Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Chapter 11

Systems Virology

Why everybody wants to measure everything

Marcus J. Korth[1,2], G. Lynn Law[1,2]

[1]*Department of Microbiology, University of Washington, Seattle, WA, USA;* [2]*Washington National Primate Research Center, University of Washington, Seattle, WA, USA*

Chapter Outline

1. INTRODUCTION

Virologists have long known that viral pathogenesis must be studied from the standpoint of both the virus and the host. Nevertheless, given its relative simplicity, studying the virus has always been more tractable. As outlined in the previous chapters, virus-centric approaches have yielded a tremendous amount of information about viral genetics, viral replication cycles, and host and tissue tropisms. Along the way have come insights into host innate and adaptive immune responses and the many ways in which viruses antagonize these responses while exploiting other cellular processes to their advantage. In the last decade, however, new opportunities to study the host response have emerged. In 1990, the National Institutes of Health and the Department of Energy announced a plan to map and sequence the human genome. Eleven years later, the first draft of the genome was released, and in 2003, the project was declared complete.

The sequencing of the human genome dramatically changed the field of viral pathogenesis. Virologists were now able to move beyond virus-centric or single-gene approaches and instead investigate the host response to infection on a genome-wide scale. With the human genome sequence in hand, it became possible to predict the complete constellation of human genes, their corresponding mRNA transcripts, and encoded protein products. This information spurred the development of methods to measure global gene expression and protein abundance, which in turn mandated the development of computational methods to interpret the resulting avalanche of data (Sidebar 1).

In this chapter, we focus on the insights into viral pathogenesis that are provided by examining the host transcriptional response, including the dynamics of innate and acquired immunity, diagnostic signatures, and the identification of targets for antiviral drugs. We also touch briefly on data interpretation and on protein and metabolite profiling. Although the methods used to measure protein and metabolite abundance (i.e., chromatography and mass spectrometry) differ from those used to measure gene expression, downstream data analysis approaches are similar, and the integration of gene expression and protein and metabolite abundance data brings us closer to a true systems level understanding of virus–host interactions. The use of large-scale genetic and protein–protein interaction screens to identify host proteins that promote

Viral Pathogenesis. http://dx.doi.org/10.1016/B978-0-12-800964-2.00011-2

Sidebar 1 The evolution of systems biology

The sequencing of the human genome and the advent of high-throughput molecular profiling are widely credited with giving rise to systems biology. In one sense this may be true. The convergence of genome sequence information, profiling technologies, and computational advances made it possible to examine biological systems on a scale never before possible. However, the concepts underlying systems biology, and an understanding of the need to comprehend complete systems, have deeper roots. The notion of emergent properties–properties or outcomes that cannot be predicted by an understanding of the individual parts of a system alone—dates back at least to the time of Aristotle (384–322 BC), who stated: "the whole is something over and above its parts and not just a sum of them all." Nevertheless, reductionism—the idea that complex systems can be analyzed and understood by reducing them to manageable pieces—held sway through much of modern history. In the early 1900s, views began to change. It became apparent, for example, that biological systems have hierarchies of organization, and that components of a system behave differently in isolation than when in the intact system (Trewavas, 2006).

The first known use of the term systems biology, and its proposal as a distinct discipline, is attributed to Mihajlo Mesarovic in his 1968 book *Systems Theory and Biology*. Then, as now, systems biology was met with some skepticism. A reviewer of the book noted: "There is no doubt that system-theoretic ideas seem somewhat strange, and perhaps just a little frightening, to the present generation of structurally-oriented biologists" (Rosen, 1968). The current concept of systems biology—driven by genome-based technologies and mathematics—is most often associated with Leroy Hood, founder of the Institute for Systems Biology. Hood proposed that biological systems are composed of two types of information, genes and networks of regulatory interactions, and that biology be viewed as an informational science (Ideker et al., 2001). In this view, studying biological systems requires detailed knowledge of the components of the system, systematic perturbation of the system, monitoring of gene, protein, and pathway responses, and the formulation of mathematical models to describe the system and its response to perturbation (viral infection being an example of such a perturbation). Today, systems biology has become a driving force in biology and medicine, although it is still often criticized for being too focused on data acquisition. As put by Nobel Laureate Sidney Brenner: "Everybody wants to measure everything, you'll never get anything out of it." Such criticisms may eventually be muted, however, as systems biology continues to evolve, driven currently by rapid advances in computing, mathematics, and network modeling, which will be essential to making complex systems comprehensible.

or inhibit viral survival is the subject of Chapter 12, The Virus–Host Interactome.

2. USING GENOMICS TO STUDY VIRAL PATHOGENESIS

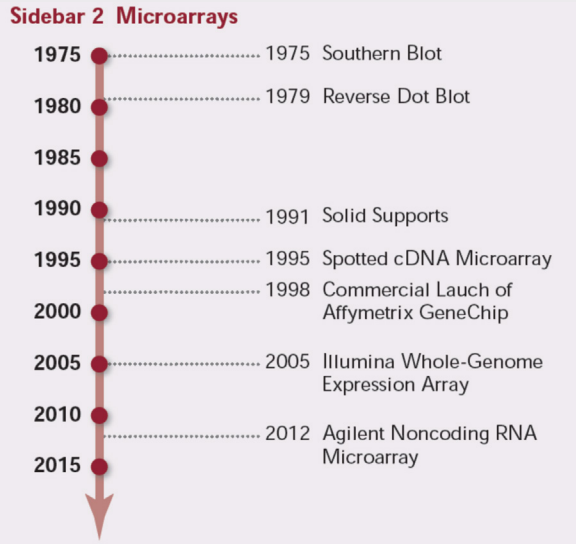
On the most basic level, profiling the host transcriptional response to viral infection entails measuring changes in the level of mRNA transcripts present in a cell population in the presence or absence of virus. Indeed, the first published application of large-scale genomic profiling in virology was essentially this simple; the study examined the transcriptional response of primary human fibroblasts at three time points following infection with human cytomegalovirus (Zhu et al., 1998). Yet even this simple experimental plan, which monitored the expression of approximately 6000 genes, revealed not only the complexity of the host response, but also the complexity of data interpretation. Changes in transcript abundance can be due to changes in synthesis, stability, or degradation and such changes may or may not correspond to changes in protein abundance or activity. The transcriptional response is dependent upon time, cell type, virus, and other parameters. Even today, the biological function of many genes is unknown. And with the number of transcripts profiled by current technologies numbering in the tens of thousands, the computational requirements for data analysis are considerable.

The initial study of the host response to human cytomegalovirus identified 258 mRNA transcripts that changed by a factor of four or more, including transcripts encoding major histocompatibility complex I surface receptors and multiple components of the pathway that produces prostaglandin E2 (an inflammatory mediator). The authors concluded: "The global analysis of changes in mRNA levels provides a catalog of genes that are modulated as a result of the host–pathogen interaction and therefore deserve further scrutiny." To a large extent, this forward-looking statement also sums up what is perhaps the main challenge associated with genomic profiling. The global analysis of transcription produces a *catalog* of differentially expressed genes, and oftentimes this catalog is extremely large. Investigators must devise strategies to sift through these catalogs and determine which genes deserve further scrutiny. Ideally, genomic profiling should also do more than produce lists; it should reveal interrelationships between genes, the structure and activity of gene networks, and the function of genes for which no role has been previously ascribed. There is also meaning to be gained from patterns within the data. As described later, this information can be used for predictive or diagnostic purposes or for computational screens for new antiviral drugs.

2.1 Transcriptional Profiling

The earliest assays for large-scale transcriptional profiling consisted of cDNAs or oligonucleotides spotted onto

Sidebar 2 Microarrays



DNA microarrays can trace their ancestry to the Southern blot, in which genomic DNA, digested into fragments by restriction enzymes, is immobilized on a permeable membrane filter for subsequent detection by labeled DNA hybridization. The Southern blot gave rise to the "reverse" dot blot, in which synthetic oligonucleotides of known sequence (called probes) were immobilized on permeable membrane supports. The collection of DNA sequences to be analyzed (called targets) were then labeled and applied to the membrane under hybridization conditions. The use of permeable membranes, however, made it difficult to control the size and shapes of the spotted DNA and limited miniaturization. These limitations were overcome by the introduction of solid supports, which provided the ability to accurately control the size, shape, and location of the spots. The robotic spotting of cDNAs onto glass slides—the first of what we now know as microarrays—was pioneered by Patrick Brown at Stanford University (Schena et al., 1995). In an interview with *Discover* magazine, Brown explained that when he applied for a grant to develop the technology: "The microarray part was thoroughly rejected…but I just decided that I would make one anyway." Importantly, the use of solid supports also facilitated the development of methods for the in situ synthesis of nucleic acids on the surface using approaches such as ink jet fabrication (spotting droplets of nucleotide reagents instead of droplets of ink) or photolithographic methods (similar to those used in the semiconductor industry). 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. Microarray technology has also been adapted for the profiling of noncoding RNA expression, DNA methylation, and single nucleotide polymorphisms, as well as for promoter analysis and the detection of genome-wide DNA copy number variation.

nitrocellulose membranes. These first "arrays" have since been replaced by commercially available platforms in which tens of thousands of oligonucleotides are arrayed at high density on glass slides or other solid supports. (Protein microarrays and protein–protein interaction profiling are discussed in Chapter 12, The Virus–Host Interactome.) Yet even the wealth of information available through the use of microarrays cannot compare to the level of information obtained through the direct sequencing of RNA transcripts (RNA-seq). Next-generation sequencing, a sequencing-by-synthesis approach that has replaced the first-generation Sanger sequencing method, is capable of yielding a truly comprehensive view of the transcriptome (discussed in more detail in Section 6 of this chapter; Sidebar 2).

