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-02948371
Submitted on 24 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
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.
- Results using different ensemble approaches to combine the predictions show that ensembles can be used to improve the classification error for essential genes.

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)

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.

References