Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Quantitative -omic data empowers bottom-up systems biology

James T Yurkovich[1,2] and Bernhard O Palsson[1,2,3]

The large-scale generation of '-omic' data holds the potential to increase and deepen our understanding of biological phenomena, but the ability to synthesize information and extract knowledge from these data sets still represents a significant challenge. Bottom-up systems biology overcomes this hurdle through the integration of disparate -omic data types, and absolutely quantified experimental measurements allow for direct integration into quantitative, mechanistic models. The human red blood cell has served as a starting point for the application of systems biology approaches and has been the focus of a recent burst of generated quantitative metabolomics and proteomics data. Thus, the red blood cell represents the perfect case study through which to examine our ability to glean knowledge from the integration of multiple disparate data types.

Addresses
[1] Department of Bioengineering, University of California, San Diego, La Jolla, USA
[2] Bioinformatics and Systems Biology Program, University of California, San Diego, La Jolla, USA
[3] Department of Pediatrics, University of California, San Diego, La Jolla, CA, USA

Corresponding author: Palsson, Bernhard O (palsson@ucsd.edu)

Current Opinion in Biotechnology 2018, **51**:130–136

This review comes from a themed issue on **Systems biology**

Edited by **Nathan Price** and **Eran Segal**

https://doi.org/10.1016/j.copbio.2018.01.009

Introduction

Over the last two decades, the life sciences have witnessed a paradigm shift brought on by the development of high-throughput '-omic' technologies. With the advent of these technologies, systems biology emerged as a way to holistically integrate the new data being generated. Integrative thinking was not something previously absent in molecular biology, it was just that high-throughput -omic technologies were making the scale of these integrative inquiries much larger [1]. With the availability of full genome sequences and other data, more and more researchers embraced the promise in systems biology and began to develop ways to bridge the gap between -omic data and computational modeling efforts [2].

Some of the first cell-scale computational models were published in the late 1980s [3]. These enzyme kinetic models detailed the known metabolic network of the human red blood cell (RBC). Why study the RBC? The reasoning was simple: if systems biology cannot be successfully applied to the simplest human cell, then why attempt to study more complex ones? Indeed, simple systems are the best starting point for the application of systems biology. The RBC is therefore a logical starting point for the development and application of systems biology methods because of its simplicity and intrinsic experimental accessibility. RBCs are also of great importance for our understanding of human health and physiology — over 84% of all human cells by count are RBCs [4]. Transfusion medicine represents an integral part of healthcare, with approximately 85 million RBC units transfused worldwide annually [5]. The systems biology analysis of the health of stored RBCs is thus a productive focus from a basic and applied standpoint.

Within the last several years, -omic technologies have been exploited to study RBCs under refrigerated storage for use in transfusion medicine [6•,7•] in an attempt to understand and elucidate the underlying physiological changes that occur because of the artificial environment [8]. Concurrently, computational biologists have worked to develop new mathematical modeling frameworks that can use these data. Because of the inherent quantitative nature of these models, however, their utility is only fully realized with *quantitative data*. In this context, *quantitative* data implies the use of standards to absolutely quantify the abundance of measured species; the output is data with quantified units (e.g., g/L, mM), rather than *qualitative data* that have relative units (e.g., arbitrary units, relative signal). While there have been several important studies that have used qualitative data effectively, the future of systems biology modeling efforts will hinge upon the availability of high quality quantitative data.

In this article, we discuss some of the recent work in -omic data generation and corresponding computational methodologies. In particular, we focus on how the use of quantitative data aids modeling efforts and enables new questions to be asked. We review studies on a variety of organisms, using the RBC as a case study throughout. We first discuss -omic data types and new experimental techniques that will likely prove to be valuable for the

field; we then survey a variety of computational modeling approaches that integrate these data types and review important advances to date; finally, we close with perspectives on where the field of systems biology is with respect to the integration of -omic data and what the next steps might be.

A variety of -omic data types describe cellular physiology

A cell is a system of interconnected complex systems described by a variety of -omic data types [9••]. Metabolomics data provide a snapshot of the cellular biochemistry that details energy production [10]. Fluxomics measurements — the use of isotopic tracers (e.g. ^{13}C) — yield an understanding of the flux state of a metabolic network [11]. Proteomics data allow for an understanding of the abundance, localization, and interactions of proteins, the cellular machinery underlying all metabolic processes and regulatory mechanisms [12]. Lipidomics technologies have enabled the in-depth characterization of the cellular membrane, including signaling, transport, and respiration mechanisms [13]. Researchers utilize one or more of these techniques to interrogate their system of interest, leading to rich information and important phenomenological observations. Recently, the RBC has been the source of much -omic data, yielding advances in analytic experimental techniques, rich data sets, and driving computational method development (Figure 1).

