



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.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

# Lost in the world of functional genomics, systems biology, and translational research: Is there life after the Milstein award?

Michael G. Katze<sup>\*</sup>, Marcus J. Korth

*Department of Microbiology and Washington National Primate Research Center, University of Washington, Seattle, WA 98195-8070, USA*

Available online 27 July 2007

## Abstract

We have always wanted to save the world from the scourges of virus infection by developing better drugs and vaccines. But fully understanding the intricacies of virus–host interactions, the first step in achieving this goal, requires the ability to view the process on a grand scale. The advent of high-throughput technologies, such as DNA microarrays and mass spectrometry, provided the first opportunities to obtain such a view. Here, we describe our efforts to use these tools to focus on the changes in cellular gene expression and protein abundance that occur in response to virus infection. By examining these changes in a comprehensive manner, we have been able to discover exciting new insights into innate immunity, interferon and cytokine signaling, and the strategies used by viruses to overcome these cellular defenses. Functional genomics may yet save the world from killer viruses.

© 2007 Elsevier Ltd. All rights reserved.

*Keywords:* Genomics; Hepatitis C; Influenza; Microarray; Proteomics

## 1. Introduction

At the time I received the Milstein award in September 1999, my laboratory was embarking on an exciting new venture. I had become fascinated by the novel opportunities that were becoming available through the sequencing of the human genome and the advent of high-throughput technologies such as DNA microarrays and mass spectrometry [1]. Because these technologies provided a way to obtain a nearly comprehensive view of a biological system, I was convinced they offered the best chance to learn everything there is to know about the complex interplay between viruses and the cells they infect. Over the past eight years, my laboratory has worked tirelessly, and even obsessively, to incorporate genomic, proteomic, and information technologies into our research. Although we certainly do not yet know everything there is to know about virus–host interactions, we have made

tremendous progress, and I remain convinced that genomic and proteomic approaches will be a fundamental force for pushing the field of virology forward in the years to come. In this paper, we provide a sampling of how we are using these technologies, the discoveries we have made, and how the technologies themselves have continued to evolve.

## 2. Dark ages of discovery

Our initial foray into the world of genomics began with cDNA microarrays that were spotted in-house using a commercially available collection of cDNA clones. Getting this new technology up and running required months of work. But after seemingly endless frustration, Gary Geiss was finally able to use these arrays, representing all of 1500 human genes, to evaluate the changes in cellular gene expression that occurred in a CD4<sup>+</sup> T-cell line infected with human immunodeficiency virus type 1 (HIV-1) [2]. This was a straightforward study in which we analyzed two time points after infection, and we were thrilled to observe the temporal regulation of genes involved in T-cell signaling,

<sup>\*</sup> Corresponding author at: Department of Microbiology, University of Washington, Box 358070, Seattle, WA 98195-8070, USA.  
Tel.: +1 206 732 6136; fax: +1 206 732 6055.

*E-mail address:* [honey@u.washington.edu](mailto:honey@u.washington.edu) (M.G. Katze).

subcellular trafficking, and transcriptional regulation. Despite how simplistic this study now appears, it was the first ever to apply genomics to AIDS research and it was the cover article when published in *Virology*. With the technology now in place, we were eager to apply it to a broad range of experimental systems. The possibilities seemed endless (as indeed they are), and we rapidly began “doing arrays” on cells infected with a variety of different viruses or exposed to double-stranded RNA (dsRNA), interferon, or other treatments (Fig. 1).

Our next study was a quantum leap in sophistication. Our microarrays now contained over 4600 cDNAs and our experimental system was more complex. This time, we profiled the cellular gene expression changes that occurred in HeLa cells infected with influenza virus, again analyzing two time points after infection [3]. In addition, we evaluated the cellular response to a heat-inactivated virus, allowing us to identify gene expression changes that were independent of viral replication. Genes involved in protein synthesis, transcriptional regulation, and cytokine signaling were induced by the replicating virus, and many of these changes suggested an active cellular antiviral response. Interestingly, viral replication also resulted in the down-regulation of many cellular genes, particularly at the later time point. The gene expression pattern induced by the inactivated virus included the induction of the cellular metallothionein genes, which may represent a protective response to virus-induced oxidative stress.

