BBA - Molecular Basis of Disease xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

BBA - Molecular Basis of Disease

journal homepage: www.elsevier.com/locate/bbadis

Metabolic modeling helps interpret transcriptomic changes during malaria☆

Yan Tang[a], Anuj Gupta[b], Swetha Garimalla[c], , The MaHPIC Consortium[d], Mary R. Galinski[e,f], Mark P. Styczynski[a], Luis L. Fonseca[b], Eberhard O. Voit[b,*]

[a] School of Chemical and Biomolecular Engineering, Georgia Tech, Atlanta, GA 30332, USA
[b] Department of Biomedical Engineering, Georgia Tech, Atlanta, GA 30332, USA
[c] School of Biological Sciences, Georgia Tech, Atlanta, GA 30332, USA
[d] Malaria Host-Pathogen Interaction Center, USA
[e] Emory Vaccine Center at Yerkes, Emory University, 954 Gatewood Road, EVC 003, Atlanta, GA 30329, USA
[f] Department of Medicine, Division of Infectious Diseases, Department of Microbiology and Immunology, Emory University School of Medicine, Atlanta, GA 30329, USA

ARTICLE INFO

Keywords:
Biochemical systems theory
Dynamic model
Generalized mass action system
Malaria
Metabolic modeling
Transcriptomics

ABSTRACT

Disease represents a specific case of malfunctioning within a complex system. Whereas it is often feasible to observe and possibly treat the symptoms of a disease, it is much more challenging to identify and characterize its molecular root causes. Even in infectious diseases that are caused by a known parasite, it is often impossible to pinpoint exactly which molecular profiles of components or processes are directly or indirectly altered. However, a deep understanding of such profiles is a prerequisite for rational, efficacious treatments. Modern omics methodologies are permitting large-scale scans of some molecular profiles, but these scans often yield results that are not intuitive and difficult to interpret. For instance, the comparison of healthy and diseased transcriptome profiles may point to certain sets of involved genes, but a host of post-transcriptional processes and regulatory mechanisms renders predictions regarding metabolic or physiological consequences of the observed changes in gene expression unreliable. Here we present proof of concept that dynamic models of metabolic pathway systems may offer a tool for interpreting transcriptomic profiles measured during disease. We illustrate this strategy with the interpretation of expression data of genes coding for enzymes associated with purine metabolism. These data were obtained during infections of rhesus macaques (*Macaca mulatta*) with the malaria parasite *Plasmodium cynomolgi* or *P. coatneyi*. The model-based interpretation reveals clear patterns of flux redistribution within the purine pathway that are consistent between the two malaria pathogens and are even reflected in data from humans infected with *P. falciparum*. This article is part of a Special Issue entitled: Accelerating Precision Medicine through Genetic and Genomic Big Data Analysis edited by Yudong Cai & Tao Huang.

1. Introduction

In contrast to many other complex diseases, such as cancer, Crohn's disease, or metabolic syndrome, infectious diseases have the distinction of a clear root cause: a pathogen has invaded the body and was not stopped by the host's natural immune defenses. If the pathogen can be eliminated with medical or pharmaceutical means, the disease has a straightforward cure. However, in many cases this is not directly possible, or it requires a relatively long period of time, during which the patient is at risk of deteriorating, with possibly lethal consequence. In these cases, the root cause becomes almost immaterial, and it is the complex system of interactions between the pathogen and the host that needs to move to the center of attention [1]. The intriguing aspect of these interactions is that we often have no real understanding of which specific subsystems in the host or the pathogen are turned on or off, so that any molecular characterization of the disease, or any attempt of a targeted intervention, becomes an enormous challenge. As an illustration, we use in this paper a case study of malaria, which afflicts > 200 million people world-wide and kills about half a million individuals per year, many of whom are children [2]. While malaria is initially a disease of the blood, it quickly affects other tissues and organ functions and triggers uncounted responses of the host's defense systems. To measure the molecular state of a person is therefore an unsurmountable problem. The Malaria Host Pathogen Interaction Center (MaHPIC; www.systemsbiology.emory.edu/) actually offers the great advantage of allowing us to collect very large datasets on infections of different

☆ This article is part of a Special Issue entitled: Accelerating Precision Medicine through Genetic and Genomic Big Data Analysis edited by Yudong Cai & Tao Huang.

* Corresponding author.
E-mail address: Eberhard.voit@bme.gatech.edu (E.O. Voit).
URL: undefined

http://dx.doi.org/10.1016/j.bbadis.2017.10.023
Received 6 August 2017; Received in revised form 27 September 2017; Accepted 17 October 2017

monkey species with different malaria parasites. These non-human primates (NHPs) present with symptoms of malaria that are very similar to those in humans and permit experiments that would not be possible in humans, due to ethical and other considerations.

