



Published in final edited form as:

Nat Rev Microbiol. 2013 July ; 11(7): 455–466. doi:10.1038/nrmicro3036.

SYSTEMS VIROLOGY: HOST-DIRECTED APPROACHES TO VIRAL PATHOGENESIS AND DRUG TARGETING

G. Lynn Law¹, Marcus J. Korth¹, Arndt G. Benecke^{1,2}, and Michael G. Katze¹

¹Department of Microbiology and Washington National Primate Research Center, University of Washington, Seattle, WA 98195, USA

²Université Pierre et Marie Curie, Centre National de la Recherche Scientifique UMR7224, Paris, France

Preface

High-throughput molecular profiling and computational biology are changing the face of virology, providing a new appreciation of the importance of the host in viral pathogenesis and offering unprecedented opportunities for better diagnostics, therapeutics and vaccines. Here, we provide a snapshot of the evolution of systems virology, from global gene expression profiling and signatures of disease outcome, to geometry-based computational methods that promise to yield novel therapeutic targets, personalized medicine and a deeper understanding of how viruses cause disease. To realize these goals, pipets and petri dishes need to join forces with the powers of mathematics and computational biology.

Introduction

Anyone who has taken an undergraduate virology course is familiar with subject matter focused on the structure of viral genomes and the molecular events associated with multi-step viral life cycles. The field of virology has done a remarkable job of characterizing and categorizing viruses and of defining the steps of viral attachment, entry, replication and release. Moreover, an understanding of viral protein function has paved the way for the development of antiviral drugs that target viral enzymatic activities. However, many of these drugs function poorly at best, and the virus-centric approach has not proven to be well suited for deciphering the complex and multifaceted virus-host interactions that underlie viral recognition, innate immune signalling and disease outcome. Within the past decade, tools have become available to chart a new course, one directed at obtaining comprehensive systems-level views of the host response and the interplay between virus and host.

Correspondence to: Michael G. Katze, Ph.D., Department of Microbiology, University of Washington, Box 358070, Seattle, WA 98195-8070, honey@u.washington.edu, Phone: 206-732-6136, Fax: 206-732-6056.

Further information

Katze laboratory: <http://viromics.washington.edu/>
Systems Virology Center: <https://www.systemsvirology.org>
Systems Influenza Center: <http://www.systemsinfluenza.org/>
Virus Pathogen Resource: <http://www.viprbc.org>
Influenza Research Database: <http://www.fludb.org>
Collaborative Cross: <http://www.csbio.unc.edu/CCstatus>

Systems virology is a phrase coined to describe the application of systems biology approaches to the field of virology¹. Systems biology is highly interdisciplinary in character, requiring the combined talents of biologists, mathematicians and computer scientists, and has as its goal a comprehensive understanding of biological systems. In the case of systems virology, these biological systems may range from virus-infected cells to tissues to whole organisms. Systems-level analyses use high-throughput technologies to measure system-wide changes in biological components such as DNA, RNA, proteins and metabolites, and are dependent on the quality of the resulting data sets (which are often noisy) and subsequent data integration and modelling. Ideally, high-throughput data derived from these and other measurements are integrated and analyzed using mathematical algorithms to generate predictive models of the system. Once a model has been developed, subsequent experimental perturbations of the system (for example, viral mutants or targeted inhibition of host genes or pathways) are used to yield refinements to the model and to increase its predictive capacity (FIG. 1)²⁻⁴.

This holistic host-directed approach stands in contrast to the more traditional reductionist approaches that focus on a pre-determined small set of molecules (genes, proteins or metabolites). Although often criticized for not being hypothesis-driven, systems-level (or discovery-based) analyses are instead increasingly being acknowledged as potent hypothesis generators. Moreover, for dynamical systems such as those involved in the host response to viral infection, systems-level analyses are considered that only way to understand emergent properties; that is, properties or biological outcomes that cannot be predicted by an understanding of the individual parts of a system alone, but rather only become apparent with knowledge of the specific organization and interactions between components⁵. Because of this, systems virology is an essential and synergistic complement to traditional virology approaches.

This Review focuses on the host response to virus infection and discusses the evolution and significant findings of systems virology, including the identification of gene expression signatures that are predictive of viral pathogenesis and vaccine efficacy, insights into how viruses disrupt cellular metabolism, and the mapping of virus-host interactomes. These accomplishments did not come from a single experiment or study, but rather from a body of work undertaken over several years by different investigators. The field has seen a progression from genomic-based approaches to measurements of proteins and metabolites and the embracing of host genetic variation as a means to better understand disease processes rather than as a source of frustration. Moving forward, systems virology must also embrace computational approaches capable of integrating this information to construct robust models of virus-host interactions that incorporate multiple dimensions and scales^{6, 7}. We cite examples of studies that are moving in this direction and outline what the next phase of systems virology must encompass to reach its full potential.

Gene expression signatures

With the completion of the human genome project and the advent of microarrays capable of measuring RNA transcripts at a genome-wide scale, the first systems-level analyses became a reality. Over the past 12 years, DNA microarrays have evolved from hundreds of cDNAs

spotted on nylon membranes to glass slides containing high-density oligonucleotides that encompass entire genomes (BOX 1). Microarrays (then covering only 1,500 human genes) were first used in virology to evaluate the changes in cellular gene expression that occurred in a CD4+ T cell line infected with HIV⁸. Since then, the use of microarrays to evaluate changes in host gene expression in response to virus infection has become commonplace. In many cases, these studies remain small and narrowly focused, and although they provide glimpses into global responses, low sample numbers make it difficult to determine reproducibility. In addition, they do not provide the robust data sets that are needed for computational modelling that could result in deep insights into system architecture or behaviour. This shortcoming has perhaps fed skepticism regarding the ability of genome-wide expression profiling to yield transformative discoveries. Below, we provide examples of more comprehensive studies resulting in genomic signatures that have increased our understanding of viral pathogenesis and the characteristics of the host response required for immune protection.

Signatures of highly pathogenic respiratory viruses

Influenza virus is well known for its ability to rapidly evolve new variants through genetic mutation and genome reassortment, yielding strains that can vary widely in virulence and transmissibility. Most strains cause relatively mild respiratory disease, whereas others, such as the 1918 pandemic virus and highly pathogenic avian H5N1 strains such as A/VN/1203/04, can cause severe and often fatal infections⁹. Over the past 8 years, the field has used DNA microarrays and functional analyses to define the virus-host interactions that regulate influenza virus pathogenesis¹⁰. These studies have identified gene expression signatures that correlate with viral virulence and have revealed that the timing and magnitude of the host response is a critical determinant of eventual outcome of infection.

This phenomenon was first demonstrated in studies that used a combination of mouse and macaque infection models and genome-wide transcriptional profiling to measure the host response to the reconstructed 1918 pandemic virus. In these animal models, the virus causes a rapidly fatal infection marked by severe lung pathology, intense neutrophil infiltration, and the rapid and sustained induction of pro-inflammatory cytokine and chemokine genes^{11, 12}, an event often referred to as a cytokine storm¹³. By contrast, macaques infected with a highly pathogenic avian H5N1 strain show a rapid and intense induction of interferon and innate immune genes that eventually resolves as the animals recover¹⁴. Genomic analyses have also revealed that macaques infected with the 1918 pandemic strain¹⁵ and mice infected with A/VN/1203/04 H5N1 virus show a strong induction of genes encoding inflammasome components (for example, NLRP3 and IL-1 β)¹⁶. This H5N1 virus is particularly virulent in mice, and although the inflammasome is part of the innate immune response to influenza A viruses¹⁷⁻¹⁹, the excessive activation of this response seems to be detrimental in this species.

As high-throughput data have accumulated in public databases, it has become possible to use this information to carry out meta-analyses. This strategy has been used to analyze data from a compendium of published studies that used mouse models to measure host transcriptional responses to lethal or non-lethal strains of influenza virus, respiratory

syncytial virus or SARS coronavirus²⁰. Two alternative methods were used to generate gene expression signatures that are predictive of high or low pathogenicity (defined here as 100% mortality or survival, respectively). The first signature consists of 74 genes the expression of which changes in the opposite direction with respect to mock infections (referred to as a 'digital' relationship). The second signature consists of 57 genes that are differentially expressed between high- and low-pathogenic infections, without reference to mock infections (referred to as an 'analogue' relationship). Most genes in the analogue signature are differentially expressed during both lethal and non-lethal infections, but high pathogenicity corresponds with a higher degree of differential expression.

When the two signatures were tested for their ability to predict pathogenicity, the best predictor of a highly pathogenic infection was the analogue signature (FIG. 2). Significantly, this meta-analysis did not take time after infection into account; that is, data from samples isolated at 1 or 5 days post-infection were treated equally. However, the majority of samples that were correctly identified as being from either a low or high pathogenic infection were from early or late time points, respectively. It is therefore likely that taking time post-infection into account would yield an even more accurate signature.

From these studies, it is now apparent that highly pathogenic respiratory viruses induce or suppress the expression of many of the same genes as mildly pathogenic viruses, but to a greater degree (and with different kinetics). Therefore, knowing the identity of genes that are differentially expressed in response to infection only provides part of the information needed to predict pathogenicity. The magnitude and timing of the host response are critical determinants of eventual disease outcome, and this may have important implications for antiviral therapy. To date, efforts to target the host response with a variety of anti-inflammatory drugs have been largely unsuccessful¹³, and it is likely that effective host-directed therapy will depend not only on the target, but on the timing at which elements of the host response are suppressed or enhanced. Moreover, these findings point to the need for computational approaches that can describe nonlinear relationships (see below) and account for various factors associated with large multivariate datasets²¹. This also suggests the need for a paradigm shift in biomarker discovery to one that looks at sets of quantitative molecular measurements.

Signatures of vaccine efficacy

The application of systems-level analyses to vaccine research, variously termed systems vaccinology²² or vaccinomics²³, has led to the identification of molecular signatures predictive of vaccine immunogenicity and to new insights into the mechanisms of action of vaccines. In one of the first large-scale uses of this strategy, gene expression profiling and computational methods were used to identify gene expression signatures predictive of the strength of the adaptive immune response in humans vaccinated with the yellow fever vaccine, YF-17D²⁴.