Our early use of microarrays also included the first systematic attempt to define the full repertoire of dsRNA-regulated genes [4], and we used arrays, in conjunction with polysome fractionation, to identify cellular and viral mRNAs that increased in abundance and which were selectively recruited to polyribosomes following influenza virus infection [5]. Our dsRNA study identified 175 dsRNA-stimulated genes, which encoded proteins involved in a broad range of cellular functions and metabolic pathways. In addition, this study showed for the first time that many genes are also down-regulated by dsRNA. Our polysome fractionation studies, headed by John Kash, revealed that many of the cellular mRNAs that are recruited to polyribosomes during influenza virus infection are involved in the control of cell proliferation and survival, mitogenesis, or the inflammatory response. In addition, many of these mRNAs contain a putative binding site for GRSF-1, a cellular mRNA-binding protein that also specifically interacts with a discrete and conserved region of the viral 5′ untranslated region [6].

### 3. Welcome to the world of bioinformatics and computational biology

As our use of microarrays extended to an increasing number of experimental systems, we quickly found ourselves having to cope with an avalanche of data. To manage the huge amounts of information generated from multiple array experiments, our bioinformatics group, then

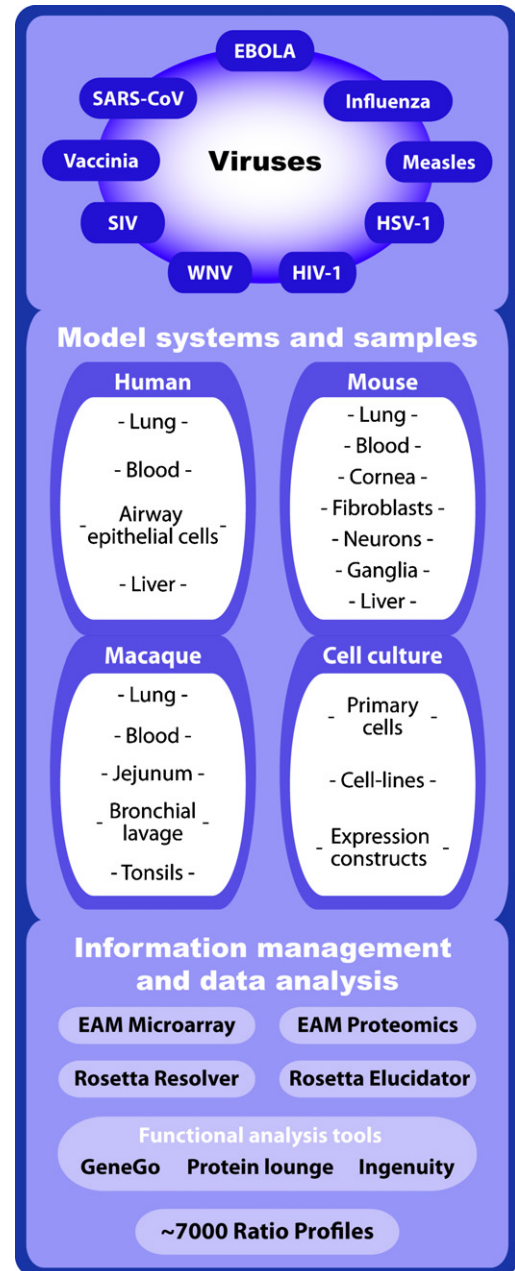


Fig. 1. Representation of the range of viruses and experimental systems we have evaluated using genomic and proteomic technologies. Highlights of experiments related to the use of many of these experimental systems, particularly for hepatitis C virus and influenza, are summarized in this review. SARS-CoV: severe acute respiratory syndrome-associated coronavirus; SIV: simian immunodeficiency virus; WNV: West Nile virus; HIV-1: human immunodeficiency virus type 1; HSV-1: herpes simplex virus type 1; EAM: Expression Array Manager.

headed by Jeff Furlong, developed a custom gene expression database, called Expression Array Manager [7]. In addition to serving as a central data repository, Expression Array Manager functions as our laboratory information management system, provides a streamlined data analysis pipeline, and provides a mechanism for publishing our gene expression data to the scientific community (available at <http://expression.viromics.washington.edu/>).

