A Systems Biology Approach to Infectious Disease Research: Innovating the Pathogen-Host Research Paradigm

Alan Aderem,[a] Joshua N. Adkins,[b] Charles Ansong,[b] James Galagan,[c,d,e] Shari Kaiser,[a] Marcus J. Korth,[f] G. Lynn Law,[f] Jason G. McDermott,[b] Sean C. Proll,[f] Carrie Rosenberger,[a] Gary Schoolnik,[g] and Michael G. Katze[f,h]

Institute for Systems Biology, Seattle, Washington, USA[a]; Fundamental and Computational Sciences Division, Pacific Northwest National Laboratory, Richland, Washington, USA[b]; Department of Biomedical Engineering, Boston University, Boston, Massachusetts, USA[c]; Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA[d]; National Emerging Infectious Diseases Laboratory, Boston University, Boston, Massachusetts, USA[e]; Department of Microbiology, School of Medicine, University of Washington, Seattle, Washington, USA[f]; Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, California, USA[g]; and Washington National Primate Research Center, University of Washington, Seattle, Washington, USA[h]

ABSTRACT The twentieth century was marked by extraordinary advances in our understanding of microbes and infectious disease, but pandemics remain, food and waterborne illnesses are frequent, multidrug-resistant microbes are on the rise, and the needed drugs and vaccines have not been developed. The scientific approaches of the past—including the intense focus on individual genes and proteins typical of molecular biology—have not been sufficient to address these challenges. The first decade of the twenty-first century has seen remarkable innovations in technology and computational methods. These new tools provide nearly comprehensive views of complex biological systems and can provide a correspondingly deeper understanding of pathogen-host interactions. To take full advantage of these innovations, the National Institute of Allergy and Infectious Diseases recently initiated the Systems Biology Program for Infectious Disease Research. As participants of the Systems Biology Program, we think that the time is at hand to redefine the pathogen-host research paradigm.

Infectious diseases continue to be major health concerns worldwide: hepatitis C, AIDS, malaria, and tuberculosis (TB) are ongoing pandemics. A third of the world population is currently infected with the TB bacillus; millions of people are infected by influenza virus annually; and even though therapeutic drugs have slowed the threat of HIV-1, there is still no definitive cure or viable vaccine in sight. Although influenza virus vaccines provide protection, they must be reformulated annually, and the time lag from the spread of a pandemic virus to the availability of a vaccine is far too long. A universal vaccine against influenza virus is needed to limit high levels of morbidity and mortality. Newly emerging infectious diseases, such as swine H1N1 influenza, avian H5N1 influenza, severe acute respiratory syndrome (SARS), and dengue fever are a constant threat; and bacterial infections, such as the recurrent *Salmonella* outbreaks in the food industry, are costly to society.

New research strategies and tactics must be employed to minimize these threats and improve global health. In response to this need, the National Institute of Allergy and Infectious Diseases (NIAID) has sponsored the Systems Biology Program for Infectious Disease Research. This program consists of four centers, each with its own unique focus on a variety of viral and bacterial pathogens (Fig. 1). In this perspective, we discuss why the time is right for applying a systems biology approach to infectious disease research and how we are implementing this approach to overcome specific obstacles. This venture requires a new way of thinking about how best to address the challenges of infectious disease research.

DEFINING SYSTEMS BIOLOGY

Intrinsic to systems biology is the notion that biological systems have "emergent properties": that is, their sum is greater than their individual parts, and the biological outcomes of a system cannot be predicted by traditional reductionist methods that study only the individual components (1). Instead, an understanding of biological systems requires the integration of high-throughput multiomics data (transcriptomics, proteomics, metabolomics, lipidomics, etc.), which are used to construct predictive models of the networks and dynamic interactions between the biological components of the complex pathogen-host system (2). Through iterative rounds of model development, testing, and filling in the gaps with experimental data, models are refined to provide predictions that can ultimately be used to help identify therapeutic targets and improve clinical outcomes. This iterative cycle of perturbation biology is key to the systems biology approach (Fig. 2) and requires the establishment of efficient interdisciplinary collaborations, the building of integrated computational infrastructures, and the generation of self-consistent, comprehensive, and model-specific experimental data (3).