Transcriptional profiling of peripheral blood mononuclear cells from vaccinated subjects revealed that YF-17D induces the expression of genes encoding proteins that are associated with viral recognition and transcription factors that regulate type I interferons. Although also characteristic of the transcriptional response to active virus infection, this response did not

correlate with subsequent CD8⁺ T cell or neutralizing antibody responses, which are thought to mediate protection. However, an alternative computational and classification method, discriminant analysis via mixed integer programming (DAMIP), identified a signature (consisting of complement system and stress response genes) that is highly accurate in predicting subsequent CD8⁺ T cell activation. This method also identified a separate signature (which includes TNF receptor superfamily members) that accurately predicts neutralizing antibody expression during infection. The robustness of these signatures was verified through the analysis of samples from a different group of subjects vaccinated with a different lot of vaccine, thereby identifying new correlates of vaccine immunogenicity.

More recently, similar systems approaches have been used to evaluate innate and adaptive immune responses to vaccination against influenza virus, with the goal of identifying early gene expression signatures that correlate with immunogenicity²⁵. Over a 3-year period, a series of clinical studies was undertaken in which young adults were vaccinated with either inactivated influenza vaccine (TIV) or live-attenuated influenza vaccine (LAIV). Molecular signatures for predicting antibody responses were identified by combining gene expression profiling, antibody response data, real-time PCR analysis and DAMIP. The resulting predictive signature consisted of genes with known roles in antibody response and other genes with previously unidentified roles in antibody or B cell responses. For example, one gene from the predictive signature, *CAMK4*, encodes CaMKIV kinase, a protein known at the time to be involved in multiple immune system processes. However, it was not known if it had a role in antibody responses²⁵. To demonstrate the ability of the systems biology approaches used in this study to identify biologically significant targets, *CAMK4*-knockout mice were vaccinated with TIV and exhibited significantly higher antibody titers on days 7, 14, and 28 days after vaccination compared with wild-type mice, thus confirming the hypothesis that *CAMK4* is important in regulating B cell responses. Although further investigations are needed to confirm many of these signature predictions, these studies demonstrate that systems approaches can both identify biological targets and generate new testable hypotheses related to the mechanism of vaccine action.

An expanding view of the transcriptome

Until recently, transcriptional profiling depended on the use of microarrays to measure the expression of well-annotated protein-coding genes. The advent of next-generation sequencing (BOX 2), however, has brought the ability to rapidly sequence the entire RNA complement of cells or tissues. This has led to a much expanded concept of the host transcriptome, as most recently revealed by the Encyclopedia of DNA Elements (ENCODE) project²⁶. It is now apparent that as much as three-quarters of the human genome is capable of being transcribed and that cells contain vast numbers and varieties of non-protein-coding RNAs²⁷. Some of these noncoding RNAs, such as microRNAs, have been well studied and are known to have roles in virus infection²⁸. For most others, functionality is less clear, but there is growing evidence that long noncoding RNAs also play parts in transcriptional and epigenetic gene regulation and disease²⁹.

RNA-seq analysis of the host response to SARS coronavirus infection has revealed the differential expression of a variety of host long (greater than 200 nucleotides) noncoding

RNAs in lung samples from virus-infected mice³⁰. Many of these RNAs have similar expression patterns *in vitro* during influenza virus infection or type I interferon treatment, suggesting that they may be involved in regulating the innate immune response to a variety of viruses³⁰. Expanding these analyses to include the sequencing of small RNAs also revealed the differential expression of over 200 small RNAs, such as small nuclear RNAs (snoRNAs) and piwi-associated small RNAs (piRNA), in response to SARS coronavirus or influenza virus infection³¹. Similarly, RNA-seq has revealed that an HIV-infected CD4+ T cell line exhibits the differential expression of host microRNAs, snoRNAs and pseudo genes compared with uninfected cells³². Viral mRNA constitutes a surprisingly large portion of the total RNA in HIV-infected CD4+ T cells in this study (nearly 40% by 24 hours after infection), and reads mapping to the viral genome have revealed novel viral RNA splice variants. A correlative analysis that combined mRNA-seq and small RNA-seq data showed additional roles for host microRNAs in T cell activation and transcriptional and cell cycle regulation during HIV infection³³.

Together, these studies attest to the power of RNA-seq to provide entirely new views of the transcriptional landscape and the previously unanticipated changes in transcription that occur in response to virus infection. Such insights are not limited to host transcription, as a combination of RNA-seq and mass spectrometry recently revealed that human cytomegalovirus (a 240-kb DNA virus) produces hundreds of previously unidentified transcripts and short proteins that may have functional, regulatory or antigenic properties³⁴. Although the functional significance of changes in noncoding RNA expression are only beginning to be examined²⁹, a better grasp of noncoding RNA expression and function will certainly be necessary for a complete understanding of viral pathogenesis, innate and adaptive immune responses, and a more general conception of gene regulation. Unfortunately, despite the advancements coming out of RNA-seq, the extensive computing infrastructure needed to handle large data files and the computational prowess required to align, assemble and analyze short sequence reads continues to put the approach out of reach for most laboratories.

Beyond the transcriptome

Of course, gene expression profiling provides only one measure of the host response to infection. In recent years, advances in systems-wide technologies have facilitated a ‘multi-omics’ approach that includes proteomics, metabolomics, lipidomics and virus-host protein interactomes. All of these measurements are adding to our understanding of the host response to viral infection, and new abilities to evaluate the role of host genetics and epigenetics are adding additional layers of complexity.

Alterations in cellular metabolism

Viruses have long been known to cause changes in host metabolism; however, the full extent of such changes was not clear until systems approaches were used to evaluate the metabolomic reprogramming that results from human cytomegalovirus (HCMV) infection³⁵. Using liquid chromatography-mass spectrometry to directly measure the levels of over 160 different metabolites, it was discovered that HCMV infection induces large increases in numerous metabolites, including glycolytic and TCA cycle intermediates, amino acids,

NADH and pyrimidine. The metabolic signature induced by HCMV infection is readily distinguishable from the signature associated with the transition of quiescent cells into the G1 phase of the cell cycle, revealing the replacement of cellular metabolic homeostasis with a HCMV-specific metabolic programme. Systems-level metabolic flux profiling [G] produced a first-of-a-kind metabolic map showing linkages between compounds and quantitative information about metabolic activity. The map is biochemically revealing, indicating a global upregulation of metabolism by HCMV, with the greatest increase in the TCA cycle and its efflux to feed fatty acid biosynthesis³⁶.

To determine whether this reprogramming of the host metabolome is cell type- or virus-specific, fibroblast and epithelial cells were used to compare the host response to two strains of HCMV and two strains of herpes simplex virus (HSV-1)³⁷. All four viruses produced significant changes in approximately 50% of the metabolome. Interestingly, the changes are consistent across different strains of the same virus and across cell types, but differ markedly between HCMV and HSV-1, demonstrating that these viruses induce distinct metabolic programmes. The findings derived from these systems-level approaches generated the hypotheses needed to drive additional focused studies in which more traditional methods were used to investigate the molecular mechanisms underlying the virus-specific hijacking of the host metabolome and the relevance of these mechanisms to potential therapeutic interventions³⁸⁻⁴².

Genomic and lipidomic analyses have also revealed that infection of primary bone marrow-derived macrophages with mouse cytomegalovirus results in a downregulation of metabolites involved in the cholesterol metabolic pathway⁴³. The lowering of cholesterol levels is mediated through the interferon-dependent downregulation of SREBP2, a transcription factor that regulates sterol biosynthesis. Pharmacologic or RNAi-mediated inhibition of the sterol pathway was shown to result in increased protection against virus infection in cell culture and in mice, demonstrating the potential benefit of targeting a host metabolic pathway as an antiviral strategy.

Virus-host interactomes

Based on the premise that viral proteins interact with cellular factors to promote efficient viral replication and pathogenesis, system-wide siRNA or shRNA [screens, yeast two-hybrid libraries and bioinformatic methods are being used to construct and describe virus-host interactomes and in turn identify cellular targets for therapeutic intervention. Such interactomes have been generated for numerous viruses, including influenza virus, HIV, dengue virus, hepatitis C virus (HCV), herpes viruses and SARS coronavirus, and have yielded 'hitlists' of cellular factors that may be important in viral pathogenesis⁴⁴⁻⁵¹. Although the number of interactions identified by these studies is impressive, few of the genes identified have been subjected to functional analyses to confirm their role in virus replication. Moreover, there is very little overlap in the host factors identified by the different screens. This could be due to a variety of factors, including differences in screening systems, cell types and viruses, as well as the methods used to identify interacting partners. The benefits and shortcomings of these studies, as well as factors affecting their outcome, have been the subject of several detailed reviews⁵²⁻⁵⁵.

To fully realize the collective information residing in this vast collection of data, it will be necessary to develop computational and mathematical methods capable of fully integrating interactomes that have been constructed using different methods or have been obtained from different biological systems, as well as their associated metadata. In a first attempt at such integration, meta-analysis of virus-host interactome data for five distinct viruses identified both common and virus-specific human protein targets⁵⁶. Common targets include proteins involved in the cell cycle, apoptosis, the unfolded protein response and nuclear transport. Many of the common host targets identified are multifunctional hubs – that is, they have multiple functions or roles within the cell. When coupled with the fact that many viral proteins interact with multiple host proteins, this combination of factors may explain how viruses, with their relatively small genomes, are capable of dysregulating so many aspects of host biology. Unfortunately, because these common host targets often have numerous cellular functions, focusing on them as drug targets may be problematic⁵².