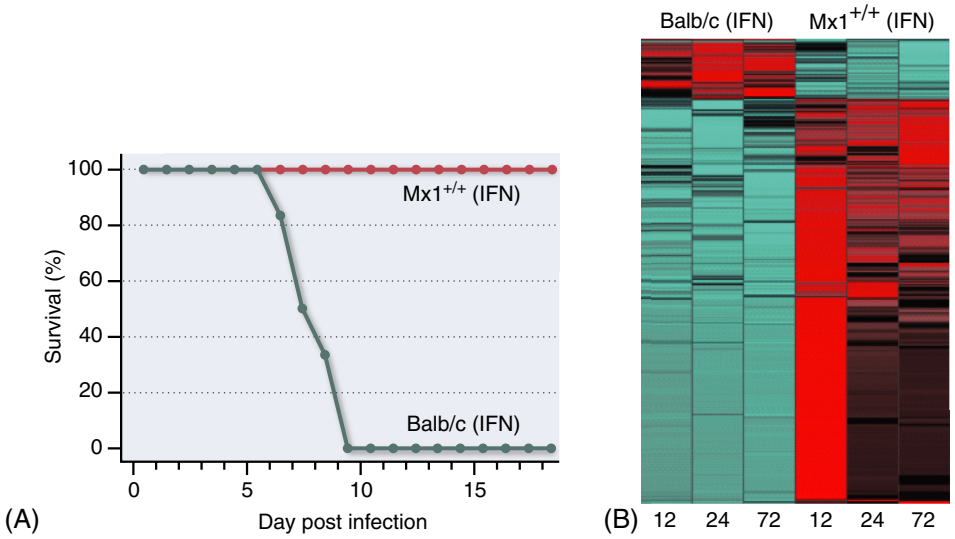
Whether generated by microarray analysis or RNA-seq, a transcriptional profile is the pattern of gene expression that is observed in a biological system. Most often, a comparison is made between a normal or resting state and one or more time points following perturbation of the system. The simpler the biological system, the easier it is to manipulate and to generate and interpret transcriptional data. For virology, a simple system means infecting cultured cells. Cell culture experiments facilitate the rapid examination of multiple time points or viruses, and they are useful for examining and modeling intracellular signaling in response to infection. However, the limited phenotypic parameters that can be studied using cultured cells—viral replication or cytopathic effect—makes it difficult to study viral pathogenesis. In contrast, with an animal model, it is possible to measure a variety of disease parameters—e.g., virus yields, clinical signs, gross pathology, histopathology, and time to recovery or death—and the cells or tissues that are examined come from their natural environment. Of course, these benefits are balanced by the cost, ethics, and complexity of animal models.

2.2 Interpreting Transcriptional Profiles

The interpretation of the transcriptional profiles associated with viral infection, or other perturbation, can take a variety of forms. In all cases, it begins with the computational determination of genes that are differentially expressed between conditions (or time points) to a statistically significant degree. This information alone can yield important scientific insights. For example, transcriptional profiling of interferon-treated cultured cells, or tissues from whole animals, has been used to define the hundreds of genes that are induced by this innate immune cytokine, many of which are important players in the innate antiviral response. With

an understanding of interferon-induced gene expression in hand, it is then possible to evaluate the extent to which a virus may induce interferon signaling, or alternatively, encode mechanisms to counteract the interferon response.

Similarly, transcriptional profiling of gene knockout cells or animals can be used to evaluate the role of specific genes in the host response to infection. These studies reveal both the changes in gene expression that occur as a result of the gene knockout as well as the effect of the knockout on the host response. Often, these studies also reveal the existence of compensatory signaling mechanisms, which can complicate data interpretation. A slight twist on this approach has been used to study *Mx1*, an interferon-induced gene that confers resistance to influenza virus infection. *Mx1* is lacking in most laboratory strains of mice, and these animals are highly susceptible to many strains of influenza virus, including the reconstructed 1918 pandemic virus. Transcriptional profiling of the lungs of 1918 virus-infected wild-type BALB/c mice (which lack a functional *Mx1*), and *knock-in* mice carrying a functional *Mx1*, demonstrates profound differences in gene expression, which correlate with reduced mortality in $Mx1^{+/+}$ mice (Cilloniz et al., 2012). Treatment of $Mx1^{+/+}$ mice with interferon prior to infection increases survival to 100% and is associated with an increase in the expression of genes related to intercellular signaling or cellular movement and a decrease in the expression of inflammatory cytokine and chemokine genes (Figure 1). Here, transcriptional profiling, attached to a relatively straightforward study design, yielded information on the function of *Mx1* in mediating resistance to influenza virus and on the role of interferon signaling in this process.



(C)

Biological function	P-value
Cell-to-cell signaling and interaction	1.91E-06 – 7.17E-02
Cellular movement	1.17E-05 – 7.17E-02
Antigen presentation	3.5E-04 – 7.09E-02
Cellular growth and proliferation	2.90E-03 – 7.17E-02
Cellular development	5.56E-03 –7.06E-02
Cell signaling	5.79E-03 – 6.09E-02
Molecular transport	5.79E-03 – 6.67E-02
Vitamin and mineral metabolism	5.79E-03 – 6.52E-02
Amino acid metabolism	1.05E-02 – 3.42E-02
Gene expression	1.05E-02 – 6.52E-02

FIGURE 1 Interferon-treated $Mx1^{+/+}$ mice are resistant to lethal 1918 pandemic influenza virus infection. (A) Survival plot of a total of 18 mice (9 animals/mouse strain; BALB/c mice lack a functional *Mx1* gene). (B) Heat map illustrating 2071 differentially expressed genes associated with complete protection of $Mx1^{+/+}$ animals compared with the wild-type animals during interferon treatment; cutoff values were ≥twofold change and $P \leq 0.01$ (ANOVA), false discovery rate corrected. Red indicates up-regulation and blue indicates down-regulation. (C) Top 10 biological functions, as determined by using Ingenuity Pathway Analysis, assigned to the 2071 differentially expressed genes. *Adapted from Cilloniz et al. (2012), with permission.*

In the example above, differentially expressed genes are viewed as a heat map, where relative increases or decreases in gene expression are depicted using color gradations, and genes that exhibit similar patterns of expression are grouped together (through the use of various clustering algorithms). Once differentially expressed genes are identified, it is important to know the biological or biochemical function of their encoded products. For many genes, this information can be obtained from a variety of public (e.g., DAVID, maintained by the National Institute of Allergy and Infectious Diseases) or proprietary (e.g., Ingenuity Pathway Analysis) databases. These databases bring together information about individual genes from multiple sources, including the scientific literature and other databases. DAVID, for example, includes over 40 annotation categories, including gene ontology, protein–protein interactions, protein functional domains, disease associations, and tissue expression.

Bioinformatics tools are then used to sift through this information to deconstruct large lists of differentially expressed genes into functionally related groups. Again as seen in the example above, grouping genes in this way helps to interpret and understand the biological information available in high-throughput data. It is important to keep in mind, however, that many genes lack functional annotation or they are only poorly annotated. This means that considerable information may be lost if data analyses do not move beyond a survey of genes with already known function. Fortunately, one of the broader outcomes of transcriptional profiling studies should be an increase in functional annotation. This will come in part as investigators use network modeling approaches (described later) to piece together the interactions or common expression patterns that link various genes and which can be used to infer gene function.

2.3 Dynamics of the Host Response to Influenza Virus

Evaluating how transcriptional profiles change over time provides insights into how the timing and magnitude of the host response impacts viral pathogenesis and disease outcome. This has been particularly well studied in the case of influenza virus. This virus is renowned for its ability to continually generate new variants, and although most variants cause a relatively mild respiratory disease, others can cause severe and even fatal infections. The 1918 pandemic virus, for example, caused over 50 million deaths worldwide, making it responsible for one of the deadliest infectious disease outbreaks in human history. Since the late 1990s, highly pathogenic avian H5N1 influenza viruses have caused sporadic infections in humans (popularized in the news media as bird flu) with a mortality rate estimated at 60%.