The technologies, computational methods, and genome information needed to successfully implement this approach now exist. Technologies such as next-generation sequencing have opened the door to sequencing total transcriptomes, encompassing microRNAs, long noncoding RNAs, and mRNAs. This has raised the possibility that a previously unappreciated class of RNAs—the long noncoding RNAs—may play important roles in the host response to virus infection (4). Proteomic technologies and accompanying informatics pipelines are evolving rapidly, with throughput and sensitivity approaching that of microarrays. Metabolomics, glycomics, lipidomics, and phosphoproteomics are young and comparatively undeveloped, but the promise is

Published 1 February 2011

Citation Aderem, A., J. N. Adkins, C. Ansong, J. Galagan, S. Kaiser, et al. 2011. A systems biology approach to infectious disease research: innovating the pathogen-host research paradigm. mBio 2(1):e00325-10. doi:10.1128/mBio.00325-10.

Address correspondence to M. G. Katze at honey@uw.edu.

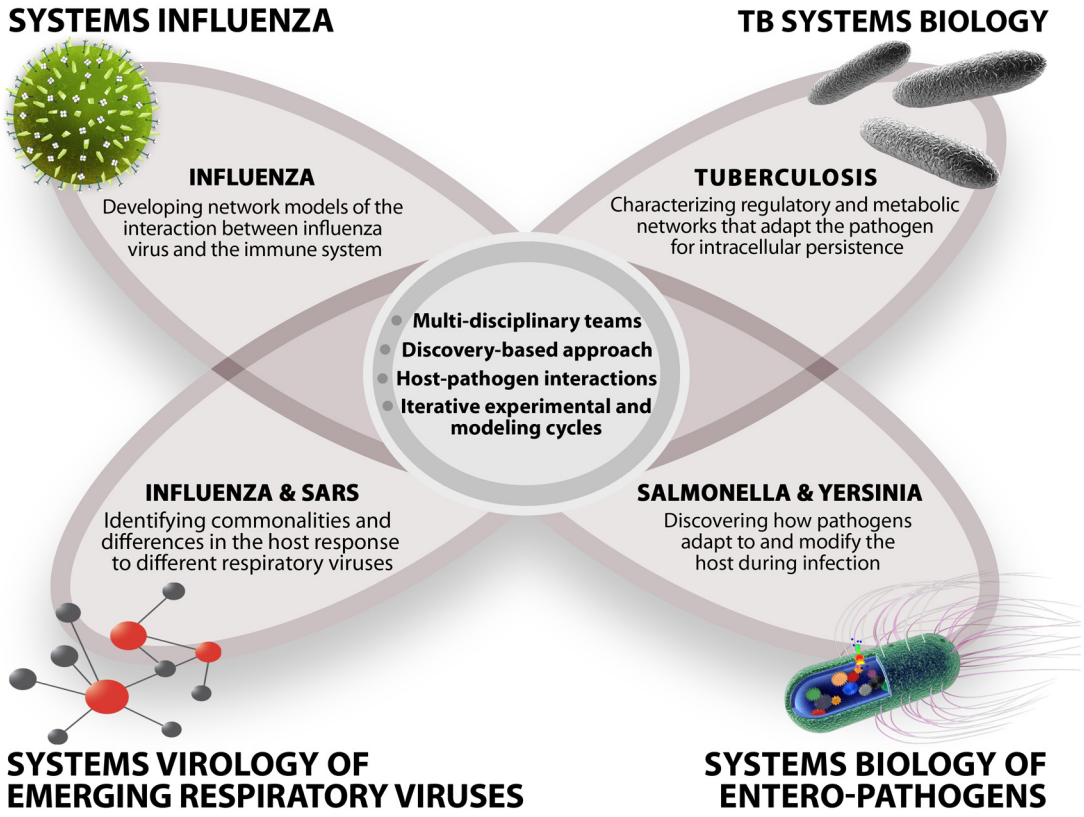

FIG 1 Common and distinct elements of the four NIAID Systems Biology Centers. The National Institute of Allergy and Infectious Diseases (NIAID) sponsors the Systems Biology Program for Infectious Disease Research. Each of the four centers focuses on unique aspects of the host-pathogen response while using several common approaches and techniques (center circle).

unlimited and the potential unprecedented. Moreover, the appropriate *in vitro* and *in vivo* experimental systems to generate useful pathogen-host interaction models are now available. Each of the four NIAID Systems Biology Centers has taken a unique approach to implementing this new research paradigm as outlined below.

