Improving prediction of essential genes using context-specific metabolic network ensembles

Pablo Rodríguez-Mier, Nathalie Poupin, Carlo de Blasio, Laurent Le Cam, Fabien Jourdan

To cite this version:

Pablo Rodríguez-Mier, Nathalie Poupin, Carlo de Blasio, Laurent Le Cam, Fabien Jourdan. Improving prediction of essential genes using context-specific metabolic network ensembles. European RFMF Metabomeeting 2020, Jan 2020, Toulouse, France. hal-02948371

HAL Id: hal-02948371
https://hal.inrae.fr/hal-02948371v1
Submitted on 24 Sep 2020

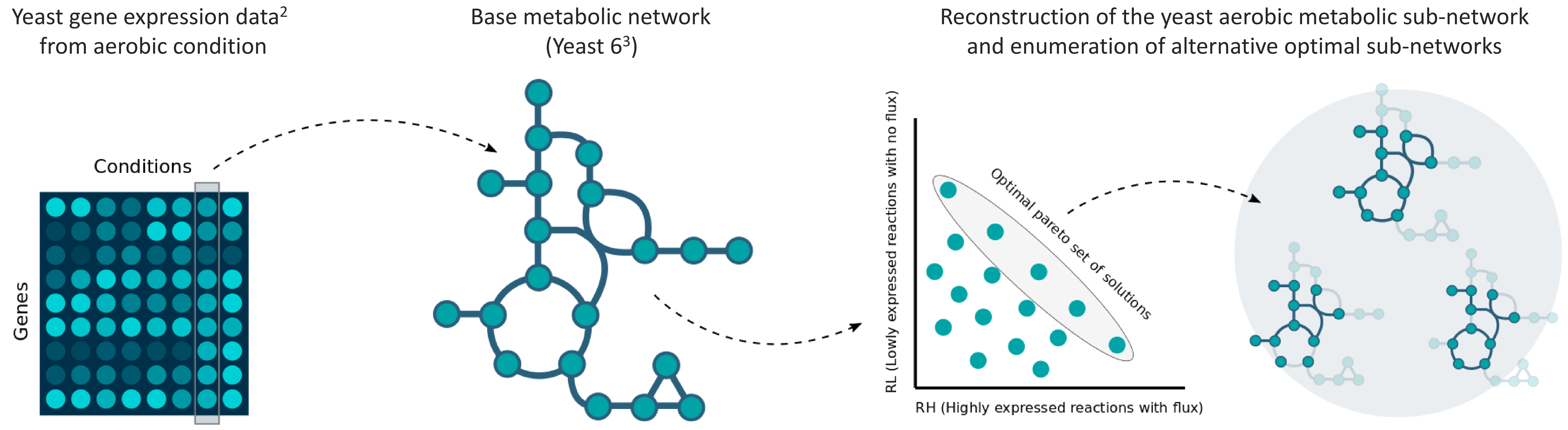
IMPROVING PREDICTION OF ESSENTIAL GENES USING CONTEXT-SPECIFIC METABOLIC NETWORK ENSEMBLES

PABLO RODRIGUEZ-MIER[a], NATHALIE POUPIN[a], CARLO DE BLASIO[b], LAURENT LE CAM[b], FABIEN JOURDAN[a]
[a]Toxalim (Research Center in Food Toxicology), Université de Toulouse, INRAE, ENVT, INP-Purpan, UPS, 31300 Toulouse, France
[b]Institut de Recherche en Cancérologie de Montpellier, Campus Val d'Aurelle, 34298 Montpellier, France

Motivation

- Context-specific constraint-based modelling techniques are useful to reconstruct the metabolic network for a particular tissue/condition and make predictions.
- Current methods generate only one network per condition using gene expression, although there is usually no single network that better fits the data.
- We hypothesize that using an ensemble of alternative optimal networks could improve on average the predictions made by a single network.
- To test this hypothesis, we curated a dataset of essential aerobic genes that are directly implicated in the metabolism of the *Saccharomyces Cerevisiae*.
- Then, we reconstructed thousands of metabolic networks using a modification of iMAT[1] to enumerate alternative solutions based on two MIP cut methods.
- Results using different ensemble approaches to combine the predictions show that ensembles can be used to improve the classification error for essential genes.