Transcriptional profiling of lung tissue from mice or macaques experimentally infected with highly pathogenic influenza viruses has revealed that these viruses elicit a rapid induction of proinflammatory cytokine and chemokine genes, an event often referred to as a cytokine storm. These genes remain highly expressed until the death of the animal. In contrast, animals infected with less-pathogenic viruses exhibit a rapid induction of interferon and innate immune response genes, a response that resolves over time as the animals recover. In macaques, the 1918 virus also induces a disproportionate induction of genes associated with the inflammasome, a group of genes that are thought to be part of a protective innate immune response to influenza viruses. Excessive activation of the inflammasome, however, appears to be detrimental rather than protective. In mice, highly pathogenic avian H5N1 viruses also down-regulate anti-inflammatory genes, including *Alox5*, responsible for the biogenesis of lipoxins, and *Socs2*, encoding a suppressor of cytokine signaling that can be induced by lipoxins to control the inflammatory response. This excessive and sustained inflammatory response appears to promote immunopathology.

An innovative study using automated image analysis, gene expression profiling, and flow cytometry has shed additional insight into the relative contributions of direct viral damage or immunopathology to lethal influenza viral infection (Brandes et al., 2013). Using what is referred to as a top-down systems analysis approach (the breaking down of a complex system into finer details), transcriptional profiles were first obtained from whole lung and then from individual immune cell populations isolated from the lungs of mice infected with lethal or nonlethal influenza virus. Rather than focusing on specific gene expression changes, transcriptional data were organized into 50 modules consisting of genes with coordinate patterns of expression (Figure 2). A module of genes annotated as highly proinflammatory was uniquely associated with lethal infection. This inflammatory signature was shown to largely originate from neutrophils that rapidly migrate to the site of infection (which is poorly contained in the case of the lethal virus). The authors suggest that this initiates a chemokine-driven feed-forward pathway in which the first neutrophils at the scene release chemokines that attract additional neutrophils, resulting in a rapidly escalating inflammatory response and tissue damage. In support of this finding, attenuation of this self-amplifying process by experimental reduction (but not elimination) of neutrophil numbers increased survival without changing viral spread.

An alternative computational methodology incorporating a geometric representation method (singular value decomposition-multidimensional scaling) has been used to visualize and quantify the kinetics of the host transcriptional response to wild-type and attenuated variants of highly pathogenic avian H5N1 viruses (Tchitchek et al., 2013). This approach was used to analyze 230 transcriptomic and 198 proteomic profiles derived from the lungs

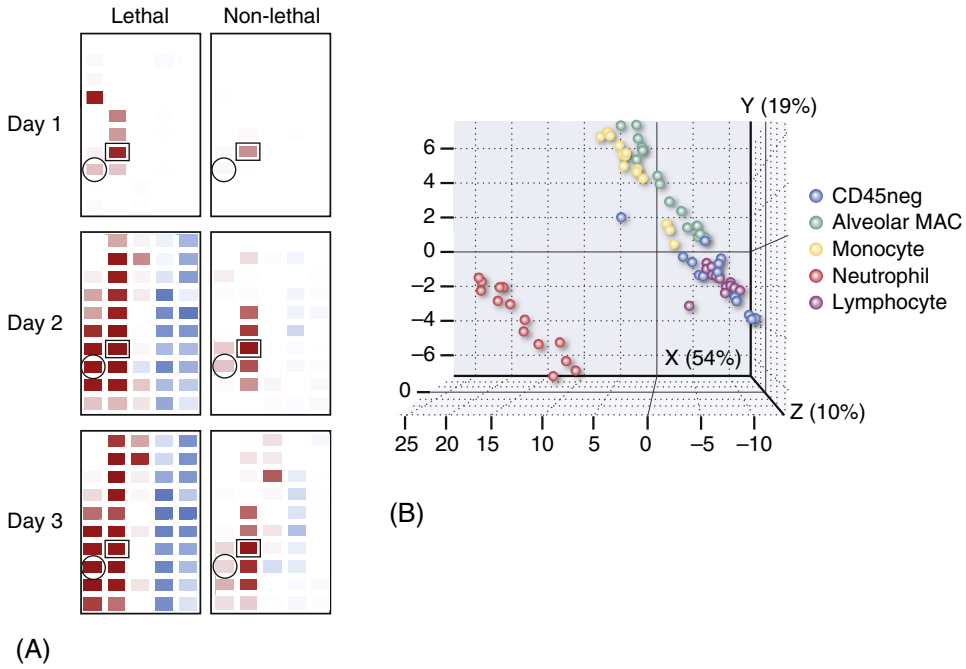


FIGURE 2 Shared and virus-specific innate inflammatory signatures characterize the response to influenza virus infection. (A) Modular map analysis reveals early shared antiviral signatures (module bounded by a rectangle) and virus-specific inflammatory signatures (module bounded by a circle). (B) Neutrophil infiltrates largely account for the lethality-associated inflammatory signature. Shown is principal component analysis of sorted cell microarrays based on genes in the inflammatory module. Amplitude in X largely depends upon whether samples were obtained from neutrophils. *Adapted from Brandes et al. (2013), with permission.*

of infected mice. The analysis revealed that the wild-type and mutant viruses elicit the differential expression of many of the same genes, but that it is the magnitude (the degree to which a gene is differentially expressed) and velocity of the initial host response (the speed at which these changes occur) that best correlates with pathogenic outcome.

2.4 Dynamics of Innate Immunity and AIDS

Transcriptional profiling studies have also shed light on the importance of the dynamics of immune activation in AIDS. This has come primarily from studies of various nonhuman primate species infected with simian immunodeficiency virus (SIV). Natural hosts for SIV, such as sooty mangabeys and African green monkeys, do not develop AIDS when naturally or experimentally infected with the virus. In contrast, Asian monkeys, such as rhesus or pig-tail macaques, develop AIDS following SIV infection. Several independent transcriptional profiling studies have shown that natural hosts for SIV exhibit an innate immune response to the virus that is comparable to that exhibited by Asian macaques. Initial levels of viral replication are also comparable. However, in natural hosts the innate immune response is of limited duration, whereas in macaques the response is sustained (Figure 3) (Jacquelin et al., 2009). This leads to chronic immune activation, which is associated with progression to AIDS in SIV-infected macaques and in HIV-infected individuals.

Because innate antiviral responses are often accompanied by inflammatory reactions, negative regulatory mechanisms, such as the induction of *IL-10*, are necessary to return antiviral responses to baseline to prevent the harmful effects of immune activation. The question remains as to why natural host species are able to resolve the initial response to SIV, whereas macaques (or most HIV-infected individuals) cannot. One hypothesis, proposed as the West Coast Model of immune activation, suggests that the kinetics of activation holds the key (Benecke et al., 2012). The West Coast Model likens virus infection to a wave on the beach, and immune activation as the mounting of a surf board. Mount too early or too late and bad things happen. If correct, the model has important implications for AIDS vaccine design as it suggests that the question of whether an adaptive immune response can be mounted against SIV (or HIV) could be irrelevant. Rather, it may be beneficial to devise strategies to modulate the timing or duration of the response, or to prevent chronic immune recognition of the virus. (Additional information on the kinetics of HIV and SIV infection and lessons learned from natural host species can be found in Chapter 9, HIV and AIDS)

3. DIAGNOSTICS AND PROGNOSTICS

The concept of using gene expression profiles for diagnostic purposes was first proposed in cancer biology, when it was