In an interesting twist on these studies, virus-host interactome data, together with data on host transcriptional changes resulting from the expression of 123 viral open-reading frames derived from DNA tumour viruses, was used to predict cellular genomic variations (for example, mutations, deletions or translocations) that can give rise to cancer⁵⁷. By defining the rewiring of cellular networks and pathways caused by the viral proteins, and identifying a list of host proteins central to the rewiring, the systematic identification of host targets of DNA tumour viruses was found to be as successful as traditional large-scale cataloging of tumour mutations for cancer gene identification. This suggests that disease phenotypes, whether resulting from virus infection or cancer, may be the result of network perturbations rather than individual genetic or genomic variations.

Host genetics and epigenetics

Host genetic variation has typically been thought of as a confounding factor that limits the ability to draw conclusions from data obtained using out bred (for example, human and nonhuman primate) populations. More recently, attempts are being made to better understand how genetic diversity influences infection outcome and how knowledge of genetic diversity can be incorporated into the construction of robust and predictive network models. One particularly exciting example is the Collaborative Cross (CC) mouse resource. The CC is a unique panel of multiparental recombinant inbred mouse strains designed to capture the level of genetic diversity found in out bred populations and provides a resource for systematically identifying individual and multiple host genetic traits that contribute to complex immune phenotypes and disease outcome (BOX 3).

As an example, a genetically diverse panel of pre-Collaborative Cross mice (not fully inbred) determined to have severe or mild responses to influenza virus infection, have been used to identify expression quantitative trait loci (eQTL) associated with the host response to infection. Twenty-one high-confidence eQTL were identified, 17 of which were confirmed using mice from the eight Collaborative Cross founder strains⁵⁸. Many of these genes have known functions related to immunity or the host response to infection, such as *IFi27l2a*, *Clec16a*, *Pde7a* and *Tcf7l1*, whereas for others such as *Sik1* and *Senp5*, their role in influenza infection is not yet clear. Structural equation modelling [G] was used to identify

potential regulatory relationships between additional genes and the validated eQTL, suggesting that these genes and corresponding sub-networks play important roles in the host response by either promoting a protective (mild response) or a pathologic (severe response) to influenza infection depending on the specific genetics of the host. Thus using this genetically diverse population, high-confidence gene candidates involved in regulating the host response to influenza infection were identified, allowing for future investigations into their utility as therapeutic targets.

It is also becoming apparent that epigenetic mechanisms have a role in regulating the outcome of virus infection, and methods for genome-scale mapping of DNA methylation⁵⁹ and histone modification⁶⁰ are now available. Such epigenetic modifications contribute to chromatin structure and organization, which in turn influence transcriptional activity, the immune response⁶¹ and viral latency⁶². Viruses may also use epigenetic control mechanisms to their advantage; the NS1 protein of H3N2 influenza virus, for example, acts as a histone mimic to suppress the expression of antiviral genes⁶³. Characterizing and understanding epigenetic mechanisms might therefore be an essential requirement for the construction of gene regulatory networks^{64, 65}.

Putting the pieces together

As more and more high-throughput data become available, systems virology is poised to enter a new phase to fulfill its initial promise of revolutionizing our understanding of virus-host interactions. To do this, the field must move beyond just the listing of molecules that are differentially expressed upon viral infection. Instead, the relationships between key molecules must be defined. Such relationships may be cause-and-effect relationships (for example, transcription factors and their target genes), the result of co-expression, or due to genetic or direct physical interactions. Here we give examples of several methods being used to further our understanding of viral-host interaction networks and discuss key computational challenges that must be addressed.

Network modelling and analysis explores relationships among molecules and analyzes the structure and organization of the relationships to predict the behaviour of the network or system. For example, the context likelihood of relatedness (CLR) method is used to predict genes that are highly interconnected (referred to as hubs) or that exhibit a high degree of betweenness centrality (referred to as bottlenecks). Genes with high betweenness centrality exhibit fewer connections than hub genes, but because they are located between (and connect or bridge) multiple subnetworks, they can play a powerful role in controlling network signaling (FIG. 3). Bottleneck genes often function as key genes in the regulation of disease progression and are therefore attractive targets for further experimentation^{66, 67}. An alternative method, co-regulation network analysis (PCluster)⁶⁸, has been combined with genome-wide expression profiling and yeast two-hybrid analysis to identify relationships between gene expression and direct physical interactions revealing previously unrecognized roles for several cellular and viral proteins in the host response to H1N1 influenza virus⁶⁹. These proteins include a network of RNA-binding proteins, components of the WNT signalling pathway and viral polymerase subunits. However, as these types of analyses often

only infer correlations between network components, additional studies are required to verify model predictions.

These network analysis methods have also been used to analyze the topology [G] of networks derived from proteomic and lipidomic profiling data, resulting in the identification of two mitochondrial fatty acid oxidation enzymes, DCI and HADHB, as bottleneck genes and possible targets through which HCV disrupts cellular metabolic homeostasis⁷⁰. The importance of DCI (and of cellular metabolic homeostasis in general) during HCV infection was then confirmed by additional studies, which included pharmacologic inhibition of fatty acid oxidation and targeted siRNA knockdown techniques, both of which demonstrated that DCI is required for productive HCV infection in hepatoma cell lines^{71, 72}. Similarly, analysis of interaction networks between HCV Core, NS4B and host proteins has been used to identify potential host anti-HCV therapeutic targets, including alpha enolase, paxillin and a solute carrier protein (SLC25A5)⁷³. An understanding network topology provides the opportunity to identify potential targets for therapeutic intervention as well as insights into possible off-target effects on network signalling that may be induced by drug treatment.

Although much can be learned from the construction and topological analysis of host-pathogen interaction network utilizing samples from whole tissues, such networks provide a generalized picture of the changes that occur in the host during the course of infection owing to the heterogeneity of cell types present in most tissues. For example, it is difficult to delineate from these types of networks the signalling events that may occur between infected lung epithelial cells and cells of the immune system, both in and outside of the infected tissue. These intercellular interactions are also controlled by signal transduction pathways, which communicate signals from the extracellular environment to intracellular effector processes. A much better understanding is needed of intercellular signaling processes, the cells that are involved, and the directionality of their effects on infection outcome.

A few studies have begun to explore cell-type specific and intercellular signaling on a system-wide scale. For example, flow cytometry and gene expression data from bronchoalveolar lavage (BAL) fluid from young-adult and aged macaques infected with 2009 pandemic H1N1 influenza virus were analyzed in conjunction with data from the Immune Response in silico (IRIS) database. This database contains cell-type specific gene expression patterns associated with various types of immune cells. By computationally comparing differentially expressed genes in BAL with cell-type specific gene expression patterns in the IRIS database, it was possible to identify genes associated with specific immune cell types including activated dendritic cells, CD4 and CD8 T cells and naïve B cells. In particular genes associated with T and B cell markers were more highly upregulated in young adult animals⁷⁴. Recent studies of mouse models of intestinal inflammation induced by TNF treatment provide a good example of how systems approaches can be used to evaluate signalling between cell types in a complex tissue environment^{75, 76}. Phosphoprotein, cytokine, chemokine and flow cytometry measurements from various immune cell types over time and under diverse conditions were combined to construct statistically robust multivariate regression models that related the phosphoprotein signals, cytokines and cell types to specific phenotypes. These models helped to elucidate key molecular and cellular processes governing epithelial cell apoptosis and proliferation in

response to TNF treatment. For example, monocyte chemotactic protein 1 (MCP1) was predicted by the model to be especially protective against TNF-induced apoptosis, which was confirmed by treating with anti-MCP1 antibody prior to TNF administration. The model also indicated that plasmacytoid dendritic cells might be particularly important, a prediction that was confirmed by depleting these cells from mice under conditions that produced the most severe epithelial cell apoptosis; this reverted the TNF-induced phenotype to the mildest outcome.

In addition to constructing network models that span intra- and intercellular signalling processes, it will also be necessary to consider nonlinear relationships such as how the network functions over time (i.e., the dynamics of the system). This is particularly true in light of evidence (discussed above) that the magnitude and timing of the host response to respiratory viruses are critical determinants of eventual disease outcome. Similar evidence is accumulating that infectious outcome of HIV is also related to the activation dynamics of host gene regulation^{77, 78}. However, high-throughput data are typically static and often not adequate for modelling dynamical systems. To help overcome this limitation in analyzing network dynamics, inference methodologies are being devised to reinterpret activity differences due to system perturbation as differences in observation time. For example, changes in pathogen-induced gene expression that are associated with genetic variability, both on the side of the pathogen (e.g., mutant viruses) and the host (as occurs in the Collaborative Cross mouse model), can principally be used to indirectly infer the critical dynamics of a system without having to measure the system over time. This innovative approach has only recently become feasible with major breakthroughs in the theory of dynamical systems and geometric high-dimensional analysis methodology^{79, 80}.

As virology continues to transition into a more quantitative science, increasing attention must be paid not only to network dynamics but to other nonlinear interactions, such as cooperative or synergistic relationships, which characterize so much of biology. Current biomarker discovery and the identification of molecular predictors of efficacy of adjuvants, vaccines or more generally drugs are largely unsuccessful because nonlinear interactions between molecules, as well as genetic diversity in populations, are not taken into sufficient consideration. Geometric methods (i.e., methods such as principal component analysis that are used to identify structure in data by identifying spatial and temporal relationships) are increasingly being used in the analysis of high-throughput molecular data. In particular, novel combinations of geometric methods, such as those based on singular value decomposition (SVD) and multidimensional scaling (MDS)^{81, 82} are beginning to be used in systems virology to better understand nonlinear interactions between variables and isolate those from biological noise.

SVD-MDS analysis of transcriptomic data derived from liver biopsies obtained from HCV-infected liver transplant patients, in combination with categorical analysis (to take into account variables such as age, time post-transplant, fibrosis score, etc.), has been used to identify a molecular signature for patients at risk of developing severe fibrosis⁸³. SVD-MDS and co-abundance networks (which relate molecules on the basis of their abundance profiles) were also used to integrate proteomic and metabolomic data sets obtained from the same cohort of patients. This strategy identified a potential role for oxidative stress in rapid

fibrosis progression post-transplant and identified serum metabolites that may prove useful as biomarkers for predicting progression to fibrosis⁸⁴. This understanding of network structure can now be used to simulate human liver metabolism using novel flux-balance modelling approaches to better comprehend and eventually treat disease⁸⁵. Additional geometric approaches have been described and should prove useful for effectively bridging different technologies^{86, 87} and for integrating diverse types of data, thereby better leveraging data already available in public databases and repositories. Finally, geometric methods, and links between geometry, information theory and probability theory⁷, will also help to identify causal relationships⁸⁸, which is clearly a currently unmet challenge. Unlike purely statistical approaches, geometric methods can be used to integrate different individual measures for the purposes of comparison and combination into coherent objects that identify relationships between genes, transcripts or proteins.