NIAID SYSTEMS BIOLOGY CENTERS

The TB Systems Biology Center (http://www.broadinstitute.org/annotation/tbsysbio) at Stanford is using cell culture models, together with a combination of biochemical and transcriptional profiling methods, to characterize the state of the *Mycobacterium tuberculosis* bacillus and the host at each stage of its infectious cycle and during transitions between stages. Each of the three stages is orchestrated by a distinct genetic program that directs the expression of stage-specific antigens and the activation of stage-specific metabolic and biosynthetic pathways. There is an unusually close interaction between the pathogen and the host immune system, which depending on the outcome of this interaction, can lead to a stable, inactive, and asymptomatic infection that is compatible with a normal life span (latency) or to rapid destruction of lung tissue and death (reactivation disease). Chromatin immunoprecipitation followed by sequencing (ChIP-Seq) is one of the key technologies used to identify the regulatory networks that govern these transitions. Incorporation of these data with results from multidimensional assays that interrogate the status of DNA, RNA, proteins, lipids, and metabolites will lead to the construction of predictive models useful in the development of stage-specific therapeutics (8).

The Systems Virology Center (http://www.systemsvirology.org) at the University of Washington is comprehensively analyzing and modeling the molecular and cellular events, pathogen-host interactions, and cellular response networks induced or altered during the course of respiratory virus infection. This research focuses on highly pathogenic H5N1 avian influenza virus and severe acute respiratory syndrome-associated coronavirus (SARS-CoV). For each virus, the host response to highly pathogenic wild-type viruses and to engineered viruses with reduced levels of pathogenicity is being analyzed and modeled. These studies use cell culture infection systems, which are the most amenable systems for computation modeling, as well as mouse and macaque infection models. Animal models enable the incorporation of disease-relevant complexity and the validation, extension, and refinement of findings obtained from cell culture systems. Designing and performing experiments so that comparisons can be made across the model systems are an integral but labor-intensive part of the process. The direct communication between biologists and modelers—constant within all of the Systems Biology Centers—ensures that the experimental designs are suitable in terms of biological sample size, reproducibility, and consistency from high-throughput assays to construction of predictive models and the resulting biological interpretations. The question remains as to whether current computational and modeling techniques are robust enough to generate relevant models given the complexities of these systems. However, only by attempting these types of studies will the required methodologies be developed.