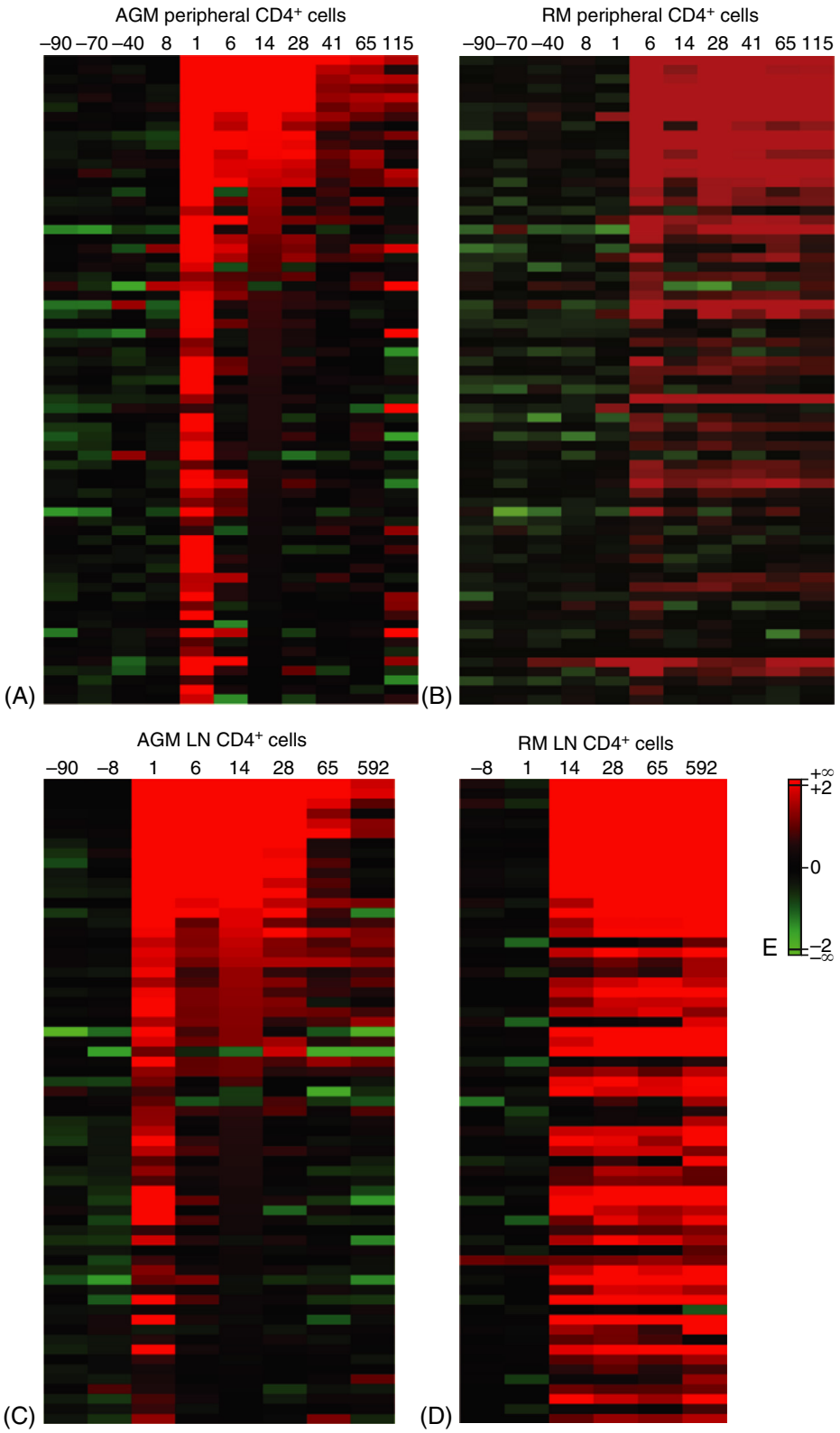


FIGURE 3 Type I interferon-stimulated gene (ISG) expression in blood and lymph node CD4+ cells obtained from monkeys infected with simian immunodeficiency virus. The genes of the ISG cluster that were significantly regulated ($P < 0.05$) in at least one of the two species are represented as heat maps. Mean values of the log2Q of type I ISG expression in peripheral (A and B, respectively) and lymph node (LN) CD4+ cells (C and D, respectively) from six African green monkeys (AGM) and six rhesus macaques (RM) are shown. Many of these genes can also be induced by type II interferon. (E) The color scheme indicates the log2Q. Red indicates up-regulation and green indicates down-regulation. *Adapted from Jacquelin et al. (2009), with permission.*

Sidebar 3 Systems biology in cancer research

Many of the systems approaches that are being used in virology were pioneered in the field of cancer research. Cancer biologists were quick to embrace systems biology as a way to understand how genetic and epigenetic aberrations perturb intracellular signaling networks thereby leading to carcinogenesis. Systems approaches have already yielded clinical benefits to patients, for example, by giving rise to new classification schemes for breast and pancreatic cancer, which in turn allow clinicians to identify patients most likely to benefit from a particular therapy. The cancer research field has been aggressive in developing large-scale systems biology resources, such as The Cancer Genome Atlas (TCGA), the International Cancer Genome Consortium, and The Cancer Proteome Atlas. The field is also driving the development of new mathematical and modeling methods for studying the relationship between intracellular signaling and the behavior of cells at the tissue level, and for the imaging of cells to determine how their spatial orientation and interactions with the environment interplay with gene expression patterns and tumor behavior. Many of these new approaches will no doubt make their way into virology research.

It should be noted, however, that cancer research also draws on virology. A good example is the use of DNA tumor virus proteins for cancer gene discovery (Rozenblatt-Rosen et al., 2012). In this study, proteins from four types of tumor viruses, papillomavirus, Epstein–Barr virus, adenovirus, and polyomavirus, were tested for their ability to interact with 13,000 human gene products. The effect of expressing individual viral genes in cell culture was also assessed by microarray analysis. These data were used to build a virus–host perturbation network that reveals genes and pathways commonly affected by the tumor virus proteins and that are likely to contribute to cancer.

discovered that breast tumor transcriptional profiles could be used for tumor classification. Because different viruses elicit different host transcriptional responses, it might also be expected that gene expression profiles can be used for diagnostic purposes or for making predictions about infection or therapeutic outcome. In clinical virology, most diagnostic tests rely on the ability to detect a particular virus in the bloodstream, either through detection of viral antigens or viral genome sequences. In contrast, host genomic signatures can be used to detect unknown as well as known viruses, or as in cancer, be used as guides for choosing therapy or as predictors of therapeutic or disease outcome (Sidebar 3).

3.1 Genomic Markers of Virus-Induced Liver Disease

One area in which the search for host diagnostic markers has been particularly intense is in hepatitis virus-induced liver disease and hepatocellular carcinoma. Liver disease caused by hepatitis C virus (HCV), for example, can take decades to develop and is often asymptomatic until disease has progressed to cirrhosis or hepatocellular carcinoma. At that point, transplantation may be the only remaining treatment option. Diagnosis is additionally hampered by the lack of reliable noninvasive detection methods as an alternative to percutaneous liver biopsy. A biomarker or genomic profile that could be detected in the blood is therefore highly desirable for early diagnosis of liver disease.

A variety of individual candidate blood biomarkers have been identified, such as aspartate and alanine aminotransferases, albumin, and alkaline phosphatase. These biomarkers provide information about liver function, but values can sometimes be normal in people with liver disease or damage, and these tests do not provide information about disease etiology. Unfortunately, diagnostic genomic signatures in blood have been hard to come by, and studies have therefore focused on the identification of such markers in liver tissue. Although profiling liver gene expression does not get around the need for a liver biopsy, the hope is that genomic profiles will be predictive in advance of tissue injury. There have been a number of reports of liver gene expression signatures, ranging from a half dozen to several hundred genes, which may be prognostic for hepatocellular carcinoma and that may be useful for risk-adjusted surveillance approaches (Hannivoort et al., 2012).

Progress in identifying genomic markers for the diagnosis and prognosis of HCV-induced liver disease has been hampered by the length of time from infection to disease, which makes prospective studies difficult. There is a group of patients, however, in which this time period is compressed. When individuals receive a liver transplant because of end-stage liver disease caused by HCV infection, the transplanted liver quickly becomes infected. While many transplant recipients show no biochemical or histological evidence of liver injury in the first 10 years after transplantation, approximately one-third develop rapidly progressive fibrosis, with the onset of cirrhosis occurring in as little as 5 years after transplantation.

Liver transplant recipients therefore represent a unique study population for discovering gene expression changes that may be predictive of disease progression. In what is perhaps the largest study focused on this population, advanced computational approaches, such as singular value decomposition-multidimensional scaling, were used to analyze transcriptional data obtained from serial liver biopsies from 57 patients (Rasmussen et al., 2012). This study revealed that within 3 months of transplantation over 400 genes were differentially expressed between progressors (who developed adverse clinical outcomes 4–7 years after transplantation)

and nonprogressors. This included the down-regulation of genes associated with immune and inflammatory responses, cell cycle progression, and metabolic functions in patients who progressed to severe liver disease. A companion study identified a proteomic signature, indicative of oxidative stress, which could also distinguish progressors from nonprogressors (Diamond et al., 2012). Significantly, these studies showed that transcriptional or proteomic markers of disease progression can be detected prior to histological evidence of severe liver injury. Such markers may therefore eventually be the basis for a diagnostic test that can identify patients at high risk of disease progression after transplantation, and perhaps more broadly outside of the transplant setting.