Most of our core genomic and proteomic data analysis functions are now performed using the Rosetta Biosoftware systems Resolver and Elucidator. These systems are designed for managing and analyzing large amounts of gene expression and mass spectrometry data and both contain a suite of higher-order analysis tools. To explore genomic and proteomic data in greater depth, we also use a variety of tools that provide a diversity of analytical choices. As an example, we use Spotfire DecisionSite to format data for further analysis using tools such as Ingenuity Pathway Analysis or MetaCore (GeneGo). These tools allow us to leverage proprietary biological content databases and examine gene expression (or protein abundance) data in the context of complex pathways to identify connections between differentially expressed genes. Sean Proll, Bryan Paeper, and Jon Rue provide the laboratory with expert guidance on the use of these tools and ensure that data moves efficiently through this complex pipeline.

#### **4. Hepatitis C virus: from cell culture to liver transplant biopsies**

We had earlier discovered that hepatitis C virus (HCV) uses its nonstructural 5A (NS5A) protein to repress the interferon-induced protein kinase, PKR, a primary mediator of the antiviral effects of interferon [8]. Our first application of gene expression profiling to HCV research was therefore to evaluate the effects of interferon treatment on cellular gene expression and to determine whether NS5A was able to alter this pattern of expression [9]. Gary Geiss again headed these studies in which we examined the effects of interferon treatment by using several types of human cells, including HeLa cells, liver cell lines, and primary fetal hepatocytes. In response to interferon, 50 of the approximately 4600 genes examined were consistently induced in each of these cell types and another 60 were induced in a cell type-specific manner. A search for interferon-stimulated response elements (ISREs) in genomic DNA located upstream of interferon-stimulated genes revealed both previously identified and novel putative ISREs. The expression of NS5A partially blocked the interferon-mediated induction of interferon-stimulated genes and inhibited the induction of a reporter gene driven from an ISRE-containing promoter, further suggesting that NS5A may play a role in HCV resistance to interferon. This study also set the stage for our use of genomics to better understand the interferon response and the varied mechanisms used by viruses to evade the effects of interferon.

##### *4.1. Birth of the NIDA center on functional genomics and HCV-associated liver disease*

What really accelerated our HCV genomics program was being awarded a National Institute on Drug Abuse (NIDA) P30 Center to do translational and clinically relevant genomics studies (<http://nida.viromics.washington.edu/>).

Maria Smith headed many of our early studies, the first of which looked for novel tumor markers in surgical liver samples from patients with hepatocellular carcinoma (HCC) [10]. This study, which by now used microarrays containing over 13,000 human cDNAs, revealed a set of 50 potential HCC marker genes that were up-regulated in the majority of the tumors analyzed. This HCC marker set contained several cancer-related genes as well as a set of genes encoding secreted or plasma proteins that may provide potential serological markers for HCC.

We performed a similar set of analyses on surgical material and core biopsy specimens obtained from HCV-infected patients with liver cirrhosis [11]. Importantly, this study also included an analysis of normal liver samples to determine normal physiologic variation in gene expression. To identify markers associated with cirrhosis, genes that were differentially expressed in normal liver, or in HCC, were subtracted from the set of genes differentially expressed in cirrhotic livers. The resulting gene set included genes expressed in activated lymphocytes infiltrating the cirrhotic liver and genes involved in the remodeling of the extracellular matrix, cell-cell interactions associated with cytoskeleton rearrangements, the anti-apoptotic pathway of Bcl-2 signaling, and the interferon response. Together, this analysis identified several potential gene expression markers of HCV-associated liver disease and contributed to our rapidly expanding database of experiments describing HCV pathogenesis.