The complexity of host responses to parasites poses a grand challenge. Namely, it is not even clear what exactly should optimally be measured, even if one had the luxury of being able to obtain measurements frequently. Of course, some experimental targets are quite evident. For instance, one can certainly measure the numbers of pathogens on a regular basis during an infection in order to characterize the extent of the disease. Such measurements can then be used to develop models of specific aspects of the disease [3–5]. One can also measure generic physiological markers like body temperature and blood cell counts, which reflect the severity of the disease at a high level. However, if the goal is to identify specific drug targets, a much deeper understanding of the molecular events during an infection is required. This necessity of a better characterization of processes is problematic, because it is generally much more difficult to measure processes than states. As a consequence, disease investigations must usually resort to measuring molecular profiles. The good news is that the –omics revolution has rendered it possible to assess molecular profiles incomparably more comprehensively than just a couple of decades ago. For instance, we can relatively easily and reliably measure the expression of most genes, and in the process distinguish between host and pathogen genes. Modern mass spectrometry has rendered it possible to establish profiles of many thousands of native and foreign metabolites and their break-down products, even though it is not always entirely clear how such high-throughput results are to be interpreted. While not yet as definitive as genome analyses, proteomics, metabolomics, and lipidomics offer a glimpse into the abundances of subclasses of metabolites, proteins, and lipids. Taken together, modern biology allows us to convert small volumes of biological samples into enormous datasets.

The sheer sizes of –omics datasets pose challenges that are relatively new to the field of biology [6]. Namely, it is no longer easy to discern valid information or true signals in the data from uncertainties, variability, and noise. Every diseased individual is different, and many differences in gene expression within a patient cohort may simply be manifestations of their genetic make-up and health histories. As a pertinent example in our case study, which we will discuss later in this article, seemingly similar macaques responded very differently to malaria infections, with some suffering relatively lightly, some very severely, and one not even surviving [7]. Clearly, humans and monkeys are dynamical entities whose features change over time. Also, of course, all experiments are burdened with certain inaccuracies, which may not be fully characterized. As a consequence of these and other complications, the expression of a given gene or protein at a given time point may be suggestive of a biomarker of disease, but it may also be a spurious event.

In this article, we describe, as a proof of principle, a computational strategy for approaching the complex questions raised in the previous paragraphs. We use malaria as an example and discuss how different types of experiments and computational analyses have shed light on unforeseen aspects of the disease. However, we certainly do not claim to have obtained complete or definitive answers to the questions we had asked at the beginning of our analysis. In other words, this article focuses on strategies rather than results.

2. Material and methods

2.1. A brief background on malaria

In all types of malaria, the sporozoite form of the *Plasmodium* parasite enters the human or NHP host through a mosquito bite. Moving quickly with the blood stream, the sporozoites soon reach the liver, where they infect hepatocytes. Depending on the species, the pathogens may remain in the liver for a long time in the form of hypnozoites, or they multiply aggressively and over time release thousands of merozoites into the bloodstream. The merozoites invade red blood cells (RBCs), where they multiply and, after a day or two, depending on the species, are released and infect other RBCs. Eventually, some of the parasites transform into sexual forms, called gametocytes, which are taken up by another mosquito. They mate within the mosquito, thereby completing the pathogen's life cycle.

The responses of the human and NHP hosts to the *Plasmodium* parasite are multifold, complicated and in very many aspects ill understood. It is not surprising that the invasion and destruction of RBCs typically leads to anemia. However, the details of how the disease affects the erythropoietic system and thus the dynamics of new RBC formation, as well as their aging and removal, are not well understood. A second clearly affected host subsystem is the immune system. The presence of the pathogen very definitively triggers numerous cellular and humoral immune responses, but the chains of events leading to these responses have remained obscure.

The reasons for these gaps in our understanding are manifold. First and foremost, the immune system is exceedingly complicated. It contains uncounted components in the form of different immune cells and specific proteins, such as immunoglobulins and cytokines, whose roles are not always exactly known. Confounding the situation is the difficulty of obtaining sufficiently large high-quality datasets. Yet another puzzle is the observation that the host often seems to clear an infection, only to suffer a relapse, which has different characteristics than the initial infection. Even this very coarse overview of a few aspects of malaria will render it evident that the disease is systemic and that large numbers of physiological subsystems interact in a life-or-death effort to control the disease.

2.2. Generic data-based characterization of a complex disease

Extracting information from large omics datasets has been compared to “drinking from a firehose.” Yet, even comprehensive attempts to measure pertinent data are not always sufficient. Within the context of our case study, our Malaria Host Pathogen Interaction Center (MaHPIC; www.systemsbiology.emory.edu/) has been collecting frequent –omics datasets to characterize the process accompanying malarial infections in NHPs. Although well-equipped and well-funded, this effort has been encountering complicated obstacles that are typical for investigations of complex diseases.

Even specifically with respect to the –omics of blood, which is much easier to obtain than measurements from other tissue samples, the following limitations arose. First, issues of ethics and animal welfare restrict blood draws from macaques, for example, to 10 ml/kg/month, or to 6 ml/kg/month if the animal is anemic. This regulation results in a spacing of measurement time points that obviously precludes the assessment of immediate metabolic host responses to the emergence of pathogens in the bloodstream. In fact, one is led to assume that metabolism, measured in this manner, is always in a steady state. It is permissible to obtain blood from the monkeys daily through standard procedures involving ear pricks, where no anesthesia is required, but it surprisingly turned out that blood from this source is metabolically quite different from venous blood [8].

Second, multi-omics approaches are often envisioned to include genomics, proteomics, metabolomics, and maybe other measurements from the same source at the same time. In our case, it is of course possible to subject blood samples to these different omics measurements. However, these measurements shed light on different blood components. RBCs, which are affected most directly, have no nuclei or mitochondria and therefore no DNA. Thus, “blood genomics” automatically and necessarily excludes about 99% of all blood cells, as it is restricted to white blood cells and parasites. By contrast, plasma proteomics is overshadowed by typical proteins like albumins, while membrane-proteomic measurements from infected RBCs are dominated