3.2 Discrimination between Viral and Bacterial Respiratory Infection

The etiologic diagnosis of respiratory infections is challenging. Moreover, because many of these infections are caused by bacterial pathogens, physicians will often treat patients with antibiotics even without a confirmed diagnosis in an attempt to provide speedy resolution of symptoms. It is therefore important to develop a means to rapidly distinguish between viral and bacterial infections as well as to identify the specific etiologic agent. Even though respiratory pathogens are typically confined to the respiratory tract, there is growing evidence that different immune cell types induce gene expression signatures in the blood that may be used to accurately diagnose acute respiratory viral infection.

In an initial study aimed at identifying such signatures, volunteers were experimentally infected with rhinovirus, respiratory syncytial virus, or influenza virus and blood samples were taken at set intervals following challenge (Zaas et al., 2009). Microarray analysis of blood gene expression patterns identified a 30-gene "acute respiratory viral" signature that was common to symptomatic individuals from all three viral challenges and which could distinguish between symptomatic individuals and uninfected controls. The signature could also accurately distinguish persons with influenza A virus infection from healthy controls in an independent community-based cohort. In addition, when used to analyze publicly available gene expression data from the blood of patients with bacterial respiratory infection, the signature can accurately distinguish viral from bacterial infection.

More recently, this same 30-gene set has been incorporated into a reverse transcription polymerase chain reaction (RT-PCR) assay (Zaas et al., 2013). RT-PCR is an established diagnostic platform, and moving the acute respiratory signature to this platform represents an important step toward eventual clinical use. The assay was tested in a cohort of 102 individuals arriving at an emergency room

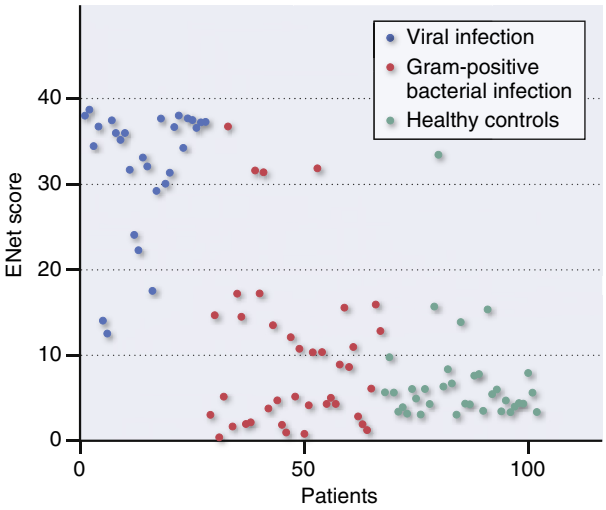


FIGURE 4 A reverse transcription polymerase chain reaction gene expression classifier accurately classifies individuals presenting to the emergency department with viral infection (blue) and distinguishes them from those presenting with Gram-positive bacterial infection (red) and healthy controls (green). ENet score (*y* axis) is a measure of the probability of having viral infection, with a score of 20 indicative of 50% probability of detection. *Adapted from Zass et al. (2013), with permission.*

with fever, and who were confirmed by standard microbiological assays to have a viral or bacterial infection. The RT-PCR assay showed 94% accuracy in distinguishing viral from bacterial infections, suggesting that measuring the expression of a small set of genes in blood samples can be used to classify viral respiratory illness in a real-world setting (Figure 4).

The stage is therefore set for using patient gene expression signatures in viral diagnostics and prognostics. However, there is still work to be done in terms of identifying the most appropriate and minimal set of signatures for these assays, and for improving specificity to provide diagnosis of specific viral agents. Nevertheless, because these approaches provide additional and complementary information to that provided by microbiological assays, gene expression profiling does not need to be considered as a substitute for current diagnostic methods. Instead, a combined approach is likely to yield benefits in terms of rapid triage, the evaluation of febrile illnesses without clear etiology, and for understanding disease pathogenesis.

Unfortunately, in addition to scientific challenges, the use of genomic profiles in clinical settings is facing increased regulatory hurdles as well. This comes primarily in response to the use of faulty (possibly fraudulent) genomic markers to select therapy for patients enrolled in a clinical trial to test alternative chemotherapy approaches to treat nonsmall-cell lung cancer (Kurzrock et al., 2014). In the aftermath to this trial, the Institute of Medicine (an arm of the National Academy of Sciences that provides advice to policy makers) has recommended that diagnostic tests that use genomic data be viewed as devices rather than as laboratory-based tests.

This designation requires that genomic-based diagnostics be subject to additional regulatory controls and be overseen by the Food and Drug Administration (FDA). These increased regulatory burdens are likely to result in increased development costs and delays in testing and implementation.

4. HOST-RESPONSE NETWORKS

Diagnostics and prognostics can be developed on the basis of gene expression signatures. So long as such signatures are accurate predictors, they do not necessarily have to impart any insight into the underlying mechanisms of viral pathogenesis (though often they do). Similarly, an examination of heat maps and functional annotations can on their own yield considerable information about the host response to infection. To use large-scale transcriptional information to gain insight into how things work at a molecular level, or to identify regulatory or drug targets, it is necessary to organize this information in different ways to look for interrelationships among the data. One of the most useful ways to visualize these relationships is in the form of biological networks. Such networks can be built using gene coexpression, the direct interaction of encoded proteins, or shared regulatory mechanisms, such as the binding of transcription factors to target genes.

4.1 Network Hubs and Bottlenecks

Biological networks are typically represented by graphs that contain nodes (genes or proteins) and lines (referred to as edges) connecting the nodes. In a gene coexpression network, for example, the edges are determined using statistical measures of expression correlation. The resulting network can then be analyzed using various computational methods to link network topology—the arrangement and connections of the components of a network—with biological properties.

One of the most common methods used to identify important elements of a network is centrality analysis. Many of the concepts of centrality analysis were first developed for analyzing social networks, and indeed, biological and social networks have many similarities. When analyzing a social network, centrality analysis might be used to determine the most influential person in the network. When analyzing a biological network, the same types of analyses can be used to identify key players in biological processes.

Centrality analysis is used to look for nodes in the network that are highly interconnected or that regulate (or restrict) the flow of information. In biological networks, genes that are highly interconnected—referred to as hubs—are often functionally important. Similarly, genes that connect or bridge multiple subnetworks—referred to as bottleneck genes—are positioned to play powerful roles in regulating network signaling even though they may have fewer connections than hub genes. Hub and bottleneck genes have both been shown to be significantly more likely to be essential for microbial virulence than their nonbottleneck or nonhub counterparts.

These same methods can be used to analyze networks derived from proteomic and lipidomic profiling. In such an analysis of HCV-infected hepatoma cells, two mitochondrial fatty acid oxidation enzymes, DCI (Enoyl-CoA Delta Isomerase 1)and HADHB (Hydroxyacyl-CoA Dehydrogenase/3-Ketoacyl-CoA Thiolase/Enoyl-CoA Hydratase), were identified as network bottlenecks and possible points of control through which HCV disrupts cellular metabolic homeostasis (Figure 5) (Diamond et al., 2010). Targeting DCI for knockdown by siRNA techniques subsequently showed that DCI is required for productive HCV infection in cultured cells. In this case, network analysis provided the information needed to sort through large numbers of proteins and to focus in on a specific target for further scrutiny. Changes in fatty acid oxidation have also been more broadly linked to inflammatory processes, dendritic cell maturation, and regulation of the immune response, suggesting that modulation of fatty acid oxidation may represent a target for antiviral therapy.

4.2 Targets for Host-Directed Antiviral Therapies

An understanding of network topology therefore provides the opportunity to identify potential host targets for therapeutic intervention. Most current antiviral drugs are directed against specific viral protein targets. Such drugs are narrow in spectrum, meaning they are effective only against a specific virus, and they are vulnerable to the emergence of viral resistance (through mutation of the viral genome). Moreover, most medically important viruses have small genomes, so the number of potential targets is limited. In contrast, treating viral infection by targeting the host increases the number of targets and decreases the likelihood that resistant viruses will emerge. Targeting host factors may also increase the likelihood that the drug will be effective against a wider spectrum of viral pathogens.

The analysis of coexpression networks using approaches such as centrality analysis is one way to identify potential host targets (such as DCI in the example above). However, combining network analysis with other information, such as protein–protein interaction data, may improve target identification. This was demonstrated using proteomic data derived from HCV-infected hepatoma cells and from liver tissue obtained from HCV-infected patients (McDermott et al., 2012). When protein–protein interaction data (e.g., host proteins known to interact with HCV proteins) were integrated into protein coabundance networks, network topology analysis of the integrated network provided improved discrimination of bottleneck and hub proteins.