Perhaps one of our most promising applications of genomics to HCV-associated liver disease is our work with serial liver biopsy samples from HCV-infected liver transplant patients. Liver transplant recipients infected with HCV develop recurrent hepatitis soon after transplantation and, in some cases, progress to fibrosis within two years of the transplant. Our goals are to identify molecular processes influencing liver disease progression and to find potential gene expression markers of early fibrosis. To achieve these goals, we are working closely with clinicians in the liver transplant unit at the University of Washington, including Anne Larson, Robert Carithers, and James Perkins, to collect biopsy samples and to integrate the gene expression data obtained from these samples with clinical observations.

Our initial analyses were performed on serial liver biopsy specimens obtained from 13 transplant recipients at 0, 3, 6, and 12 months after transplantation [12]. Gene expression data were compared with clinical observations and with gene expression data obtained from 55 nontransplant HCV-infected and uninfected liver samples. Our analyses revealed several specific gene expression patterns, the first of which was unique for the transplant recipients regardless of their infection status. The corresponding genes encoded stress response proteins and blood proteins involved in coagulation that were differentially expressed in response to post-transplantation graft recovery. The second pattern was specific to HCV-infected samples and included the increased expression of genes encoding components of the interferon-mediated antiviral response and immune system. This

pattern was absent or suppressed in the patients who developed early fibrosis, indicating that disease progression might result from an impaired liver response to infection. Finally, we identified gene expression patterns that were specific for the 12-month biopsy specimens of all four HCV-infected patients who developed early fibrosis after transplantation. The identified gene expression patterns may prove useful for diagnostic and prognostic applications in HCV-infected patients, including predicting early progression to fibrosis. We are continuing to collect and analyze biopsy samples as new patients are recruited into the study, which should enable us to obtain increasing levels of confidence in the predictive power of the markers that we identify.

In other recent studies, Kathie Walters and Sharon Lederer have examined the gene expression profiles that differentiate alcohol- and HCV-induced liver cirrhosis [13] and the gene expression patterns associated with HCV-induced pathogenesis in individuals co-infected with HIV-1 [14]. We found that global gene expression patterns vary significantly depending upon the etiology of liver disease and that stages of liver cirrhosis can be differentiated based on gene expression patterns in ethanol-induced, but not HCV-induced, disease. Many of the gene expression changes specifically observed in HCV-infected cirrhotic livers are associated with activation of the innate immune response. In contrast, we found that intrahepatic global gene expression profiles do not differ between HCV- and HCV/HIV-1 co-infected individuals. However, a specific gene expression pattern that may be associated with HCV-induced pathogenesis was identified. This pattern has similarities to the gene expression profiles in transplant patients who progress to fibrosis within one year of transplantation, suggesting it may also be relevant to predicting disease progression.

#### 4.2. *The chimeric mouse model*

Although clinical samples can provide considerable insight into the changes in cellular gene expression that occur during liver disease progression, they are not necessarily ideal for studies investigating the molecular mechanisms responsible for these changes. These types of studies typically require an animal model that can be experimentally manipulated. In this regard, research in the HCV field has been significantly enhanced by the development of a mouse model of HCV infection. This model, developed by Lorne Tyrrell and Norman Kneteman at the University of Alberta, Canada, is based on the severe combined immunodeficiency disorder (SCID)-beige/albumin (Alb)-urokinase plasminogen activator (uPA) transgenic mouse. Intrasplenic injection of freshly isolated human hepatocytes into these mice leads to the repopulation of the mouse liver with human hepatocytes. The result is a mouse containing a chimeric mouse-human liver and these animals can be persistently infected with a variety of HCV genotypes [15].

