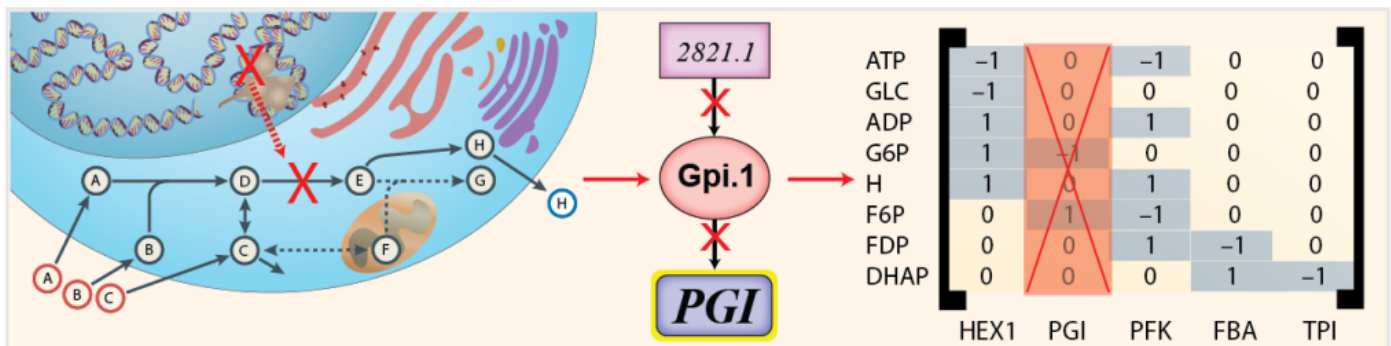


Using genome-scale models to predict biological capabilities

Rapid advances in DNA sequencing and synthesis technologies have revolutionized the study of life sciences. With these new technologies come new challenges. The massive amount of new data generated using these technologies must be curated, managed, and analyzed. In other fields this is known as the problem of “big data” or the “data deluge”. In biology, the problem has led to critiques of the scientific process itself, prompting critics to claim that modern biology is “high-throughput, low-input and no-output”.

Thus, new tools and methods are required to harness and analyze this exponentially increasing amount of biological data. One such method that has demonstrated success at integrating large datasets of diverse biological datatypes is called Constraint-based Reconstruction and Analysis (COBRA). This approach has demonstrated the ability to predict a range of cellular functions including cellular growth capabilities and the effect of gene knockouts at the genome-scale. It has been successfully applied in metabolic engineering, drug development, and studies of organismal and enzyme evolution.



COBRA tools and methods at the genome-scale have been under development since the first whole genome sequences appeared in the mid-1990s. The first genome-scale reconstruction of a metabolic network was created for *Haemophilus influenzae*, appearing four short years after this first genome sequence was established. Since then reconstructions have expanded to cover more than 100 unique organisms and increased in sophistication to the level where they enable predictive biology. COBRA methods rely on constructing a mechanistic basis for the biochemical and genetic processes that underlie cellular functions. Networks are constructed based on genome annotation, biochemical characterization, and the published scientific literature on the target organism. A network reconstruction can be converted into a mathematical format and thus lend itself to computational simulations.

The fundamentals of the COBRA approach and its uses are described in a recent primer article published by Cell Press, which lays out the constraint-based methodology in six sections. The sections are illustrated with detailed figures and the text explains the computational approaches used and how to apply them yourself. The primer also provides a table of resources for the keenly

interested reader wishing to delve deeper into the subject. Each of the six sections addresses a grand challenge in today's world of "big data" biology:

Section 1 addresses the collection and organization of disparate data types for an organism of interest and the conversion of this information into a biochemical reaction network reconstruction.

Section 2 focuses on the conversion of biochemical reconstructions into computational models that can be used to predict metabolic capabilities.

Section 3 explains the validation of qualitative model predictions and their reconciliation with experimental results to discover new biology.

Section 4 details advanced genome-scale modeling methods used to make quantitative predictions.

Section 5 highlights the integration of high-throughput "omics" data with genome-scale models.

Section 6 examines the future of genome-scale modeling and the prospect of extending these principles to processes beyond metabolism, including transcription and translation.

Working your way through this Primer will allow you to address grand questions in biology using "big data".

Publication

[Using Genome-scale Models to Predict Biological Capabilities.](#)

O'Brien EJ, Monk JM, Palsson BO

Cell. 2015 May 21