Reconstruction & enumeration of metabolic networks


Yeast gene expression data[2]
from aerobic condition

Base metabolic network
(Yeast 6[3])

Reconstruction of the yeast aerobic metabolic sub-network
and enumeration of alternative optimal sub-networks

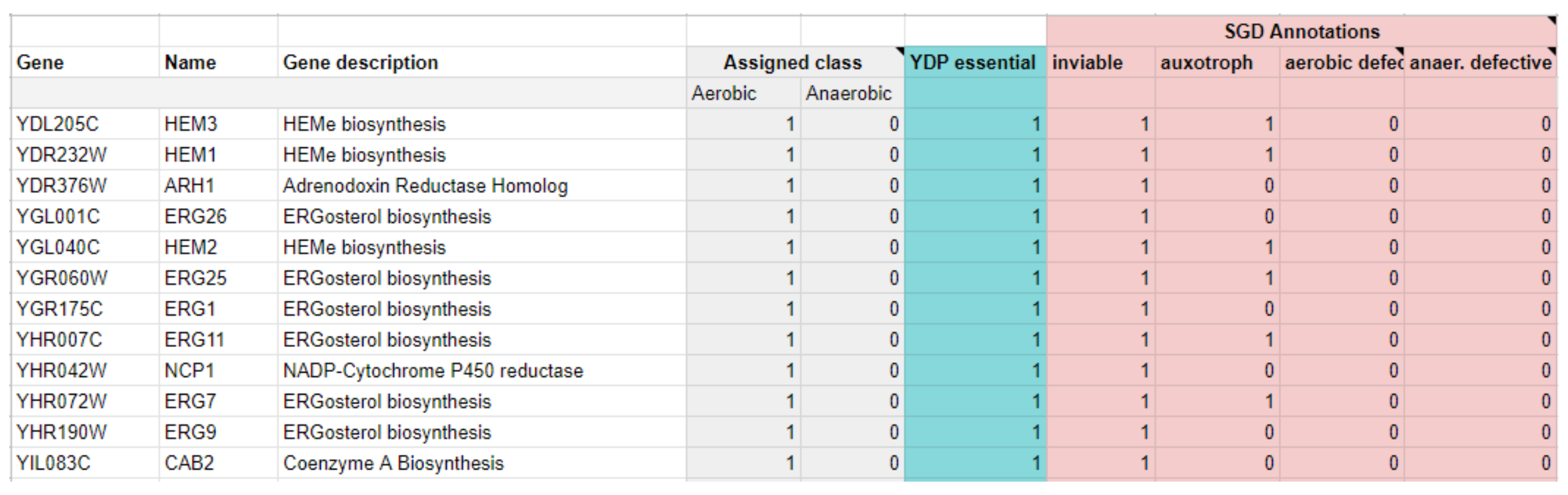
Dataset curation (essential genes under aerobic conditions for *S. Cerevisiae*)

SGD Saccharomyces GENOME DATABASE

Saccharomyces Genome Deletion Project[4]