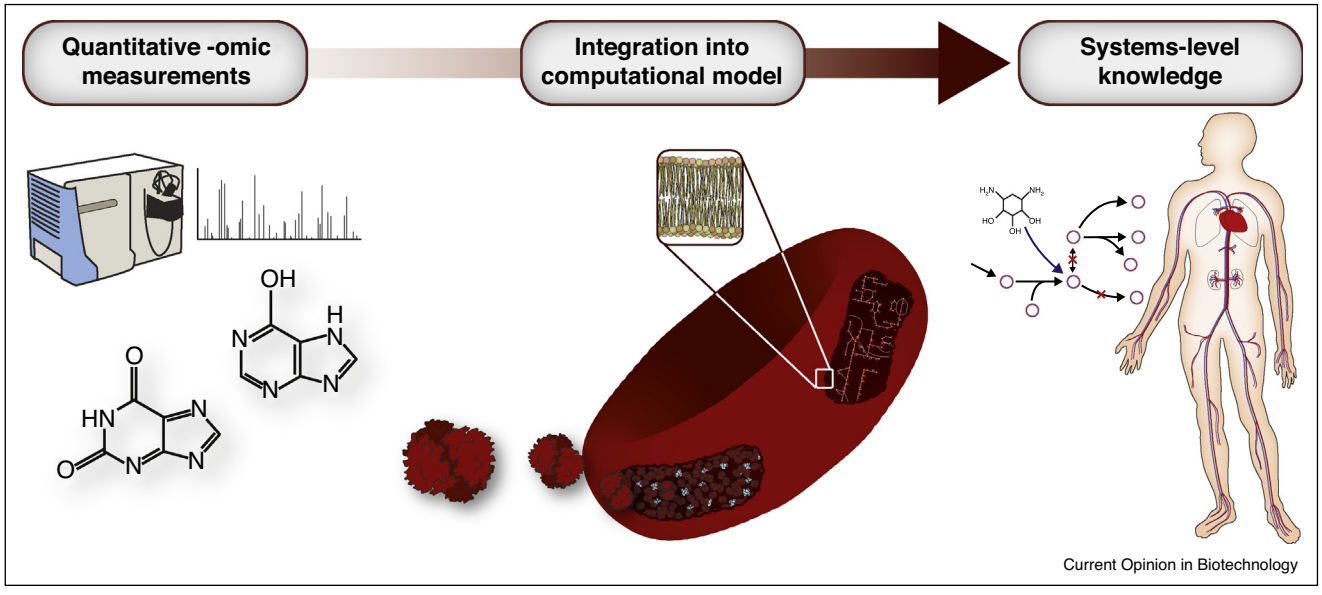
Metabolomics

During the storage of RBCs in blood bags at 4 °C, a variety of changes occur within the cells that impact their ability to carry oxygen and generate energy upon transfusion into a patient. These morphological and biochemical changes — collectively referred to as the 'storage lesion' [14] — have been thoroughly explored through the use of metabolomics data over the last decade [6•]. These studies have explored the impact of various perturbations to the storage media on the metabolic function of the RBCs over the course of the 42 day storage period by taking weekly time points. A set of metabolites was identified that serves as storage-age biomarkers [15•]. Several of these studies have provided very informative data sets [16–22], but their utility for systems biology modeling is limited due to their qualitative nature; while raw signals can provide meaningful statistical analyses [23,24], they are inherently incapable of being integrated into quantitative models.

More recently, there has been an influx of quantitative data characterizing the RBC storage process. Perhaps the most complete characterization to date was produced by Bordbar *et al.* [25•], providing a much finer resolution on the temporal dynamics observed during storage by taking time points every 3–4 days. The RBC community is embracing the trend of quantitative data generation, producing more absolutely quantified data sets [26,27]. Similar data have also been produced in other cells and organisms, such as the human platelet [28,29], *Escherichia coli* [30], and *Saccharomyces cerevisiae* [31].

With such rich data available, the onus has been on the modeling community to help realize the full potential of these data. As a result, there have been several different computational approaches that utilize quantitative metabolomics data. Some studies have relied on statistical

Figure 1



Quantitative -omic data allows for integration into quantitative mechanistic models capable of generating phenotypic predictions.