The chimeric mouse model provides a number of features that make it a useful system in which to study HCV–host interactions. Because groups of mice are transplanted with hepatocytes from different donors, we have the opportunity to analyze host-specific responses to HCV. In addition, the lack of an adaptive immune response in these animals makes it possible to distinguish between virus-mediated and immune-mediated effects on gene expression. Kathie Walters is using this model to characterize the host transcriptional response to HCV infection. Because liver samples from these animals typically have small percentages of contaminating mouse liver tissue, she first evaluated the level of cross-hybridization of mouse liver mRNA to the human probes present on the microarray [16]. The small set of genes identified (less than 2% of the genes present on the array) is either removed from subsequent data analysis or changes in the expression of these genes are validated using human-specific quantitative real-time PCR.

In an initial series of experiments, groups of mice transplanted with hepatocytes from different donors were inoculated with a single source of HCV and gene expression profiling was performed to characterize the host response to infection [17]. Although all HCV-infected animals showed evidence of an interferon response, the level of this response, both in the intensity and number of up-regulated interferon-induced genes, varied between animals depending upon the origin of the donor hepatocytes. These results indicate that host genetic factors contribute to the variation in the host response to HCV, including the activation of innate antiviral signaling pathways. Mice with weak interferon responses also tended to have high levels of intrahepatic HCV RNA, indicating that an ineffective interferon response may allow increased levels of viral replication. These animals also accumulated higher numbers of differentially expressed genes than did mice with strong interferon responses. Interestingly, we have also observed impaired interferon responses and the accumulation of increased numbers of differentially expressed genes in the intrahepatic gene expression profiles of transplant patients who develop fibrosis within one year of transplantation.

#### 4.3. *Entrance of proteomic technologies*

HCV research and the NIDA P30 Center also provided our entry point into proteomics. The ability to look at the protein content of a cell provides a strong complement to our genomic analyses, allowing us to evaluate how gene expression changes in response to virus infection translate into changes in the abundance of proteins, the molecules that directly carry out biological functions. Moreover, proteomic analysis of body fluids, such as blood, holds the promise of identifying candidate biomarkers for human diagnostic or prognostic applications.

Our initial venture into the realm of proteomics was led by Wei Yan in collaboration with Ruedi Aebersold at the Institute of Systems Biology. We began by using tandem

mass spectrometry to analyze the global proteome of the Huh7 hepatoma cell line and cultured primary and immortalized human fetal hepatocytes [18]. We also used isotope coded affinity tag (ICAT) technology to measure the changes in protein abundance that occurred in human hepatocytes in response to interferon treatment [19]. These analyses led to the generation of a liver proteome dataset consisting of 2159 unique proteins. Among the identified proteins were 78 interferon-regulated proteins that play roles in a multitude of cellular functions including antiviral defense, immune response, cell metabolism, and signal transduction. These data also contributed to the development of PeptideAtlas, a public resource for further annotating and validating the human genome by mapping identified peptide sequences to the human genome sequence [20].

For our more recent proteomic work, we have teamed with Richard Smith at Battelle, Pacific Northwest National Laboratory (PNNL). In our first studies with the Smith group, we used high mass accuracy Fourier transform ion cyclotron resonance (FTICR) mass spectrometry, coupled with the accurate mass and time (AMT) tag approach [21], to perform global proteomic analyses on Huh-7.5 cells containing a full-length HCV replicon [22]. Using this more sensitive technology, we identified over 4200 proteins. As a first-pass means to detect changes in protein abundance associated with HCV infection, we also performed a semi-quantitative comparison of total peptide identifications. This peptide-spectral count (“peptide-hits”) approach revealed HCV-related perturbations in the abundance of a variety of proteins associated with lipid metabolism and oxidative stress.