Conclusions

After having been extensively hyped as a paradigm shift, system-level approaches have been criticized for failing to rapidly fulfill their initial grand promises. Standing in the way have been numerous technical, experimental and mathematical hurdles. However, as discussed in this Review, significant progress is being made, and new computational approaches are leading the way. Also of importance is the ever-growing availability of high-throughput data in public databases; indeed, data sharing is crucial to the future success of systems virology (BOX 4). To date, bench work has been the necessary precedent to computational approaches, but we are now at a point where sufficient data are available that computational methods can be the starting point for making discoveries, generating hypotheses and in turn guiding targeted bench work. A prime example of this is the flourishing of virtual screening methods in drug discovery, which depend on systematic drug characterization efforts and public databases holding functional genomics information obtained under standard experimental conditions^{89, 90}. Moreover, the information gained from systems approaches form the basis for what has been termed P4 medicine: personalized, predictive, preventive and participatory⁹¹. In the case of infectious disease, genetic information on the individual and the pathogen, disease-predictive molecular signatures, targeted risk reduction and prophylactic measures, and active patient participation will merge into a new approach to medical care⁹².

In conclusion, contemporary virology cannot afford to simply catalogue myriad circumstantial observations. Systems-level approaches provide the opportunity to assemble the incomplete puzzle of biology in a meaningful way that will advance our understanding of how viruses cause disease and lead to improved patient care. Although there will always be a need for traditional microbiology, the success of today's undergraduates will depend on their ability to combine traditional skills with systems approaches and mathematics. The time is ripe, the data are here, and the mathematics to put them together is coming along. Take heed of the words on Plato's doorstep: "ἀγεωμετρητος μηδεις εἰσιπῶ"^a.

^aLet no one ignorant of geometry enter.

Acknowledgments

We thank Laurence Josset for generating Figure 3 and Mark Heise and Martin Ferris for providing the data used in Box 3. Research in the author's laboratory is supported by Public Health Service grants R2400011172, R2400011157, P30DA015625, P51RR00166, and U54AI081680 and by federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under contract HHSN272200800060C.

Glossary terms [G]

Betweenness centrality	Centrality is a measure of the location of a gene in a network. Genes with high betweenness centrality, referred to as bottleneck genes, are located between and connect different portions of the network.
CLIP-Seq	Cross-linking immunoprecipitation-high-throughput sequencing, a screening method used for identifying RNA sequences that interact with either RNA-binding proteins or other RNA.
Epistasis	The phenomenon where the effects of one gene are modified by one or more other genes.
Metabolic flux profiling	A measurement approach that uses liquid chromatography-tandem mass spectrometry to quantify the rate of conversion of biochemical molecules in a metabolic network after perturbing the system. Systems-levels metabolic flux profiling is a high-throughput approach to quantifying changes in metabolic activity.
Network topology	The arrangement and connections of the various components of a network.
RIP-Seq	Immunoprecipitation of RNA-binding proteins followed by high-throughput sequencing of the bound RNA.
Structural equation modeling	A multivariate analysis technique for testing and estimating causal relationships among variables.
snoRNA	Small nucleolar RNAs that guide the modification (e.g., methylation or pseudouridylation) of other RNAs, particularly ribosomal RNAs.
piRNA	Piwi-interacting RNAs. These small RNAs are thought to be involved in gene silencing through the formation of ribonucleoprotein complexes with Piwi proteins.
Unfolded protein response	A cellular stress response to the accumulation of unfolded proteins in the endoplasmic reticulum. The response is characterized by a signal transduction pathway designed to restore homeostasis by limiting protein biosynthesis and increasing the abundance of molecular chaperones involved in protein folding.
Expression quantitative trait loci (eQTL)	Genomic loci that regulate mRNA expression. eQTL are mapped by computationally connecting DNA sequence variation with variation in

gene expression, providing information on how host genetics impacts the function of molecular networks.

Complement system

Proteins found in the blood that react with one another and aid the ability of phagocytic cells to eliminate microbes. Complement proteins also play a role in the development of inflammation.

References

1. Aderem A, et al. A systems biology approach to infectious disease research: innovating the pathogen-host research paradigm. *mBio*. 2011; 2:e00325-10. [PubMed: 21285433]
2. Tisoncik JR, Katze MG. What is systems biology? *Future Microbiol*. 2010; 5:139–141. [PubMed: 20143938]
3. Katze, MG., editor. *Systems biology*. Heidelberg: Springer; 2013.
4. Macilwain C. *Systems Biology: Evolving into the Mainstream*. *Cell*. 2011; 144:839–841. [PubMed: 21414474]
5. Diercks A, Aderem A. Systems approaches to dissecting immunity. *Curr Top Microbiol Immunol*. 2012 In press.
6. Lauffenburger DA. The multiple dimensions of Integrative Biology. *Integr Biol (Camb)*. 2012; 4:9. [PubMed: 22158999]
7. Lesne A. Multiscale analysis of biological systems. IHES pre-print - 2011-04-17. 2012
8. Geiss GK, et al. Large-scale monitoring of host cell gene expression during HIV-1 infection using cDNA microarrays. *Virology*. 2000; 266:8–16. [PubMed: 10612655]
9. Fukuyama S, Kawaoka Y. The pathogenesis of influenza virus infections: the contributions of virus and host factors. *Current Opinion in Immunology*. 2011; 23:481–486. [PubMed: 21840185]
10. Korth MJTN, Benecke AG, Katze MG. Systems approaches to influenza-virus host interactions and the pathogenesis of highly virulent and pandemic viruses. *Seminars in Immunology*. 2012 **in the press**.
11. Kash JC, et al. Genomic analysis of increased host immune and cell death responses induced by 1918 influenza virus. *Nature*. 2006; 443:578–581. [PubMed: 17006449]
12. Kobasa D, et al. Aberrant innate immune response in lethal infection of macaques with the 1918 influenza virus. *Nature*. 2007; 445:319–323. [PubMed: 17230189]
13. Tisoncik JR, et al. Into the eye of the cytokine storm. *Microbiol Molec Biol Rev*. 2012; 76:16–32. [PubMed: 22390970]
14. Baskin CR, et al. Early and sustained innate immune response defines pathology and death in nonhuman primates infected by highly pathogenic influenza virus. *Proc Natl Acad Sci USA*. 2009; 106:3455–3460. [PubMed: 19218453]
15. Cilloniz C, et al. Lethal influenza virus infection in macaques is associated with early dysregulation of inflammatory related genes. *PLoS Pathog*. 2009; 5:e1000604. [PubMed: 19798428]
16. Cilloniz C, et al. Lethal dissemination of H5N1 influenza virus is associated with dysregulation of inflammation and lipoxin signaling in a mouse model of infection. *J Virol*. 2010; 84:7613–7624. [PubMed: 20504916]
17. Thomas PG, et al. The intracellular sensor NLRP3 mediates key innate and healing responses to influenza A virus via the regulation of caspase-1. *Immunity*. 2009; 30:566–575. [PubMed: 19362023]
18. Ichinohe T, Lee HK, Ogura Y, Flavell R, Iwasaki A. Inflammasome recognition of influenza virus is essential for adaptive immune responses. *J Exp Med*. 2009; 206:79–87. [PubMed: 19139171]
19. Allen IC, et al. The NLRP3 inflammasome mediates in vivo innate immunity to influenza A virus through recognition of viral RNA. *Immunity*. 2009; 30:556–565. [PubMed: 19362020]