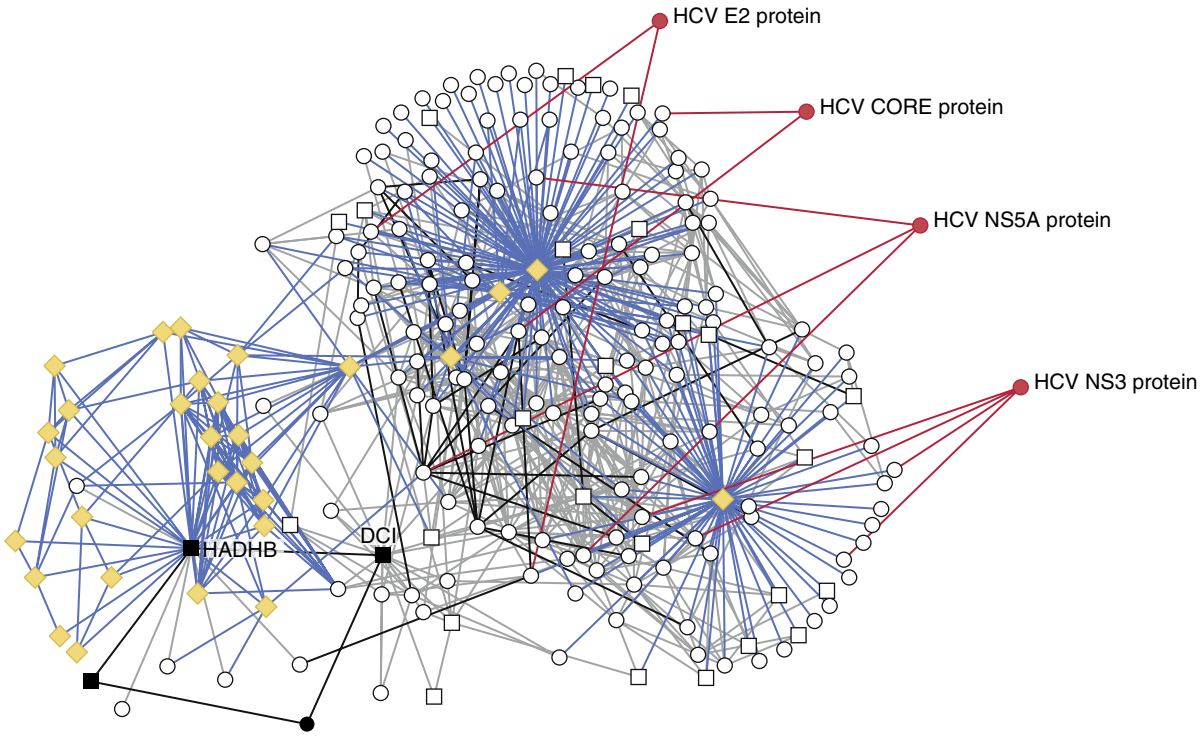


FIGURE 5 The integrated network surrounding several key bottlenecks identified by computational modeling. The neighbors of bottlenecks in the integrated network are shown. Relationships between the proteins and lipid species are gray for proteomics correlation, purple for lipidomics–proteomics correlation, black for protein–protein interactions, and red for interactions with viral proteins. Lipid species are indicated as yellow diamonds, HCV proteins are red, mitochondrial proteins are squares, and proteins involved in fatty acid β-oxidation are in black. *Adapted from Diamond et al. (2010), with permission.*

A similar analysis looked directly at host factors that interact with the HCV-encoded NS5A protein (Tripathi et al., 2013). Through its interaction with other HCV proteins and host factors, NS5A plays an important role in HCV infection, including regulating viral replication, the production of viral particles, and interferon resistance. NS5A has therefore become an attractive target for antiviral therapy. However, rather than targeting NS5A directly, it may be possible to target one or more host factors that interact with NS5A. To identify such host targets, NS5A-interacting proteins were identified through literature mining and by using a yeast two-hybrid approach (a molecular biology technique used to discover protein–protein interactions). The resulting 132 host proteins were used to build an interaction network, which was further expanded by incorporating protein–protein interactions for the proteins targeted by NS5A (resulting in 1442 proteins with 6263 interactions between them). Topological analysis was then used to identify bottleneck and hub proteins.

Intriguingly, these analyses suggest that NS5A preferentially interacts with highly central proteins in the host protein interaction network. These proteins have functions in a variety of cellular processes, including innate immunity, chemokine signaling, cell-to-cell communication, and cellular transport. Among the bottleneck proteins identified were two endoplasmic reticulum proteins, RTN1 and RTN3. These proteins are present in very low density lipoprotein transport vesicles, which have been reported to play a role in the production and release of infectious HCV. In cell culture assays, knockdown of RTN1 and RTN3 using siRNA has no effect on viral RNA levels, but significantly reduces viral titer. As regulators of viral propagation, RTN1 and RTN3 may be novel targets for anti-HCV therapy and perhaps more broadly as therapeutic targets for other viruses that depend upon lipoprotein vesicles for the release of infectious virus.

5. DRUG REPURPOSING

Developing new drugs takes an enormous amount of time, averaging 14 years from target discovery to FDA approval. Failure rates and costs are also extraordinarily high. So even once targets are identified and validated, a long road to drug development remains. One strategy to reduce the time and expense of drug development is to determine whether a drug approved to treat one disease might be repurposed to treat another. Similarly, many partially developed drug candidates could potentially be repurposed for new indications. Global transcriptomic methods have become central to drug repurposing efforts (Hurle et al., 2013), and the paradigm is finding its way into antiviral drug research.

5.1 Inverse Genomic Signatures

The inverse genomic signature approach is based on the proposition that a drug should have therapeutic benefit if it generates a gene expression profile that is the inverse of the signature associated with the disease (Figure 6; Peng et al., 2014). The approach therefore requires knowledge of the gene expression profiles induced by large numbers of drugs. Such information is being generated by the Connectivity Map project, which seeks to find connections between human diseases, gene expression profiles, and drug action. The Connectivity Map database contains over 7000 transcriptional profiles generated by treating cultured human cells with over 1300 compounds, many of which are FDA-approved drugs. Analytical tools can then be used to calculate a connectivity score, a measure of the similarity—or inverse similarity—between query gene expression signatures and profiles in the database.

When this approach was first used to identify drugs that inhibit influenza virus replication, a common 20-gene expression signature was identified from cultured human lung epithelial cells infected with different strains of human or avian influenza virus. This gene expression signature was then used to screen drug-associated profiles in the Connectivity Map database, and candidate antivirals were identified by their inverse correlation to the common signature. Eight potential antivirals were identified, six of which were subsequently determined to inhibit influenza virus replication, including the 2009 H1N1 pandemic influenza virus, which was not used to generate the 20-gene signature (Josset et al., 2010).

Because the use of genomic signatures for drug screening has the potential to dramatically reduce the time needed for drug development, the approach may be particularly beneficial when applied to emerging viral infections. Several recent studies have used the approach to screen for drugs against emerging influenza virus strains (e.g., H7N9 viruses), and Connectivity Map has also been used to identify drugs that may be effective against Middle East respiratory syndrome coronavirus (Josset et al., 2013). In all of these studies, follow-up validation using cell culture has demonstrated reductions in viral replication. However, it remains to be seen whether and how rapidly drugs (or classes of drugs) identified by this method actually make their way into clinical use. The same can be said for the cancer field, in which drug repurposing approaches using Connectivity Map have identified drug candidates against a variety of cancers. Many of these drugs have been validated in cell culture and rodent models, but have been slow to move into clinical trials.

5.2 Network-Based Approaches

As discussed earlier, networks can be used to represent regulatory and functional interactions between genes or proteins, or between combinations of genes, proteins, or metabolites. One goal of generating such networks is to identify *targets* for antiviral drugs. In contrast, network-based drug repurposing strategies aim to harness the information contained in networks to identify candidate *drugs*. Although network-based approaches can take a variety of forms, many are focused on understanding how diseases are connected to one another—through similar transcriptional or protein–protein interaction networks, for example—or