We have also taken advantage of the nanoproteomic platforms available at PNNL to perform proteomic studies on human liver biopsy samples that yield only limited amounts of protein. These studies were performed using a high-sensitivity 11.4-T FTICR mass spectrometer coupled with the AMT tag approach and our Huh-7.5 cell protein database. Using this method, we identified over 1500 proteins from only 2 µg of a protein digest obtained from a liver biopsy sample [22]. The proteins identified included many of relevance to HCV infection and liver disease, including cellular proteins involved in lipid metabolism and the interferon response. The number of proteins identified by this approach is considerably larger than what has been reported in other published studies of the liver proteome, where less-sensitive methods, requiring milligram amounts of protein lysate, have identified only a few hundred proteins. Deborah Diamond leads the proteomics research in our laboratory and has recently authored a review that more fully discusses current developments in the application of proteomics to liver disease research [23].

## 5. Influenza virus: my first love

Influenza research has been a mainstay of the laboratory for over 20 years, and as mentioned earlier, some of our

earliest genomic studies were performed on cells infected with influenza virus. Today, with worldwide attention focused on highly pathogenic avian influenza and the looming threat of a new pandemic, influenza research has taken on a renewed vigor. One of the priorities in this research is the development of improved animal models to better understand influenza pathogenesis and to develop new diagnostic, therapeutic, and vaccine strategies, all of which are important components of preparedness for the next influenza pandemic [24].

Over the past two years, we have taken on this challenge by developing a macaque influenza infection model that features genomic and proteomic analyses as a key component of the experimental design. Carole Baskin has taken a leading role in this effort and was the first to report on the use of genomic technologies to characterize experimental influenza virus infection using a pig-tailed macaque model [25]. This study was performed using a mildly pathogenic strain of virus (A/Texas/36/91) with the goal of constructing a blueprint of an uncomplicated influenza infection. Gene expression profiling performed on lung tissue and tracheobronchial lymph nodes revealed that numerous genes associated with the interferon response were differentially expressed, particularly at day 4 after inoculation. We also observed an increase in the expression of genes encoding various mediators of chemotaxis, adhesion, and the transmigration of immune cells, suggesting the trafficking of these cells into the lungs and tracheobronchial lymph nodes, which was confirmed by histopathology. In addition, this study revealed gene expression changes relevant to the processing of antigens on MHC class I complexes, particularly in lung tissue, and many genes encoding proteins relevant to T-cell function were induced at one or more time points.

The gene expression profiling performed for this study was done using human cDNA microarrays, which until recently was the only option available for studies using nonhuman primates. However, because of the extensive use of macaques as models for a wide variety of human diseases (and in AIDS research in particular), we have been part of a push to develop genomic resources focused on macaque species [26]. This effort resulted the sequencing of the rhesus macaque genome [27], and James Wallace headed an effort to use this extensive sequence information to develop an oligonucleotide microarray containing over 17,000 unique macaque sequences [28]. Additional information related to nonhuman primate genomics is also available at [macaque.org](http://macaque.org), a Web site we have developed to disseminate macaque genomic and proteomic data and resources.

The macaque oligonucleotide array was used in a follow-up study in which we focused on the early innate immune response of macaques infected with influenza virus [29]. In this study, led by Tracey Baas and Carole Baskin, we also performed genomic analyses on whole blood samples from infected animals to determine whether we could detect gene expression signatures in the blood that would correlate with

those detected in the lung. In general, genomic analyses of lung tissue showed an increase in the expression of many genes associated with interferon signaling early after infection, while other immune and cytokine responses were sustained throughout the course of the infection. We also observed a number of genes (many of which were involved in the interferon response) that were differentially expressed in a similar fashion in lung and blood samples, particularly early after infection. This finding suggests that microarray analyses of blood samples may hold promise for diagnostic applications. Finally, this study also included a first of its kind global proteomic analysis of macaque lung tissue. This analysis identified over 3500 proteins, and consistent with the gene expression data, proteomic data also revealed an increase in the abundance of many proteins involved in the innate immune response.