20. Chang ST, Tchitchek N, Ghosh D, Benecke A, Katze MG. A chemokine gene expression signature derived from meta-analysis predicts the pathogenicity of viral respiratory infections. *BMC systems biology*. 2011; 5:202. [PubMed: 22189154]
21. Khondoker MR, et al. Multi-factorial analysis of class predication error: estimating optimal number of biomarkers for various classification rules. *Journal of Bioinformatics and Computational Biology*. 2010; 08:945–965. [PubMed: 21121020]
22. Pulendran B, Li S, Nakaya HI. Systems vaccinology. *Immunity*. 2010; 33:516–529. [PubMed: 21029962]
23. Poland GA, Ovsyannikova IG, Kennedy RB, Haralambieva IH, Jacobson RM. Vaccinomics and a new paradigm for the development of preventive vaccines against viral infections. *OMICS*. 2011; 15:625–636. [PubMed: 21732819]
24. Querec TD, et al. Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. *Nat Immunol*. 2009; 10:116–125. [PubMed: 19029902]
25. Nakaya HI, et al. Systems biology of vaccination for seasonal influenza in humans. *Nat Immunol*. 2011; 12:786–795. [PubMed: 21743478]
26. Dunham I, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489:57–74. [PubMed: 22955616]
27. Djebali S, et al. Landscape of transcription in human cells. *Nature*. 2012; 489:101–108. [PubMed: 22955620]
28. Skalsky RL, et al. The Viral and Cellular MicroRNA Targetome in Lymphoblastoid Cell Lines. *PLoS Pathog*. 2012; 8:e1002484. [PubMed: 22291592]
29. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet*. 2009; 10:155–159. [PubMed: 19188922]
30. Peng X, et al. Unique signatures of long noncoding RNA expression in response to virus infection and altered innate immune signaling. *mBio*. 2010; 1:e00206–e00210. [PubMed: 20978541]
31. Peng X, et al. Integrative deep sequencing of the mouse lung transcriptome reveals differential expression of diverse classes of small RNAs in response to respiratory virus infection. *mBio*. 2011; 2:00198-11.
32. Chang ST, et al. Next-generation sequencing reveals HIV-1-mediated suppression of T cell activation and RNA processing and regulation of noncoding RNA expression in a CD4+ T cell line. *mBio*. 2011; 2:e00134-11. [PubMed: 21933919]
33. Chang ST, et al. Next-Generation Sequencing of Small RNAs from HIV-Infected Cells Identifies Phased microRNA Expression Patterns and Candidate Novel microRNAs Differentially Expressed upon Infection. *mBIO*. 2013 *In press*.
34. Stern-Ginossar N, et al. Decoding human cytomegalovirus. *Science*. 2012; 338:1088–1093. [PubMed: 23180859]
35. Munger J, Bajad SU, Collier HA, Shenk T, Rabinowitz JD. Dynamics of the cellular metabolome during human cytomegalovirus infection. *PLoS Pathog*. 2006; 2:e132. [PubMed: 17173481]
36. Munger J, et al. Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy. *Nat Biotechnol*. 2008; 26:1179–1186. [PubMed: 18820684]
37. Vastag L, Koyuncu E, Grady SL, Shenk TE, Rabinowitz JD. Divergent effects of human cytomegalovirus and herpes simplex virus-1 on cellular metabolism. *PLoS Pathog*. 2011; 7:e1002124. [PubMed: 21779165]
38. McArdle J, Schafer XL, Munger J. Inhibition of calmodulin-dependent kinase kinase blocks human cytomegalovirus-induced glycolytic activation and severely attenuates production of viral progeny. *J Virol*. 2011; 85:705–714. [PubMed: 21084482]
39. Spencer CM, Schafer XL, Moorman NJ, Munger J. Human cytomegalovirus induces the activity and expression of acetyl-coenzyme A carboxylase, a fatty acid biosynthetic enzyme whose inhibition attenuates viral replication. *J Virol*. 2011; 85:5814–5824. [PubMed: 21471234]
40. Liu ST, et al. Synaptic vesicle-like lipidome of human cytomegalovirus virions reveals a role for SNARE machinery in virion egress. *Proc Natl Acad Sci U S A*. 2011; 108:12869–12874. [PubMed: 21768361]

41. McArdle J, Moorman NJ, Munger J. HCMV targets the metabolic stress response through activation of AMPK whose activity is important for viral replication. *PLoS Pathog.* 2012; 8:e1002502. [PubMed: 22291597]
42. Grady SL, Hwang J, Vastag L, Rabinowitz JD, Shenk T. Herpes simplex virus 1 infection activates poly(ADP-ribose) polymerase and triggers the degradation of poly(ADP-ribose) glycohydrolase. *J Virol.* 2012; 86:8259–8268. [PubMed: 22623791]
43. Blanc M, et al. Host Defense against Viral Infection Involves Interferon Mediated Down-Regulation of Sterol Biosynthesis. *PLoS Biol.* 2011; 9:e1000598. [PubMed: 21408089]
44. Pfefferle S, et al. The SARS-coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors. *PLoS Pathog.* 2011; 7:e1002331. [PubMed: 22046132]
45. Brass AL, et al. Identification of host proteins required for HIV infection through a functional genomic screen. *Science.* 2008; 319:921–926. [PubMed: 18187620]
46. Krishnan MN, et al. RNA interference screen for human genes associated with West Nile virus infection. *Nature.* 2008; 455:242–245. [PubMed: 18690214]
47. Sessions OM, et al. Discovery of insect and human dengue virus host factors. *Nature.* 2009; 458:1047–1050. [PubMed: 19396146]
48. Brass AL, et al. The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. *Cell.* 2009; 139:1243–1254. [PubMed: 20064371]
49. Li Q, et al. A genome-wide genetic screen for host factors required for hepatitis C virus propagation. *Proc Natl Acad Sci U S A.* 2009; 106:16410–16415. [PubMed: 19717417]
50. The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat Genet.* 2004; 36:1133–1137. [PubMed: 15514660]
51. Uetz P, et al. Herpesviral Protein Networks and Their Interaction with the Human Proteome. *Science.* 2006; 311:239–242. [PubMed: 16339411]
52. Friedel CC, Haas J. Virus-host interactomes and global models of virus-infected cells. *Trends Microbiol.* 2011; 19:501–508. [PubMed: 21855347]
53. Ma-Lauer Y, Lei J, Hilgenfeld R, von Brunn A. Virus-host interactomes - antiviral drug discovery. *Curr Opin Virol.* 2012; 2:614–621. [PubMed: 23057872]
54. Panda D, Cherry S. Cell-based genomic screening: elucidating virus-host interactions. *Curr Opin Virol.* 2012; 2:778–786.
55. Stertz S, Shaw ML. Uncovering the global host cell requirements for influenza virus replication via RNAi screening. *Microbes Infect.* 2011; 13:516–525. [PubMed: 21276872]
56. Meyniel-Schicklin L, de Chassey B, Andre P, Lotteau V. Viruses and interactomes in translation. *Mol Cell Proteomics.* 2012; 11:M111 014738. [PubMed: 22371486]
57. Rozenblatt-Rosen O, et al. Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. *Nature.* 2012; 487:491–495. [PubMed: 22810586]
58. Bottomly D, et al. Expression quantitative trait loci for extreme host response to influenza A in pre-collaborative cross mice. *G3 (Bethesda).* 2012; 2:213–221. [PubMed: 22384400]
59. Laird PW. Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet.* 2010; 11:191–203. [PubMed: 20125086]
60. Garcia BA. Mass spectrometric analysis of histone variants and post-translational modifications. *Front Biosci (Schol Ed).* 2009; 1:142–153. [PubMed: 19482690]
61. Knight JC. Genomic modulators of the immune response. *Trends Genet.* 2012
62. Ernberg I, Karimi M, Ekstrom TJ. Epigenetic mechanisms as targets and companions of viral assaults. *Ann N Y Acad Sci.* 2011; 1230:E29–E36. [PubMed: 22417105]
63. Marazzi I, et al. Suppression of the antiviral response by an influenza histone mimic. *Nature.* 2012; 483:428–433. [PubMed: 22419161]
64. Benecke A. Gene regulatory network inference using out of equilibrium statistical mechanics. *HFSP journal.* 2008; 2:183–188. [PubMed: 19404429]
65. Berg J. Out-of-equilibrium dynamics of gene expression and the Jarzynski equality. *Phys Rev Lett.* 2008; 100:188101. [PubMed: 18518416]

66. McDermott JE, Taylor RC, Yoon H, Heffron F. Bottlenecks and hubs in inferred networks are important for virulence in *Salmonella typhimurium*. *J Comput Biol.* 2009; 16:169–180. [PubMed: 19178137]
67. Girvan M, Newman MEJ. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences.* 2002; 99:7821–7826.
68. Segal E, et al. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet.* 2003; 34:166–176. [PubMed: 12740579]
69. Shapira SD, et al. A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. *Cell.* 2009; 139:1255–1267. [PubMed: 20064372]
70. Diamond DL, et al. Temporal proteome and lipidome profiles reveal hepatitis C virus-associated reprogramming of hepatocellular metabolism and bioenergetics. *PLoS Pathog.* 2010; 6:e1000719. [PubMed: 20062526]
71. McDermott JE, et al. Topological analysis of protein co-abundance networks identifies novel host targets important for HCV infection and pathogenesis. *BMC Syst Biol.* 2012; 6:28. [PubMed: 22546282]
72. Rasmussen AL, et al. Systems virology identifies a mitochondrial fatty acid oxidation enzyme, dodecenoyl coenzyme A delta isomerase, required for hepatitis C virus replication and likely pathogenesis. *J Virol.* 2011; 85:11646–11654. [PubMed: 21917952]
73. Tripathi LP, et al. Network based analysis of hepatitis C virus Core and NS4B protein interactions. *Molecular BioSystems.* 2010; 6:2539–2553. [PubMed: 20953506]
74. Josset L, et al. Increased viral loads and exacerbated innate host response in aged macaques infected with 2009 pandemic H1N1 influenza A virus. *J Virol.* 2012; 86:11115–11127. [PubMed: 22855494]
75. Lau KS, et al. In Vivo Systems Analysis Identifies Spatial and Temporal Aspects of the Modulation of TNF- α -Induced Apoptosis and Proliferation by MAPKs. *Sci. Signal.* 2011; 4:ra16-. [PubMed: 21427409]
76. Lau KS, et al. Multi-scale in vivo systems analysis reveals the influence of immune cells on TNF- α -induced apoptosis in the intestinal epithelium. *PLoS Biol.* 2012; 10:e1001393. [PubMed: 23055830]
77. Benecke A, Gale M Jr, Katze MG. Dynamics of innate immunity are key to chronic immune activation in AIDS. *Current opinion in HIV and AIDS.* 2012; 7:79–85. [PubMed: 22156845]
78. Benecke AG. Critical Dynamics in Host-Pathogen Systems. *Current topics in microbiology and immunology.* 2012
79. Soule C. Graphic Requirements for Multistationarity. *ComPlexUs.* 2003; 1:123–133.
80. Cinquin O, Demongeot J. High-dimensional switches and the modelling of cellular differentiation. *Journal of theoretical biology.* 2005; 233:391–411. [PubMed: 15652148]
81. Becavin C, Tchitchek N, Mintsä-Eya C, Lesne A, Benecke A. Improving the efficiency of multidimensional scaling in the analysis of high-dimensional data using singular value decomposition. *Bioinformatics.* 2011; 27:1413–1421. [PubMed: 21421551]
82. Becavin C, Benecke A. New dimensionality reduction methods for the representation of high dimensional 'omics' data. *Expert review of molecular diagnostics.* 2011; 11:27–34. [PubMed: 21171918]
83. Rasmussen AL, et al. Early transcriptional programming links progression to hepatitis C virus-induced severe liver disease in transplant patients. *Hepatology.* 2012; 56:17–27. [PubMed: 22278598]
84. Diamond DL, et al. Proteome and computational analyses reveal new insights into the mechanisms of hepatitis C virus-mediated liver disease posttransplantation. *Hepatology.* 2012; 56:28–38. [PubMed: 22331615]
85. Sonnenschein N, et al. A Network Perspective on Metabolic Inconsistency. *BMC systems biology.* 2012; 6:41. [PubMed: 22583819]
86. Ponnappalli SP, Saunders MA, Van Loan CF, Alter O. A higher-order generalized singular value decomposition for comparison of global mRNA expression from multiple organisms. *PLoS one.* 2011; 6:e28072. [PubMed: 22216090]