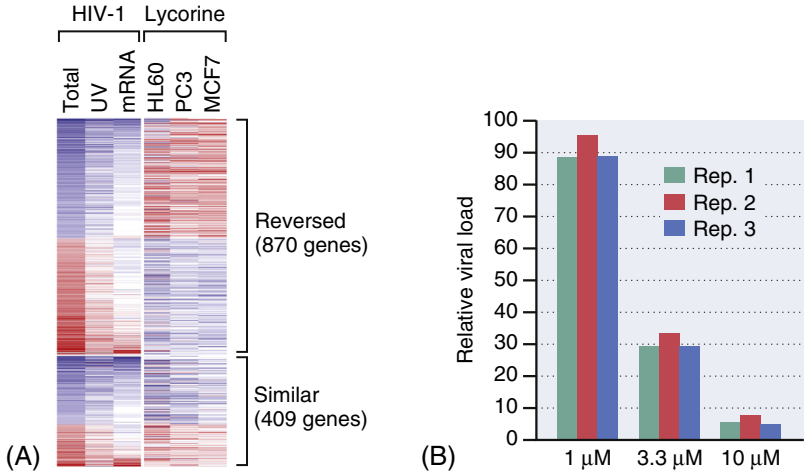


FIGURE 6 Side by side comparison of gene expression changes induced by HIV-1 infection of a human CD4+ T cell line (SUP-T1) and by treatment of three different cell lines with the drug lycorine. Gene expression-based antiviral drug repurposing predicts that a drug should have therapeutic benefit if it generates a gene expression profile that is the inverse of the signature associated with viral infection. (A) Columns represent individual conditions: Total RNA-seq of HIV-1 infection (Total) or UV-inactivated virion infection (UV), mRNA-seq of HIV-1 infection (mRNA), and lycorine treatment of three different cell lines. Lycorine-induced expression profiles were obtained from Connectivity Map. (B) Relative viral loads (quantified by qPCR) in HIV-1-infected SUP-T1 cells treated with lycorine showing inhibition of HIV-1 replication. *Adapted from Peng et al. (2014), with permission.*

how drugs are connected to one another through their mechanism of action. For example, Connectivity Map can also be used to construct "drug networks" to identify connections between drugs on the basis of shared transcriptional effects. In this case, different drugs (rather than genes or proteins) form the nodes of the network. Drugs can also be connected by side-effect similarity, which can be used to infer whether two drugs share a target.

Various networks—transcriptional, protein–protein interaction, disease, or drug—can also be integrated into multilayer networks that can be probed for drug–disease relationships. For example, an integrated analysis of gene expression data from 54 diseases (including several viral diseases) and human protein–protein interaction data yielded a network of 138 disease relationships (Suthram et al., 2010). Within this integrated network, a set of common pathways was identified, and many of the proteins in those pathways were found to be targets of existing drugs. By identifying disease relationships and shared drugs, it was possible to make predictions about the repurposing of drugs from one disease to another. With the ever-increasing amounts of genomic data being generated in infectious disease research, network-based approaches are sure to be increasingly exploited for the repurposing of drugs to fight viral infections.

6. NEW VIEWS OF THE TRANSCRIPTIONAL LANDSCAPE: LONG NONCODING RNAs AND VIRAL INFECTION

The direct sequencing of RNA transcripts is yielding exciting new views of the transcriptional landscape. The staggering amount of information available through sequencing is exemplified by the Encyclopedia of DNA Elements (ENCODE) project, funded by the National Human Genome Research Institute. The overall goal of the project is to identify and characterize all functional elements in the human genome, including cataloging of the complete repertoire of RNAs produced by human cells. The project has revealed that as much as three quarters of the human genome is capable of being transcribed and that cells contain many varieties of RNA transcripts (Djebali et al., 2012). These include polyadenylated and nonpolyadenylated transcripts, known and unannotated protein-coding transcripts, and long noncoding RNAs (>200 nucleotides) such as intergenic transcripts. Small noncoding RNAs (<200 nucleotides) are also abundant, including microRNAs (miRNAs), piwi-interacting RNAs (piRNAs), small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), and transfer RNAs. Indeed the transcriptional landscape is so complex that it has called into question the very definition of a gene!

Virologists are beginning to explore the ramifications of this newfound transcriptional complexity on viral pathogenesis. In one of the first studies to do so, RNA-seq was used to profile the host response to infection with severe acute respiratory syndrome coronavirus (SARS-CoV) (Peng et al., 2010). In the lungs of mice infected with this virus, over 10,000-long noncoding RNAs were identified, and nearly 1500 were differentially expressed in response to infection. Comparable expression profiles were observed in cell lines infected with influenza virus or treated with type I interferon, suggesting that these RNAs may be involved in the innate response to a variety of viruses. Similar analyses that focused on the sequencing of small RNAs revealed that SARS-CoV infection also induces the differential expression of different classes of small noncoding RNAs.

Distinct patterns of noncoding RNA expression have now been observed in the response to many different RNA and DNA viruses. Moreover, long noncoding RNAs are not limited to cellular transcription. For example, RNA-seq has revealed that human cytomegalovirus (HCMV), a 240-kb DNA virus, produces hundreds of previously unidentified transcripts, including alternatively spliced transcripts and long noncoding RNAs (Gatherer et al., 2011). HCMV is also one of several viruses known to encode its own miRNAs, and RNA-seq has further revealed that HCMV encodes additional novel forms of small RNAs (Stark et al., 2012). Even viral genomes are capable of producing complex transcriptional profiles.

The next important step, of course, is to determine the biological functions of these newly found long noncoding RNAs. There is already evidence that long noncoding RNAs play roles in transcriptional and epigenetic gene regulation, developmental processes, and in a variety of diseases, including neurological and immune disorders and cancer. These RNAs may function through direct interaction with specific genome sequences (thereby affecting chromatin remodeling and gene transcription), transcription factors, or other components of the transcriptional machinery. For example, the long noncoding RNA, lincRNA-Cox2, has been demonstrated to interact with heterogeneous nuclear ribonucleoproteins to mediate both the activation and repression of multiple immune response genes (Carpenter et al., 2013). Similarly, lncRNA-CMPK2 has been reported to be a negative regulator of the interferon response and is itself among over 200 long noncoding RNAs induced by interferon (Figure 7; Kambara et al., 2014). Knockdown of lncRNA-CMPK2 expression reduces HCV replication in interferon-treated hepatocytes, and lncRNA-CMPK2 is up-regulated in liver samples from HCV-infected patients, suggesting it may also play a role in modulating the interferon response in these individuals.

Unfortunately, determining the biological functions of long noncoding RNAs has so far proven challenging.

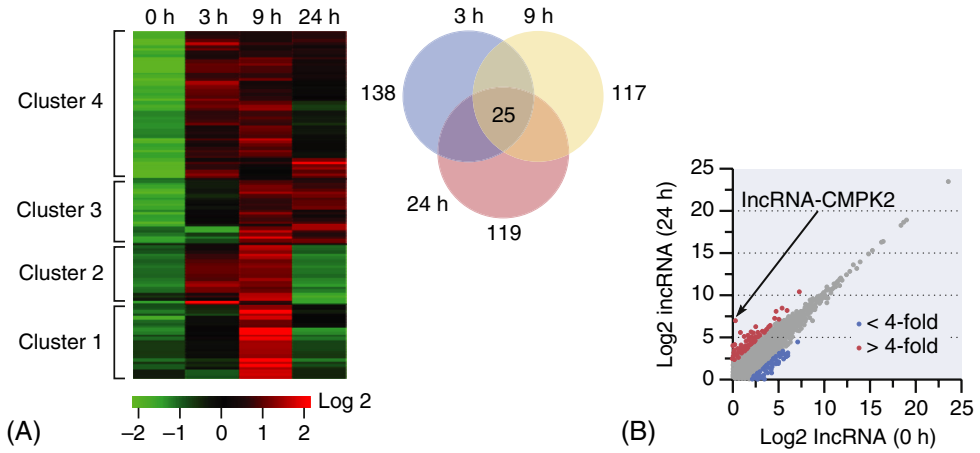


FIGURE 7 Interferon-α induces hundreds of long noncoding RNAs in human hepatocytes. (A) Heat map and Venn diagram of lncRNAs showing fourfold or greater change in expression following interferon-α treatment. (B) Scatter plot depicting the annotated lncRNAs that show a statistically significant change of fourfold or more after 24 h of interferon stimulation. The gray dots mark lncRNAs that did not show a significant change in expression. Red and blue dots correspond to up-regulated and down-regulated lncRNAs, respectively. The location of lncRNA-CMPK2, a negative regulator of the interferon response, is shown. *Adapted from Kambara et al. (2014), with permission.*

Unlike proteins, the function of long noncoding RNAs cannot presently be predicted from primary sequence or secondary structure. And with thousands of such RNAs to choose from, it is difficult to know which to pursue. In the example above, lncRNA-CMPK2 was chosen for study because of its over 100-fold induction in response to interferon stimulation. However, long noncoding RNA expression can also be integrated into network models, which may help to provide additional information as to which RNAs should be the focus for follow-up studies. With the advent of new genome editing techniques, such as CRISPR-Cas9, it will also be possible to design large-scale screens to identify long noncoding RNAs that may be required for viral replication or virus-induced cytopathic effects (for details on CRISPR-Cas9 and large-scale interaction screening, see Chapter 12, The Virus–Host Interactome).

