Applying machine learning tools to glean complex regulatory networks in metabolism

Wenfa Ng

Unaffiliated researcher, Singapore, Email: ngwenfa771@hotmail.com

Abstract

Metabolism captures the entire set of chemical reactions that occur in the cell. However, metabolism is complex and difficult to decipher using conventional biochemical techniques. Specifically, metabolism transcends multiple level of organization in biology: moving from genes, to mRNA, proteins and metabolites, each of which contains a distinct and usually cross-interacting layer of regulation. Hence, metabolism should be understood at multiple levels of complexity, but it is difficult to discern complex inter-relationships across multiple biological levels in large dataset such as those increasingly available from omics technologies. Given the ability of machine learning tools to glean relationships between multiple cross-interacting variables from large complex dataset, it may be useful in applying machine learning tools for understanding complex regulatory networks in metabolism. The goal is to understand metabolism at the single bacterial cell level across different nutritional and environmental conditions. Such an endeavour would require multiple types of data involving genome, transcriptome, proteome, and metabolome derived from different analytical techniques and instruments. Most important to the understanding of metabolism and its regulatory network is the need to obtain correlations between different types of data, such as that between transcriptome and proteome to understand how different RNA transcripts are translated to proteins and enzymes. For example, areas where there is a negative correlation between transcript counts and protein abundance may point to feedback inhibition, while an overabundance of protein over RNA transcripts would imply a positive feedback relationship. While such relationships could be gleaned for individual protein and transcript from analyzing large datasets, knowledge of the mediators of the positive or negative feedback relationship remain key to building a full mechanistic model that describes the regulatory network of metabolism in a single cell. But, herein lies the challenge, determining mediator of a phenomenon such as which repressor inhibits the translation of a transcript remains difficult for such data-driven exercise in obtaining correlations due to the inability to profile causative relationships. Hence, experiments that combine human expert knowledge of metabolism and screening tools remain essential in complementing machine learning algorithms in understanding the mediators of metabolic control at the single cell level. Since each metabolic reaction is likely controlled by a mediator, identification of the full complement of mediators for gaining full metabolic control of the cell at the system level remain a tough challenge, especially for reactions not involved in central carbon metabolism.

Keywords: genome, proteome, transcriptome, metabolome, metabolism, regulatory network, single cell, central carbon metabolism, mediator, correlations,

Subject area: biochemistry, cell biology, microbiology, bioinformatics, computational biology,

Conflicts of interest

The author declares no conflicts of interest.

Funding

No funding was used in this work.