P56 - Predicting and modeling genotype-phenotype associations in yeast metabolic networks

Helder Lopes[1], Fábio Ramalho[1], Isabel Rocha[1]

[1]Centre of Biological Engineering, Department of Biological Engineering, University of Minho, 4710-057 Campus de Gualtar, Braga, Portugal (helderlopes@ceb.uminho.pt)

Over the last 15 years, several genome-scale metabolic models (GSSMs) of *Saccharomyces cerevisiae* were reconstructed and published. The in silico representation of the interaction network between all the system components is carried out to predict the physiological behavior of a microorganism, under different environmental and genetic perturbations. However, gene knockout predictions are usually assessed and validated using merely gene essentiality data. Saccharomyces Genome Database (SGD) [1] is a powerful web-accessible resource that comprises functional structured information of budding yeast genes. SGD contains information about over 180 different observed types of phenotypes of which nearly 10% can be predicted using GSMMs. These data can provide an additional layer for curation and validation of metabolic models, as well as contribute to model improvements and to gain insights into yeast physiology.

In this study we have assessed the predictive accuracy of GSSMs based on single-gene deletions, by comparing experimental data present in SGD with computational simulations. Since the phenotypical behavior upon a gene deletion depends on the strain background, media and other environmental conditions, we performed a thoroughly characterization and (re)curation of the in vivo experiments to closely mimic these evidences in silico. Nearly 3000 different phenotypic reported cases were evaluated using two different constraint-based approaches (pFBA [2] and LMOMA [3]), which allow a direct association between genetic data and metabolic fluxes. In parallel, a Jupyter Notebook platform was also developed, aiming to serve as a possible validation tool for new yeast GSMMs, using the curated SGD-based dataset.

We observed that, despite all the recent efforts and advances in the reconstruction and annotation of GSMMs, there is still a lot of opportunities for improvements in the models' predictive ability. Most of the observed mismatches result from structural issues in network reconstructions or due to the lack of regulatory information. To address these issues, several strategies were investigated, including changes in gene-protein-reaction associations and reversibility of reactions in the network, aside from the formulation of a new biomass equation, based on the experimental determination of its macromolecular composition, to which several cofactors, that surprisingly had not been represented in the original biomass reaction, were also added. For example, this last modification led to significant improvements in the prediction of auxotroph-inducing mutations and lethal knockouts, which should enable us to more effectively engineer yeast as a cell factory.

REFERENCES: [1] Cherry JM et al. 2012; 40(Database issue):D700-5; [2] Lewis NE et al. Mol Syst Biol 2010, 6, 390; [3] Lewis, NE et al. Nat Rev Microbiol 2012, 10, 291-305.