Unquestionably, RNA-seq and the discovery of long noncoding RNAs have ushered in a new era in the study of viral pathogenesis. It will be essential to gain an understanding of the role of these RNAs during viral infection to fully understand the mechanisms by which viruses cause disease. It is also likely that a better understanding of long noncoding RNA function will lead to new therapeutic options. RNA-based therapeutics—most of which are focused on the use of antisense oligonucleotides to degrade specific mRNAs—are already in development. Indeed, an antiviral drug, fomivirsen (an antisense oligonucleotide that blocks the synthesis of a key cytomegalovirus protein), was the first drug of this type to be approved by the FDA. Antisense transcripts could be similarly used to deplete specific long noncoding RNAs, or small molecules could be used to disrupt the interaction of long noncoding RNAs with their protein or DNA partners.

7. CONCLUDING REMARKS

Ongoing advances in technology will continue to spur new approaches for studying the host transcriptional response to viral infection. While most current studies using animal models still examine the RNA profiles of whole tissues, it is becoming increasingly common to augment this approach with the analysis of isolated cell populations. But even the analysis of isolated cell types results in averaging the transcriptomes of millions of cells. RNA-seq performed on single cells is therefore emerging as a new frontier in transcriptional profiling. The ability to perform single-cell analyses has become possible through improvements to methods for cell isolation and the conversion of miniscule amounts of cellular RNA into cDNA for sequencing. Although single-cell analyses are not necessary (or appropriate) for all types of studies, the capability provides unique opportunities, such as profiling cell-to-cell variability. Such analyses have revealed surprising variability in the expression of hundreds of immune genes across single cells, as well as variation in splicing patterns (Shalek et al., 2013).

Efforts are also underway to incorporate an understanding of how the epigenome—the heritable, and potentially reversible, genome-wide chemical changes to the DNA and histone proteins of an organism—impact the host response to viral infection. Epigenetic modifications, such as DNA or histone methylation, result in changes in gene expression through alterations in chromatin structure. These modifications can occur in response to a multitude of environmental

changes, including viral infection. In addition, several viruses, such as human cytomegalovirus and Epstein–Barr virus, use epigenetic mechanisms in part for switching between latent and active infection (Ernberg et al., 2012). Although studying the involvement of epigenetic mechanisms in viral infection is still in its infancy, such studies are likely to offer new clues into disease mechanisms.

Clearly, high-throughput molecular profiling has revealed that the host response to viral infection is more complicated than ever before thought possible. Moreover, gene expression, protein abundance, and all of their attendant regulatory mechanisms are only part of the picture. As discussed in other chapters, consideration must also be given to protein–protein interactions, host metabolism, host genetics, and even the myriad microorganisms that form the human microbiome. With each new technological advance comes a new avalanche of data to a field that some argue is already suffering from information overload. Perhaps what is needed most are improved computational methods for integrating diverse types of data and for identifying new interrelationships, including cooperative or synergistic interactions. These may arrive in the form of new geometric approaches and links between geometry, information theory, and probability theory (Law et al., 2013). In the meantime, even if the complete picture will have to wait to be assembled, there are plenty of discoveries to come to keep virologists energized well into the future.

REFERENCES

Benecke A, Gale Jr M, Katze MG. Dynamics of innate immunity are key to chronic immune activation in AIDS. Curr Opin HIV AIDS 2012;7:79–85.

Brandes M, Klauschen F, Kuchen S, et al. A systems analysis identifies a feedforward inflammatory circuit leading to lethal influenza infection. Cell 2013;154:197–212.

Carpenter S, Aiello D, Atianand MK, et al. A long noncoding RNA mediates both activation and repression of immune response genes. Science 2013;341:789–92.

Cilloniz C, Pantin-Jackwood MJ, Ni C, et al. Molecular signatures associated with Mx1-mediated resistance to highly pathogenic influenza virus infection: mechanisms of survival. J Virol 2012;86:2437–46.

Diamond DL, Krasnoselsky AL, Burnum KE, 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.

Diamond DL, Syder AJ, Jacobs JM, et al. Temporal proteome and lipidome profiles reveal hepatitis C virus-associated reprogramming of hepatocellular metabolism and bioenergetics. PLoS Pathog 2010;6:e1000719.

Djebali S, Davis CA, Merkel A, et al. Landscape of transcription in human cells. Nature 2012;489:101–8.

Ernberg I, Karimi M, Ekstrom TJ. Epigenetic mechanisms as targets and companions of viral assaults. Ann N Y Acad Sci 2012;1230:E29–36.

Gatherer D, Seirafian S, Cunningham C, et al. High-resolution human cytomegalovirus transcriptome. Proc Natl Acad Sci USA 2011;108:19755–60.

Hannivoort RA, Hernandez-Gea V, Friedman SL. Genomics and proteomics in liver fibrosis and cirrhosis. Fibrogenesis Tissue Repair 2012;5:1.

Hurle MR, Yang L, Xie Q, et al. Computational drug repositioning: from data to therapeutics. Clin Pharmacol Ther 2013;93:335–41.

Ideker T, Galitski T, Hood L. A new approach to decoding life: systems biology. Annu Rev Genomics Hum Genet 2001;2:343–72.

Jacquelin B, Mayau V, Targat B, et al. Nonpathogenic SIV infection of African green monkeys induces a strong but rapidly controlled type I IFN response. J Clin Invest 2009;119:3544–55.

Josset L, Menachery VD, Gralinski LE, et al. Cell host response to infection with novel human coronavirus EMC predicts potential antivirals and important differences with SARS coronavirus. mBio 2013;4:e00165–13.

Josset L, Textoris J, Loriod B, et al. Gene expression signature-based screening identifies new broadly effective influenza A antivirals. PLoS One 2010;5.

Kambara H, Niazi F, Kostadinova L, et al. Negative regulation of the interferon response by an interferon-induced long non-coding RNA. Nucleic Acids Res 2014;42(16):10668–80.

Kurzrock R, Kantarjian H, Stewart DJ. A cancer trial scandal and its regulatory backlash. Nat Biotechnol 2014;32:27–31.

Law GL, Korth MJ, Benecke AG, et al. Systems virology: host-directed approaches to viral pathogenesis and drug targeting. Nat Rev Micro 2013;11:455–66.

McDermott JE, Diamond DL, Corley C, 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.

Peng X, Gralinski L, Armour CD, et al. Unique signatures of long noncoding RNA expression in response to virus infection and altered innate immune signaling. mBio 2010;1:e00206–10.

Peng X, Sova P, Green RR, et al. Deep sequencing of HIV infected cells: insights into nascent transcription and host-directed therapy. J Virol 2014;88:8768–82.

Rasmussen AL, Tchitchek N, Susnow NJ, et al. Early transcriptional programming links progression to hepatitis C virus-induced severe liver disease in transplant patients. Hepatology 2012;56:17–27.

Rosen R. A means toward a new holism. Science 1968;161:34–5.

Rozenblatt-Rosen O, Deo RC, Padi M, et al. Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. Nature 2012;487:491–5.

Schena M, Shalon D, Davis RW, et al. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995;270:467–70.

Shalek AK, Satija R, Adiconis X, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature 2013;498:236–40.

Stark TJ, Arnold JD, Spector DH, et al. High-resolution profiling and analysis of viral and host small RNAs during human cytomegalovirus infection. J Virol 2012;86:226–35.

Suthram S, Dudley JT, Chiang AP, et al. Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. PLoS Comput Biol 2010;6:e1000662.

Tchitchek N, Eisfeld AJ, Tisoncik-Go J, et al. Specific mutations in H5N1 mainly impact the magnitude and velocity of the host response in mice. BMC Syst Biol 2013;7:69.

Trewavas A. A brief history of systems biology. Plant Cell 2006;18:2420–30.

Tripathi LP, Kambara H, Chen YA, et al. Understanding the biological context of NS5A-host interactions in HCV infection: a network-based approach. J Proteome Res 2013;12:2537–51.

Zaas AK, Burke T, Chen M, et al. A host-based RT-PCR gene expression signature to identify acute respiratory viral infection. Sci Transl Med 2013;5:203ra126.

Zaas AK, Chen M, Varkey J, et al. Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans. Cell Host Microbe 2009;6:207–17.

Zhu H, Cong JP, Mamtora G, et al. Cellular gene expression altered by human cytomegalovirus: global monitoring with oligonucleotide arrays. Proc Natl Acad Sci USA 1998;95:14470–5.

FURTHER READING

Csermely P, Korcsmaros T, Kiss HJ, et al. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. Pharmacol Ther 2013; 138: 333–408.

Korth MJ, Tchitchek N, Benecke AG, et al. Systems approaches to influenza-virus host interactions and the pathogenesis of highly virulent and pandemic viruses. Semin Immunol 2013; 25: 228–39.

Menachery VD, Baric RS. Bugs in the system. Immunol Rev 2013; 255(1): 256–74.

Tisoncik JR, Korth MJ, Simmons CP, et al. Into the eye of the cytokine storm. Microbiol Molec Biol Rev 2012; 76(1): 16–32.