Yeast 6 metabolic genes

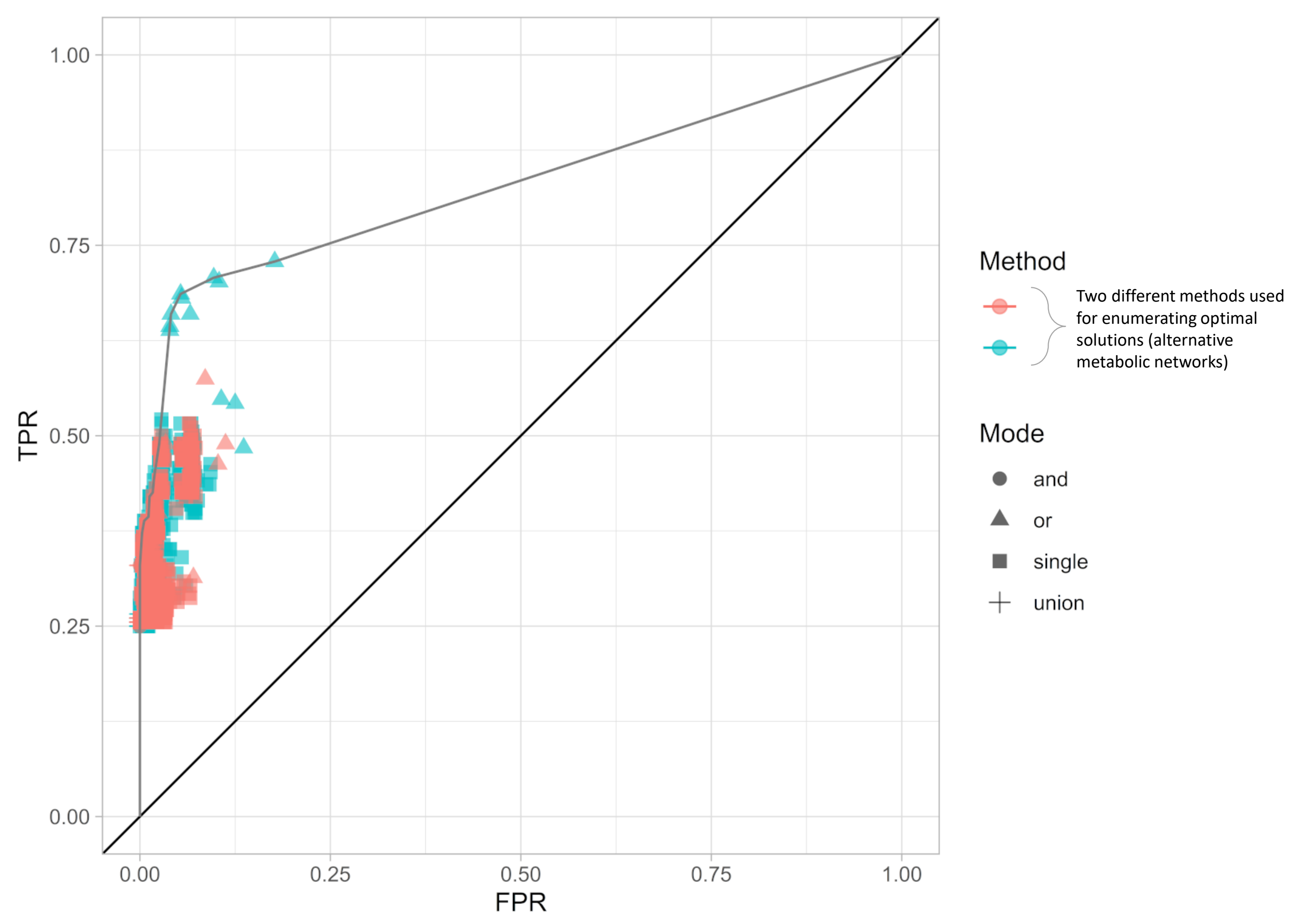
Manual curation



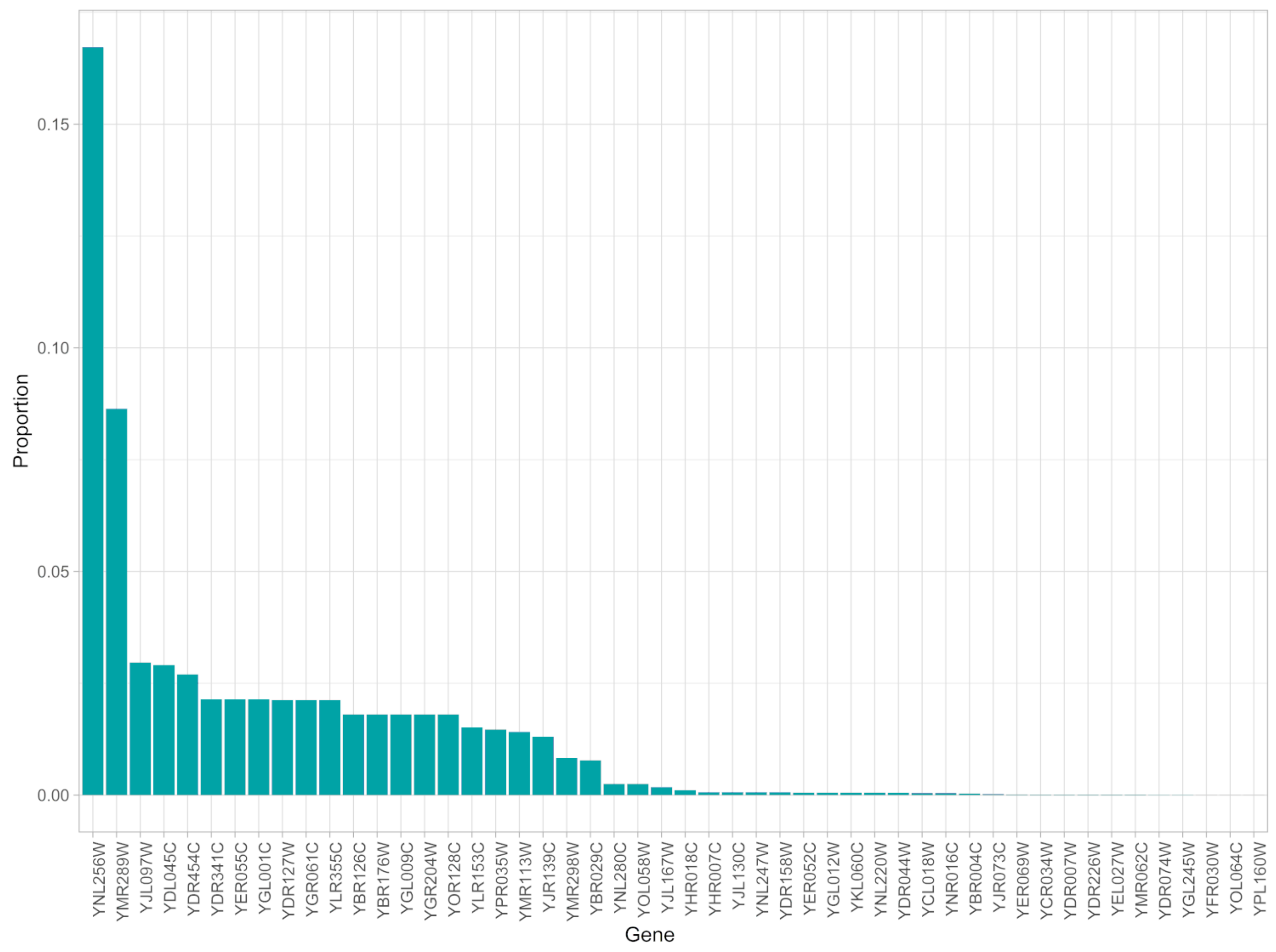
Gene	Name	Gene description	Assigned class		YDP essential	SGD Annotations			
						inviable	auxotroph	aerobic defec	anaer. defective
			Aerobic	Anaerobic					
YDL205C	HEM3	HEMe biosynthesis	1	0	1	1	1	0	0
YDR232W	HEM1	HEMe biosynthesis	1	0	1	1	1	0	0
YDR376W	ARH1	Adrenodoxin Reductase Homolog	1	0	1	1	0	0	0
YGL001C	ERG26	ERGosterol biosynthesis	1	0	1	1	0	0	0
YGL040C	HEM2	HEMe biosynthesis	1	0	1	1	1	0	0
YGR060W	ERG25	ERGosterol biosynthesis	1	0	1	1	1	0	0
YGR175C	ERG1	ERGosterol biosynthesis	1	0	1	1	0	0	0
YHR007C	ERG11	ERGosterol biosynthesis	1	0	1	1	1	0	0
YHR042W	NCP1	NADP-Cytochrome P450 reductase	1	0	1	1	0	0	0
YHR072W	ERG7	ERGosterol biosynthesis	1	0	1	1	1	0	0
YHR190W	ERG9	ERGosterol biosynthesis	1	0	1	1	0	0	0
YIL083C	CAB2	Coenzyme A Biosynthesis	1	0	1	1	0	0	0

Evaluation & conclusions

- We used three different schemes to combine the predictions of the ensembles: **or** (essential if any metabolic network predicts essential), **and** (essential if all networks predict essential), and **union** (essential if the single network generated from the union of the alternative networks predicts essential).
- We observed that the "**or** ensemble" strategy increases the **True Positive Rate** without increasing too much the **False Positive Rate**.
- Around 50 essential genes could only be detected using the "or-ensemble" of networks, from which YNL256W (FOL1) was the most common detected essential gene and YFR030W (MET10), YOL064C (MET22), YDR007W (TRP1) among others (see chart below on the right side) are among the least detected genes.



Models in ROC space. Each point corresponds to the FPR/TPR achieved by each model (single network or ensemble), using 12 different combinations of parameters (gene thresholds) to build the models.



True essential genes only detected in "**or" ensembles** and proportion of networks that detected them within the ensemble.

References

1. Shlomi, Tomer, et al. "Network-based prediction of human tissue-specific metabolism." Nature biotechnology 26.9 (2008): 1003.
2. Rintala, Eija, et al. "Low oxygen levels as a trigger for enhancement of respiratory metabolism in Saccharomyces cerevisiae." BMC genomics 10.1 (2009): 461.
3. Heavner, Benjamin D., et al. "Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance." Database 2013 (2013).
4. Winzeler, Elizabeth A., et al. "Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis." science 285.5429 (1999): 901-906.