabolic Networks

When thinking about regulatory networks two levels of complexity have to be considered: regulation of metabolic gene expression by Transcription Factors (TF) and the complete regulatory network itself, comprising the transcription regulation of TF's themselves.



Metabolic Optimization by Manipulating TF Expression

Yeastract

Over 300K **TF-target gene associations**

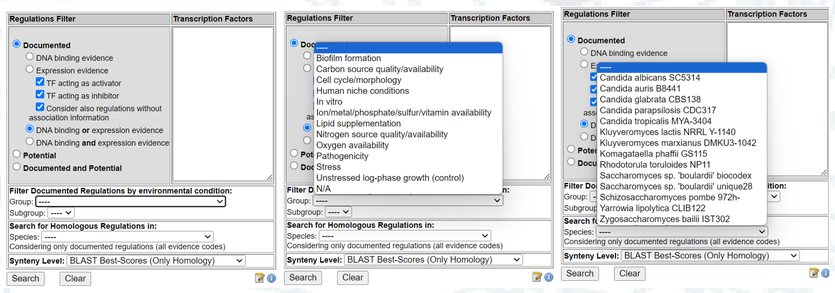
Metabolic genes from **Yeast 8.5.0**

TF-metabolic genes associations

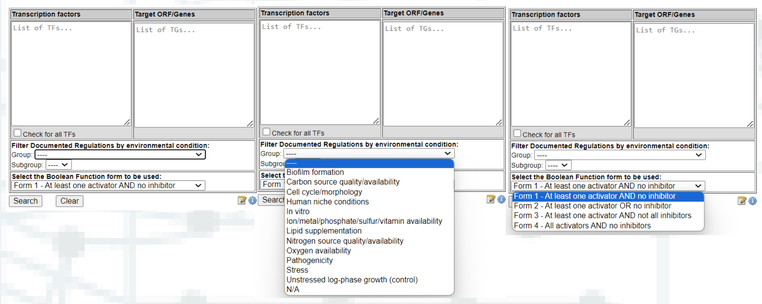
FBA analysis pre-defined objective function

Optimization

Predictions



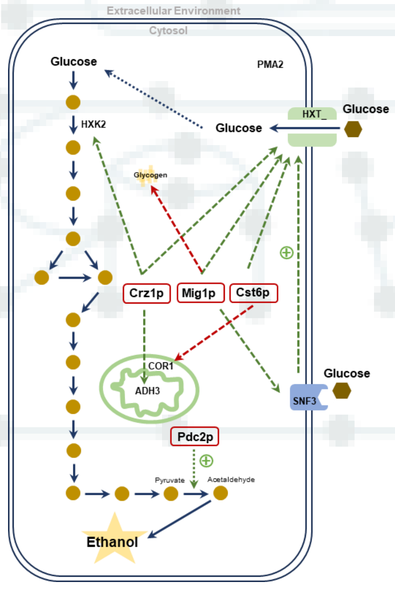
From Regulatory Associations to Boolean Networks

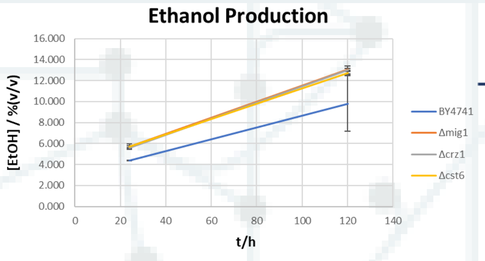


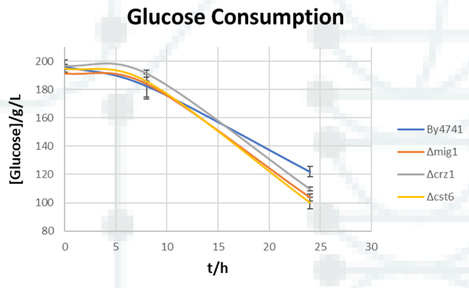
Optimizing Ethanol Production Through Transcription Factor Expression Manipulation: a Case Study

TFs	TF KO / under-expression Under(over)-expression of actv(inh) TGs						TF Over-expression Over(under)-expression of actv(inh) TGs						Best flux
	UE=0.00 (OE=1.25)	UE=0.00 (OE=1.50)	UE=0.25 (OE=1.25)	UE=0.25 (OE=1.50)	UE=0.50 (OE=1.25)	UE=0.50 (OE=1.50)	OE=1.25 (UE=0.00)	OE=1.25 (UE=0.25)	OE=1.25 (UE=0.50)	OE=1.50 (UE=0.00)	OE=1.50 (UE=0.25)	OE=1.50 (UE=0.50)	
	flux	flux	flux	flux	flux	flux	flux	flux	flux	flux	flux	flux	
Pdc2p	7.1903	97.1903	109.35	109.35	121.513	121.513	156.241	156.241	156.241	Infeasible	Infeasible	Infeasible	156.24100 | OE=1.25 (UE=0.00)
Mig1p	46.568	147.168	146.568	147.168	146.568	147.168	117.169	124.368	131.568	117.169	124.368	131.568	147.16800 | UE=0.00 (OE=1.50)
Cst6p	46.568	147.168	146.568	147.168	146.568	147.168	117.169	124.368	131.568	117.169	124.368	131.568	147.16800 | UE=0.00 (OE=1.50)
Rap1p	146.568	147.168	146.568	147.168	146.568	147.168	117.169	124.368	131.568	117.169	124.368	131.568	147.16800 | UE=0.00 (OE=1.50)
Gcr2p	146.568	147.168	146.568	147.168	146.568	147.168	117.169	124.368	131.568	117.169	124.368	131.568	147.16800 | UE=0.00 (OE=1.50)
Crz1p	46.568	147.168	146.568	147.168	146.568	147.168	117.169	124.368	131.568	117.169	124.368	131.568	147.16800 | UE=0.00 (OE=1.50)

Pdc2p
Mig1p
Cst6p
Crz1p







Pdc2p — Ongoing work

Mig1p
Cst6p
Crz1p
> BY4741

Confirmed in preliminary experimental validation!

Conclusions

Considering ethanol as a case study—an industrially relevant compound—preliminary experimental validation has shown promising results coherent with the predictions performed by manipulating TF-expression. However the most significant challenge corresponds to the next step, the inference of regulatory networks that encompass TF transcription regulation. This step is still undergoing development, shows itself to be promising based on previous results.

References

1. Teixeira MC, Viana R, Palma M, Oliveira J, Galocha M, Mota MN, Couceiro D, Pereira MG, Antunes M, Costa IV, Pais P, Parada C, Chaouiya C, Sá-Correia I, Monteiro PT. YEASTRACT+: a portal for the exploitation of global transcription regulation and metabolic model data in yeast biotechnology and pathogenesis. Nucleic Acids Res. 2023 Jan 6;51(D1):D785-D791. doi: 10.1093/nar/gkac1041. PMID: 36350610; PMCID: PMC9825512.