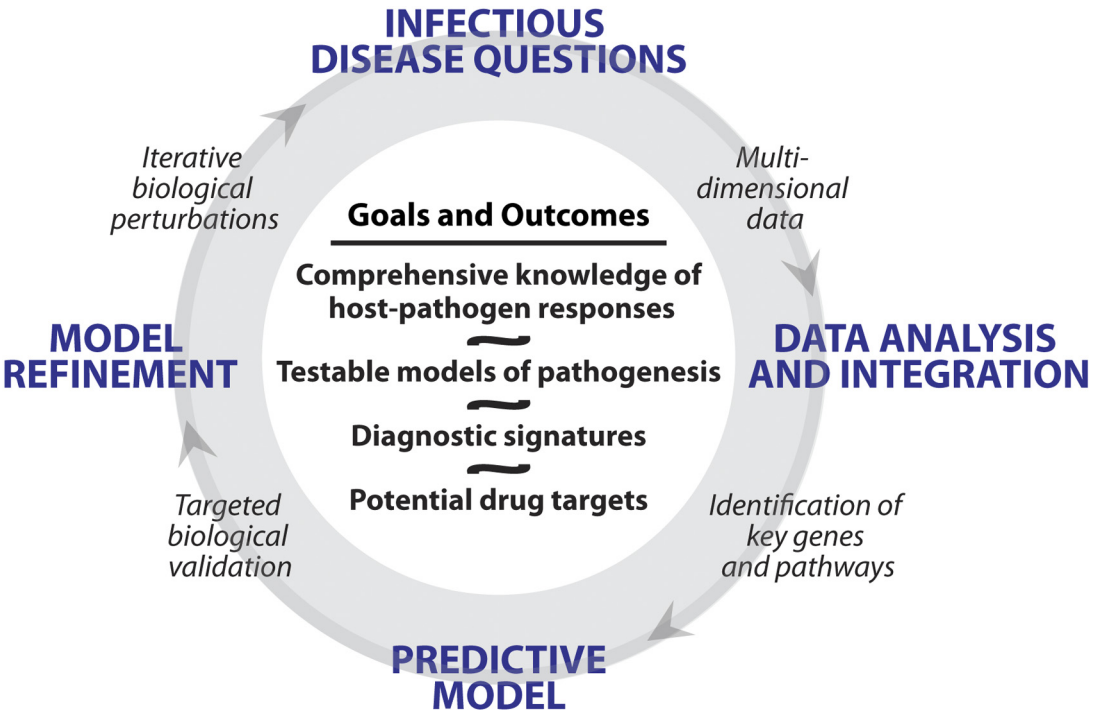

FIG 2 Iterative cycles of perturbation biology. The infectious disease questions, the first step in the cycle, determine the appropriate biological models and technologies utilized to generate multidimensional data. Data analysis and integration identify key components, pathways, and networks which allow for the construction of a predictive model. Model-predicted biological bottlenecks or key network nodes are validated by performing additional targeted experiments and data integration, resulting in a refined model. Importantly, several rounds of biological perturbations (i.e., use of mutant pathogens, cellular small interfering RNA [siRNA] knockdowns or knockout mice) are required to produce a predictive model that could be effectively utilized by the general infectious disease community. In addition to a more comprehensive understanding of the host-pathogen response and testable models, this type of perturbation biology will produce publicly disseminated multidimensional data sets and potentially both diagnostic signatures and drug targets.

The Center for Systems Influenza (http://www.systemsinfluenza.org) at the Institute for Systems Biology is using multi-omics approaches to compare the responses of both *in vitro* and *in vivo* systems to infection with influenza virus strains of various pathogenicities. The devastating mortality of the 1918 influenza pandemic was characterized by both a dysregulated innate immune response and a high susceptibility to secondary bacterial pneumonias. Computational integration of the transcriptomic, proteomic, and lipidomic data into an appropriate network model will illuminate the detailed signatures of extreme virulence and identify the molecular mechanisms underlying these responses. These studies are being extended to identify and model those host-virus interactions that predispose an infected lung to secondary bacterial invasion by *Staphylococcus aureus*. A comprehensive model of the influenza virus-host interaction, aided significantly by recent advances in sequencing technologies and advances in viral genetic manipulation, will serve to calibrate the global public health response to a newly emergent strain as well as uncover novel targets for therapeutic intervention.

The Center for Systems Biology for EnteroPathogens (http://www.sysbep.org/) at the Pacific Northwest National Laboratory is using regulatory and metabolic modeling, empowered by multi-omics data, to characterize the interactions between the host and distantly related bacterial pathogens *Salmonella enterica* serovar Typhimurium and *Yersinia* species. *Salmonella* is a leading cause of acute gastroenteritis worldwide, and *Yersinia pestis* is the causative agent of plague. As general models of infection, *Salmonella* and *Yersinia pestis* are among the best characterized with *in vitro* and *in vivo* animal models. These bacterial pathogens must use a well-orchestrated series of regulatory and metabolic changes to replicate and persist in the inhospitable host environment. Among these highly regulated processes is the process of secreting bacterial proteins into the host that modulate the host's response network. Early efforts have focused on developing methods to investigate and model the coordinated regulatory and metabolic changes occurring in the pathogen during infection and to identify and characterize the secreted bacterial proteins responsible for manipulating the host networks (5, 6). Understanding the regulatory and metabolic programs required for infection will enable the development of effective therapeutic treatments in the future.

WHAT ARE THE CHALLENGES AND HOW CAN SUCCESS BE MEASURED?

Criticism of systems biology includes the following: "it's too slow; it's too expensive; it's a fishing expedition." We readily admit that systems approaches are currently driven more by technology and computing than by a clear understanding of the data. Indeed, a major challenge is to thoroughly mine the existing data before accumulating even more information. However, critics must be cognizant of the immense challenges involved with this approach. For example, we need to create a new interdisciplinary culture along with a new vocabulary. We need to break down the traditional silos of information present in most scientific research endeavors. We need to merge the experimental with the clinical with

the computational and mathematical. We need to build bridges between diverse members of the scientific community who otherwise do not communicate with each other. The iterative cycle of experimental design, sample preparation, high-throughput assays, and data integration and analysis that is an absolute requirement for an effective systems approach is a lengthy, repetitious, and unglamorous process. But without this iterative progression, predictive and revealing models are not achievable.