87. Huang SS, Fraenkel E. Integrating proteomic, transcriptional, and interactome data reveals hidden components of signaling and regulatory networks. *Science signaling*. 2009; 2:ra40. [PubMed: 19638617]
88. Luksza M, Lassig M, Berg J. Significance analysis and statistical mechanics: an application to clustering. *Physical review letters*. 2010; 105:220601. [PubMed: 21231375]
89. Josset L, et al. Gene Expression Signature-Based Screening Identifies New Broadly Effective Influenza A Antivirals. *PLoS One*. 2010; 5:e13169. [PubMed: 20957181]
90. Yang W-LR, Lee Y-E, Chen M-H, Chao K-M, Huang C-YF. In-silico drug screening and potential target identification for hepatocellular carcinoma using Support Vector Machines based on drug screening result. *Gene*. 2013; 518:201–208. [PubMed: 23220021]
91. Hood L, Friend SH. Predictive, personalized, preventive, participatory (P4) cancer medicine. *Nat Rev Clin Oncol*. 2011; 8:184–187. [PubMed: 21364692]
92. Bengoechea JA. Infection systems biology: from reactive to proactive (P4) medicine. *Int Microbiol*. 2012; 15:55–60. [PubMed: 22847266]
93. Southern EM. DNA microarrays. History and overview. *Methods Mol Biol*. 2001; 170:1–15. [PubMed: 11357674]
94. Gorreta, F.; Carbone, W.; Barzaghi, D. *Molecular Profiling*. Espina, V.; Liotta, LA., editors. Humana Press; 2012. p. 89-105.
95. Li W, Ruan K. MicroRNA detection by microarray. *Analytical and Bioanalytical Chemistry*. 2009; 394:1117–1124. [PubMed: 19132354]
96. Fouse SD, Nagarajan RP, Costello JF. Genome-scale DNA methylation analysis. *Epigenomics*. 2010; 2:105–117. [PubMed: 20657796]
97. Lam, C-W.; Lau, K-C.; Tong, S-F. *Advances in Clinical Chemistry*. Gregory, SM., editor. Elsevier: 2010. p. 1-18.
98. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet*. 2010; 11:31–46. [PubMed: 19997069]
99. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009; 10:57–63. [PubMed: 19015660]
100. Martin JA, Wang Z. Next-generation transcriptome assembly. *Nat Rev Genet*. 2011; 12:671–682. [PubMed: 21897427]
101. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*. 2009; 10:669–680. [PubMed: 19736561]
102. König J, Zarnack K, Luscombe NM, Ule J. Protein–RNA interactions: new genomic technologies and perspectives. *Nat Rev Genet*. 2012; 13:77–83. [PubMed: 22251872]
103. Collaborative-Cross-Consortium. The genome architecture of the Collaborative Cross mouse genetic reference population. *Genetics*. 2012; 190:389–401. [PubMed: 22345608]
104. Zou F, et al. Quantitative trait locus analysis using recombinant inbred intercrossoes: theoretical and empirical considerations. *Genetics*. 2005; 170:1299–1311. [PubMed: 15879512]

Box 1 |Microarrays

DNA microarrays are widely used for global transcriptome profiling and remain the workhorse technology of systems biology. Microarrays can trace at least part of their ancestry to the Southern blot, in which DNA fragments are immobilized on a filter membrane for subsequent detection by labeled DNA hybridization, and Southern himself has provided a detailed description of the many early technological developments that together led to the first commercial microarrays⁹³. Over the past 15 years, the technology has progressed from cDNAs deposited on membranes by spotting robots, to commercial microarrays consisting of tens of thousands of oligonucleotides on glass slides or other solid supports. Today, major commercial providers of microarrays include Affymetrix, Agilent Technologies, Illumina and NimbleGen, with platforms varying in the length of oligonucleotide used, the number of oligonucleotides representing each gene, and the methods used for oligonucleotide synthesis and attachment to solid supports. Although microarrays are best known for their use in profiling the expression of protein-coding genes, the technology has also been adapted for the profiling of microRNA expression, DNA methylation and single nucleotide polymorphisms (SNPs), as well as for promoter analysis and the detection of genome-wide DNA copy number variation. Methods have also been developed to extract RNA from formalin-fixed paraffin-embedded (FFPE) samples, which have opened the door for microarray-based gene expression profiling of the large numbers of clinical samples archived using this preservation method. Microarrays can now provide considerably more information about virus-host interactions than simply the differential expression of protein-coding genes in response to virus infection^{94–97}.

Box 2 |Next-generation sequencing

Next-generation sequencing is a massively parallel sequencing-by-synthesis approach that has replaced the first-generation automated Sanger sequencing method for genome analysis⁹⁸. The approach generates hundreds of millions of bases of high-quality DNA sequence per run and has enabled the rapid sequencing of entire genomes. It is also used for transcriptome profiling, an approach referred to as RNA-Seq⁹⁹. This has a number of advantages over microarray-based profiling, including the ability to identify and quantify rare or as-yet-undiscovered transcripts and to provide information on alternative splicing, genetic variation, and gene and exon boundaries. The method of cDNA library construction used determines the type of transcripts to be sequenced: small (less than about 200 bases) RNA sequencing (for example, microRNAs, snoRNAs and piRNAs), mRNA sequencing (various types of polyadenylated transcripts) or whole-transcriptome sequencing (polyadenylated and non-polyadenylated transcripts). Once sequence reads are generated, they are aligned to a known reference genome, or if no reference genome is available, sequences can be assembled de novo. Because of the short length of sequence reads, the alignment and assembly of sequence information is a significant computational challenge¹⁰⁰. Additional potential applications of next-generation sequencing of relevance to systems virology include the detection of protein-DNA (ChIP-Seq) or protein-RNA (RIP-Seq/CLIP-Seq) [G] binding events and DNA methylation profiling (Methyl-Seq)^{59, 101, 102}.

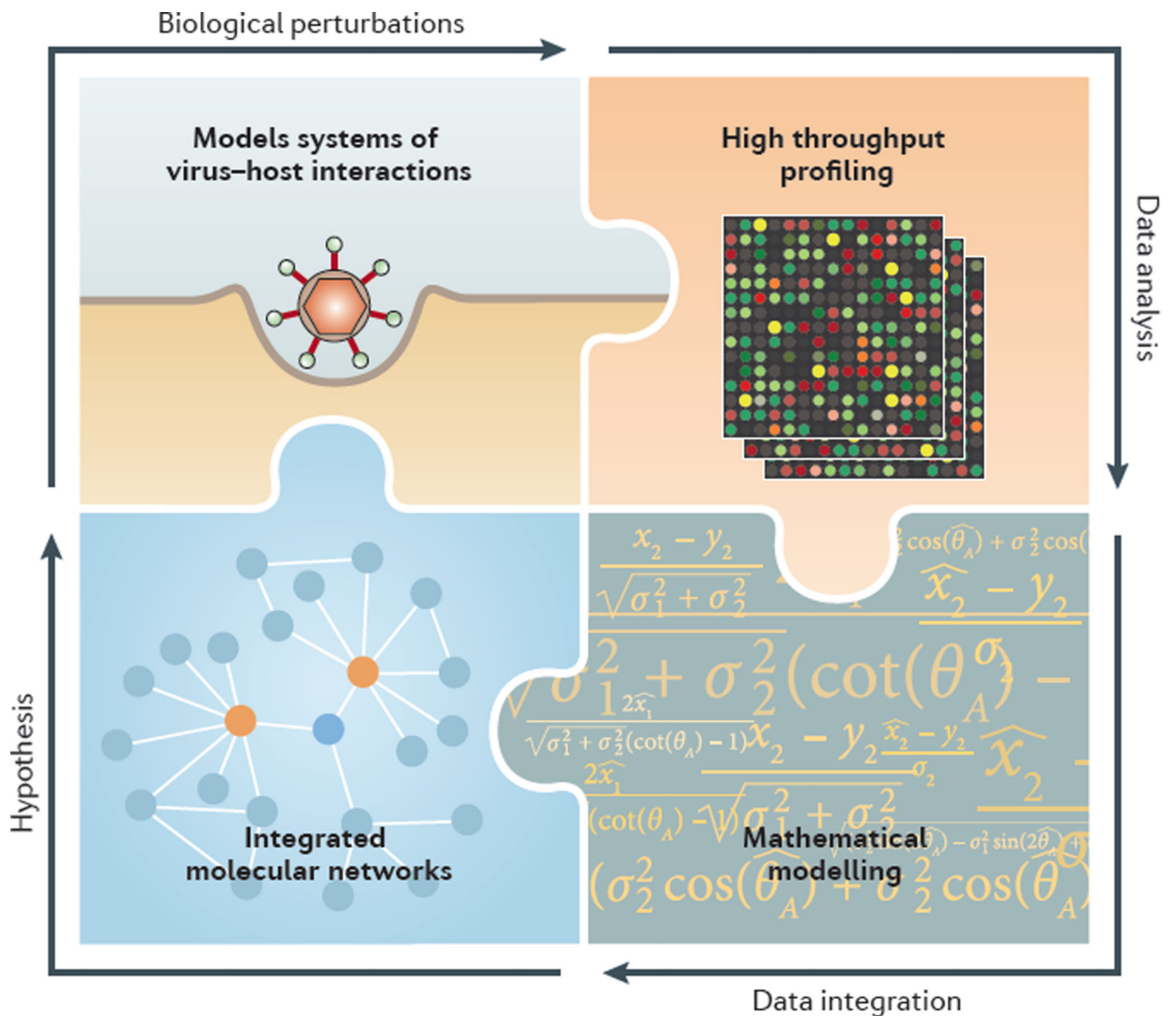
Given the advantages of next-generation sequencing, some have suggested that it will soon replace the use of microarrays for transcriptome profiling. However, the enormous amount of data generated by next-generation sequencing requires an extensive information technology and computational infrastructure for data processing, storage and analysis. These costs and complexities, coupled with the cost of the sequencing itself, make it likely that microarrays will continue to be widely used for some time.