We acknowledge that the best response to this criticism is to provide fundamental knowledge directly relevant to human health that could not otherwise be achieved through traditional scientific approaches: better cures and vaccines and better drugs and diagnostics. But just as systems biology differs in many aspects from more traditional basic research approaches, the methods for measuring success also differ. Although publications are one important measurement, an additional measure of success is the utilization of the generated resources by the scientific community, including experimental data sets, software tools, computational models, and research protocols, all made available to the scientific community through the four Centers' and NIAID websites (http://www.niaid.nih.gov/labsandresources/resources/sb/Pages/default.aspx). We call upon researchers outside the Centers to utilize the vast data sets and other resources being generated to further the knowledge and understanding of pathogen-host interactions.

As highlighted in a recent review, systems biology is already making important contributions to cancer research (7). The approach has been used to identify network motifs that can accurately predict the development of distant metastases in breast cancer patients—tumors that may develop years after the primary tumor and which frequently prove fatal. A network-centric approach has also identified specific oncogenes in certain types of B-cell lymphomas, and a multiscale mathematical model has been developed to test the efficacy of different radiation protocols, which has led to improvements in radiation therapies for colorectal cancer. The cancer field was one of the first research communities to aggressively embrace (and fund) the use of systems biology. We believe that the infectious disease community will build on the resulting substantial systems biology methodologies and insights to claim similar successes in the near future. We invite your thoughts and comments on this perspective.

ACKNOWLEDGMENTS

We thank Valentina Di Francesco for helpful comments and careful revisions of the paper and ultimately for providing the leadership of the Systems Biology Program, Peter Dudley and Irene Glowinski for editorial assistance, and Maria Giovanni for the general support of the program this paper originated from.

This project has been funded with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under contract no. HHSN272200800058C, HHSN27220080059C, and HHSN272200800060C and interagency agreement Y1-A1-8401.

REFERENCES

1. **Zak, D. E., and A. Aderem.** 2009. Systems biology of innate immunity. Immunol. Rev. **227**:264–282.
2. **Bumann, D.** 2009. System-level analysis of *Salmonella* metabolism during infection. Curr. Opin. Microbiol. **12**:559–567.
3. **Peng, X., E. Y. Chan, Y. Li, D. L. Diamond, M. J. Korth, and M. G. Katze.** 2009. Virus-host interactions: from systems biology to translational research. Curr. Opin. Microbiol. **12**:432–438.
4. **Peng, X., L. Gralinski, C. D. Armour, M. T. Ferris, M. J. Thomas, S. Proll, B. G. Bradel-Tretheway, M. J. Korth, J. C. Castle, M. C. Biery, H. K. Bouzek, D. R. Haynor, M. B. Frieman, M. Heise, C. K. Raymond, R. S. Baric, and M. G. Katze.** 2010. Unique signatures of long noncoding RNA expression in response to virus infection and altered innate immune signaling. mBio **1**(5):e00206-10.
5. **Yoon, H., J. E. McDermott, S. Porwollik, M. McClelland, and F. Heffron.** 2009. Coordinated regulation of virulence during systemic infection of *Salmonella enterica* serovar Typhimurium. PLoS Pathog. **5**:e1000306.
6. **Niemann, G. S., R. N. Brown, J. K. Gustin, A. Stufkens, A. S. Shaikh-Kidwai, J. Li, J. E. McDermott, H. M. Brewer, A. Schepmoes, R. D. Smith, J. N. Adkins, and F. Heffron.** 2011. Discovery of novel secreted virulence factors from *Salmonella enterica* serovar Typhimurium by proteomic analysis of culture supernatants. Infect. Immun. **79**:33–43.
7. **Laubenbacher, R., V. Hower, A. Jarrah, S. V. Torti, V. Shulaev, P. Mendes, F. M. Torti, and S. Akman.** 2009. A systems biology view of cancer. Biochim. Biophys. Acta **1796**:129–139.
8. **Galagan, J. E., P. Sisk, C. Stolte, B. Weiner, M. Koehrsen, F. Wymore, T. B. Reddy, J. D. Zucker, R. Engels, M. Gellesch, J. Hubble, H. Jin, L. Larson, M. Mao, M. Nitzberg, J. White, Z. K. Zachariah, G. Sherlock, B. C. Ball, and G. K. Schoolnik.** 20 May 2010. TB database 2010: overview and update. Tuberculosis (Edinburgh) **90**(4): 225–235. [Epub ahead of print.]