Box 3 | The Collaborative Cross

The Collaborative Cross is a recombinant inbred mouse genetic reference population derived from eight laboratory mouse strains⁵⁰. Such populations are used to study complex traits and are most beneficial when they contain large numbers of lines that exhibit substantial and uniform genetic variation. In the case of the Collaborative Cross, five inbred strains (A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ and NZO/HILtJ) and three wild-derived strains (CAST/EiJ, PWK/PhH and WSB/EiJ) are being used to generate a large panel of recombinant inbred (RI) lines through a randomized funnel breeding scheme¹⁰³, resulting in hundreds of independent RI lines in which genetic variation is randomly and uniformly distributed throughout the genome. Once established, RI lines can be intercrossed among themselves to generate recombinant intercross (RIX) lines or backcrossed to other non-Collaborative Cross lines to generate recombinant inbred backcross (RIB) lines. RIX lines maximize genetic diversity, and because they are outbred, they are better models of human populations and they can provide information on gene dominance and epistasis [G]¹⁰⁴. RIB lines can be used, for example, to evaluate whether specific allele combinations alter the phenotype associated with a given target gene (for example, a dominant targeted gene knockout). Comparison between viral titer in lungs and the extent of airway inflammation from a panel of pre-Collaborative Cross mice (not fully inbred) demonstrates the impact of genetic diversity on influenza virus disease phenotypes (see the figure; each diamond represents a different mouse). Phenotypes range from high viral titer and inflammation to low viral titer and little inflammation; in addition, some animals show unique phenotypes, such as high viral titer with little accompanying inflammation. The Collaborative Cross is being developed as a resource for the biomedical research community, with genotype information and mouse line availability publically available at UNC Systems Genetics.

Box 4 |The importance of data sharing

The rapid dissemination of high-throughput omics data sets to the scientific community is an important element for the future success of systems virology. It is only through meta-analyses that rely on advanced computational and mathematical approaches and the integration of large, coherent and systematic multi-dimensional data sets that the complexity of virus-host interactions can be unraveled. Owing to the immense amount of time, expertise, manpower and expense that are required to produce these data sets, it is imperative for the systems virology community to share data in a timely manner. In most cases, owing to the richness of information contained in these data sets, this can be done with little chance of compromising one's own research. Two of the National Institute of Allergy and Infectious Diseases (NIAID) sponsored Bioinformatics Resource Centers, ViPR and IRD, are charged with capturing, publically sharing, storing, integrating and visualizing systems-wide high-throughput omics data sets (both raw and processed) that detail the host response to virus infection. Importantly, in addition to capturing data types such as transcriptomic, proteomic, metabolomic, lipidomic and ChIP-Seq, their goal is to capture extensive metadata, which provide users with the experimental details needed to facilitate data interpretation. The practice of linking metadata to several forms of the omics data (raw, analyzed or modeled) enables ViPR and IRD to meet the demands of a range of end users (see the figure). For instance, at their portals, users can search for information on individual genes, download lists of differentially expressed genes or link metadata to raw data for doing meta-analysis. All laboratories, as a service to the community, should be committed to sharing these types of data sets, whether through organization-specific public websites, the appropriate NIAID Bioinformatics Resource Center, or a combination of both.



Nature Reviews | Microbiology

Figure 1. The systems virology paradigm

Using appropriate model systems of virus-host interactions, high throughput profiling techniques are used to generate multi-dimensional data. Data analysis and mathematical modelling are combined to generate comprehensive, integrated, and predictive molecular networks of biological systems and virus-host interactions. Resulting predictions and hypotheses lead to a subsequent round or cycle of biological perturbations. Completion of each cycle results in model refinement and a deeper understanding of complex biological processes. These findings and outcomes can be used directly or further refined by the scientific community for various types of disease intervention.

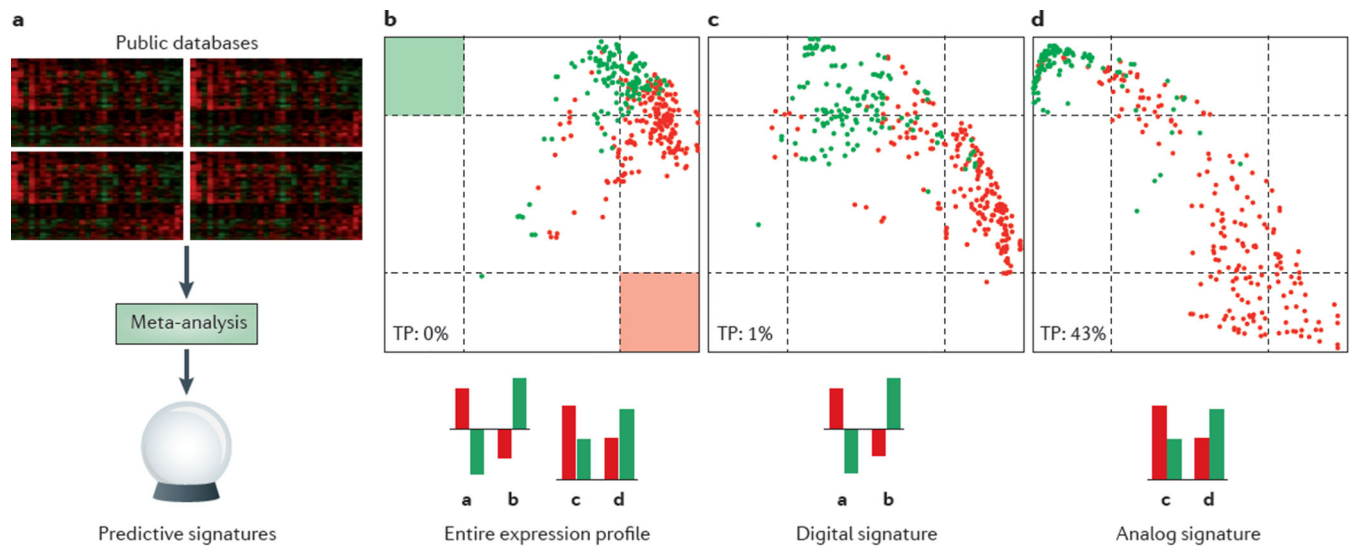
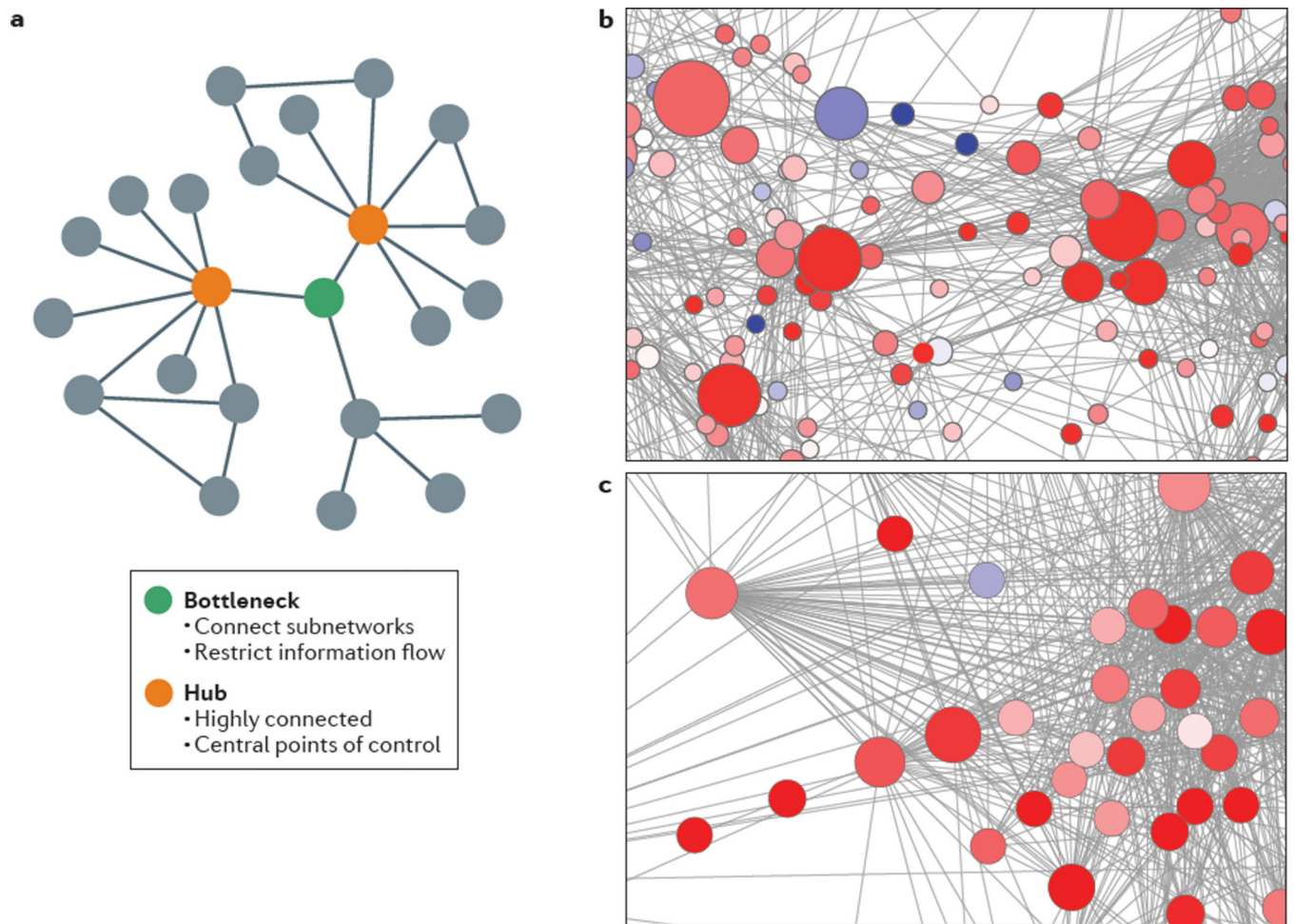


Figure 2. A 57-gene analogue signature predicts respiratory virus pathogenicity

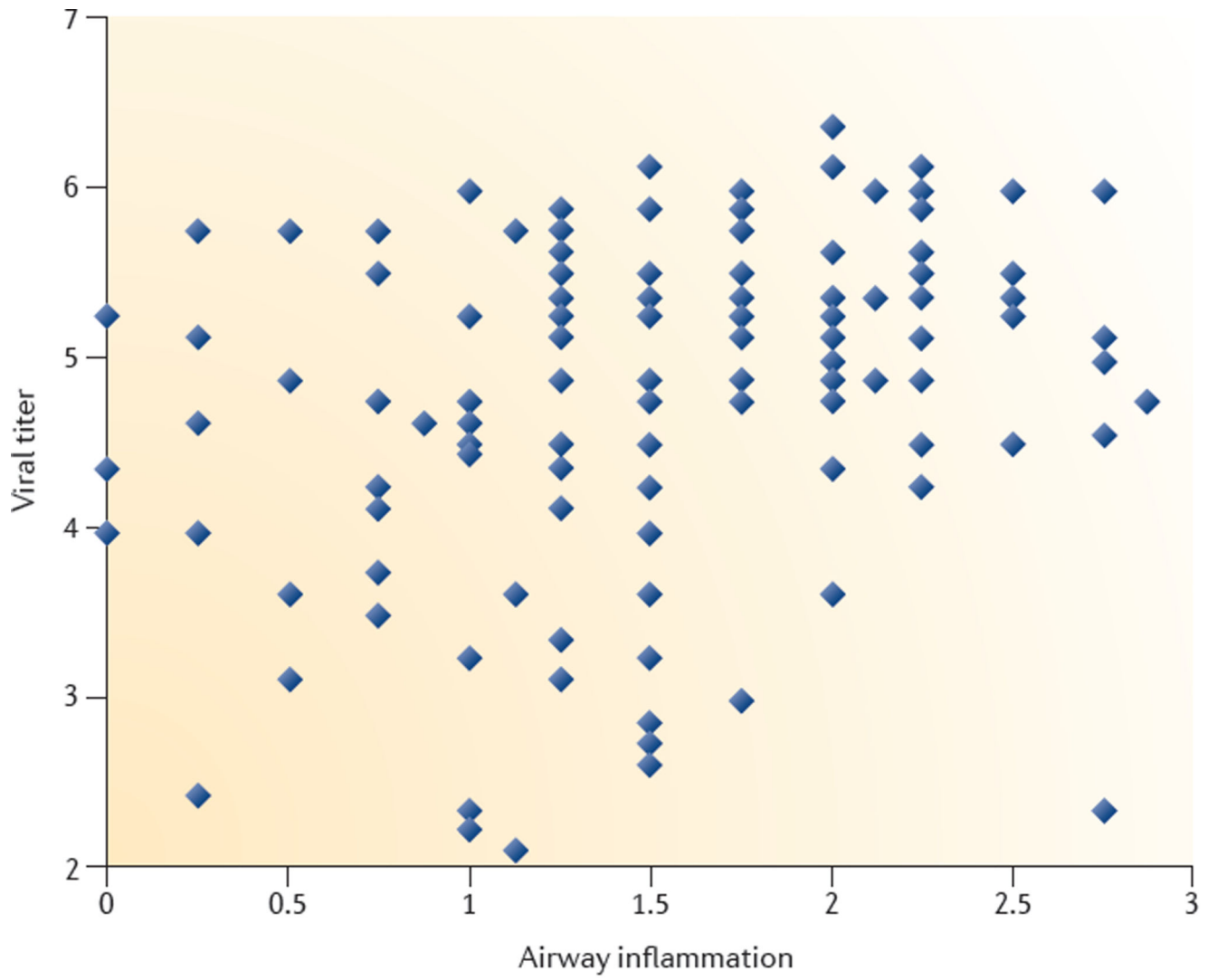
a A compendium was compiled from publically available data from transcription profiles on lung samples from mice infected with influenza virus or SARS coronavirus. A meta-analysis was performed and two signatures predictive of severity of infection (non-lethal verse lethal) were derived. The digital signature is comprise of a subsets of genes that changes in opposite direction compared to mock and the analog signature is comprise of a subsets of genes the are differentially expressed between the two groups without referencing mock. **b–d** Each dot represents a gene expression profile ; non-lethal infections are indicated by a green dot and lethal infections are indicated by a red dot. Classification results using the entire gene expression profile (b), the digital gene signature (74 genes) (c) or the analogue gene signature (57 genes) (d). Regions defined for positive identification of non-lethal (green square) or lethal (red square) infections are indicated in b. The analogue signature was more accurate at predicting lethal virus infection (TP=43%) than either the digital signature (TP=1%) or the entire gene expression profile (TP=0%).



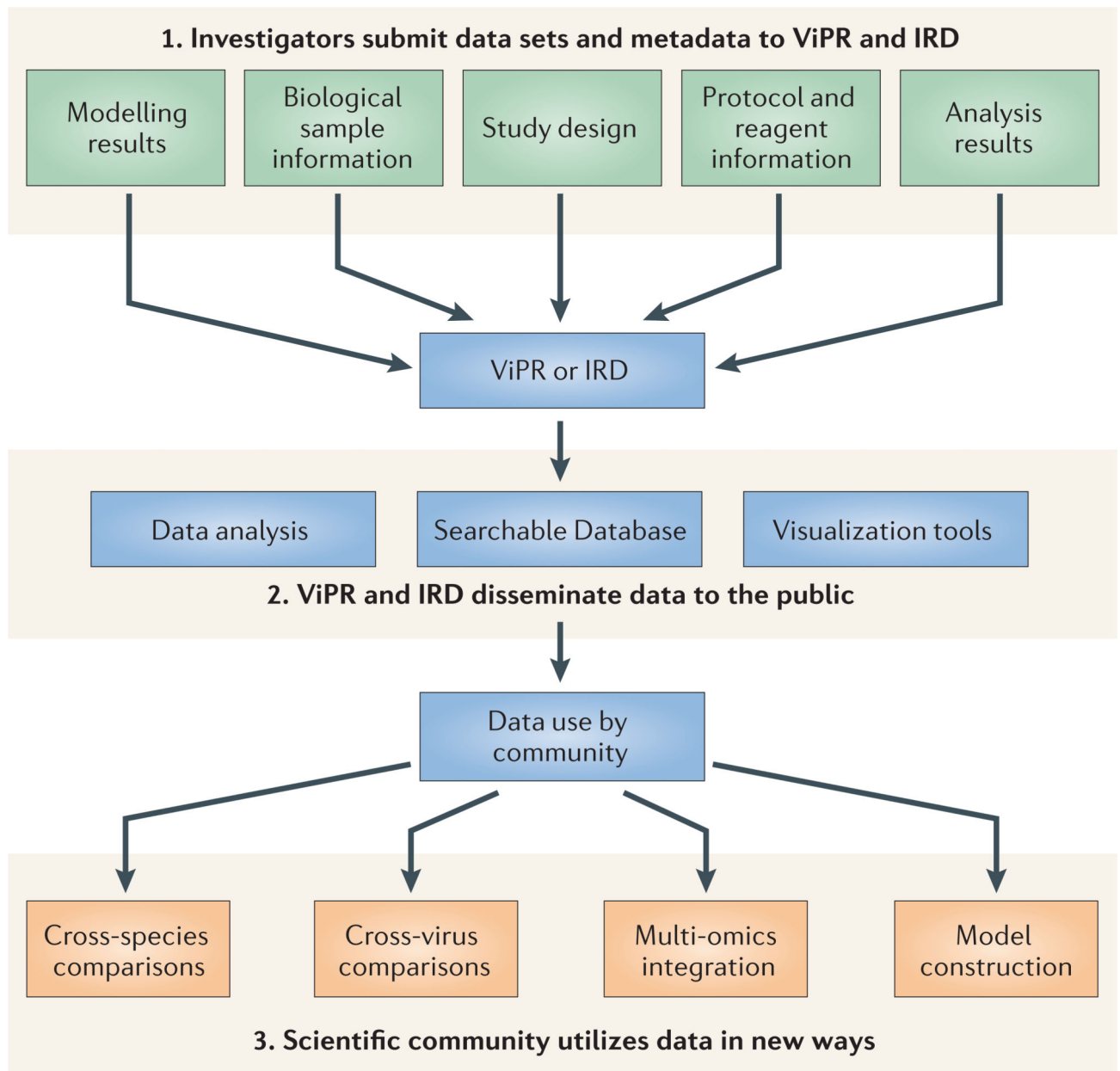
Nature Reviews | Microbiology

Figure 3. Co-regulation networks

a A schematic showing one bottleneck gene (green circle or node) and two hub genes (orange nodes) in a hypothetical network. Hubs are highly connected to other nodes in the network. Bottlenecks have a high degree of betweenness centrality and connect or act as bridges between sub-networks. **b and c** The context likelihood of relatedness method was used to infer functional associations between differentially expressed genes responding to H5N1 influenza virus infection in mice. Genes (nodes) are colored based on the ratio of gene expression values of high-dose to low-dose infection. Red indicates higher expression and blue indicates lower expression in the high dose compared to the low dose. **b.** The complexity of these networks does not allow for a visual determination of bottlenecks. Rather the betweenness centrality is calculated and the size of the node is relative to this value. The larger nodes are major bridges between different parts of the network. **c.** A major hub from a separate subnetwork with in the H5N1 influenza co-regulatory network.



Box 3 fig.



Box 4 fig.