### 5.1. *New insights into the 1918 pandemic virus*

What influenza virologist would not be interested in understanding the molecular mechanisms underlying the extreme pathogenesis of the virus responsible for the 1918 influenza pandemic? Studies related to this virus, which killed and estimated 50 to 100 million people worldwide [30], have been (and remain) a prominent fixture of the laboratory, and genomic analyses are providing important new insights into what made this virus so lethal. In our first experiments associated with the 1918 virus, we evaluated global gene expression patterns in cells infected with wild-type influenza, engineered viruses lacking all or part of the NS1 gene, or with a virus containing the NS1 gene of the 1918 pandemic virus [31]. Deletion of the NS1 gene increased the magnitude of expression of cellular genes implicated in the interferon, NF- $\kappa$ B, and other antiviral pathways, and a virus with a C-terminal deletion in its NS1 gene induced a cellular gene expression pattern intermediate between the patterns induced by wild-type and NS1 knockout viruses. In contrast, a virus containing the NS1 gene from the 1918 pandemic virus was more efficient at blocking the expression of interferon-regulated genes than its parental virus. Together, these results suggested that the cellular response to influenza virus is significantly influenced by the NS1 gene and that the 1918 NS1 is a particularly effective interferon antagonist. Of course, we have since learned that the story is much more complicated.

We next used a mouse infection model to assess the contribution of the 1918 HA and NA genes to viral pathogenesis [32]. These studies, led by John Kash, evaluated mouse-adapted A/WSN/33 viruses that were engineered to contain the HA and NA genes from the 1918 virus or from the nonlethal A/New Caledonia/20/99 virus. Microarray analyses performed on lung tissues isolated from all infected animals showed the activation of many genes involved in the inflammatory response, including cytokine, apoptosis, and lymphocyte genes. However, consistent with

histopathology analysis, the parental and 1918 HA/NA:WSN recombinant viruses showed increased expression of genes associated with activated T cells and macrophages, as well as genes involved in apoptosis, tissue injury, and oxidative damage that were not observed in mice infected with the New Caledonia HA/NA:WSN virus. These studies documented clear differences in gene expression profiles that were correlated with pulmonary disease pathology and suggested that an intrinsic property of the 1918 HA and NA proteins may be the production of a longer and more severe immune response culminating in a more destructive viral infection.

The reconstruction of all eight gene segments of the pandemic virus provided the opportunity to find out what genomics could reveal about the high-virulence phenotype of the 1918 virus. The initial characterization of this virus by Terrence Tumpey and coworkers demonstrated that it was highly virulent in mice and most animals died within five days after infection [33]. Genomic analyses, led by John Kash, showed that animals infected with the reconstructed virus showed an increased and accelerated activation of genes associated with pro-inflammatory and cell-death pathways by 24 h after infection [34]. Significantly, these genes remained activated until the death of the animal. This was in contrast to less dramatic and delayed host immune responses (and less severe disease pathology) in mice infected with influenza viruses containing only subsets of genes from the 1918 virus. These findings indicate a cooperative interaction between the 1918 influenza genes and suggest that enhanced inflammatory and cell-death responses may contribute to severe immunopathology.

Our most recent work with the reconstructed 1918 virus was done in collaboration with Yoshihiro Kawaoka. In these studies, we evaluated the host response to the reconstructed virus using a cynomolgus macaque infection model [35]. The 1918 virus replicated to high levels and spread rapidly throughout the respiratory tract of infected animals, causing severe damage and masses of infiltrating immune cells throughout the course of infection. Genomic analyses revealed that the 1918 virus triggered an aberrantly high and sustained expression of numerous genes involved in the innate immune response, including proinflammatory cytokines and chemokines. The early and sustained host response in macaques infected with the 1918 virus was similar to what we observed in mice, indicating that critical host responses that influence disease outcome may occur very early after infection. Interestingly, the 1918 virus also appeared to selectively attenuate the expression of specific innate immune response genes, including certain genes associated with the type 1 interferon response. We have not determined the mechanism for this attenuation, but it is possible that the viral NS1 gene may play a role in regulating this response. Although the details remain to be uncovered, the atypical innate immune response induced by the 1918 virus was insufficient for protection and may actually contribute to the lethality of the virus.

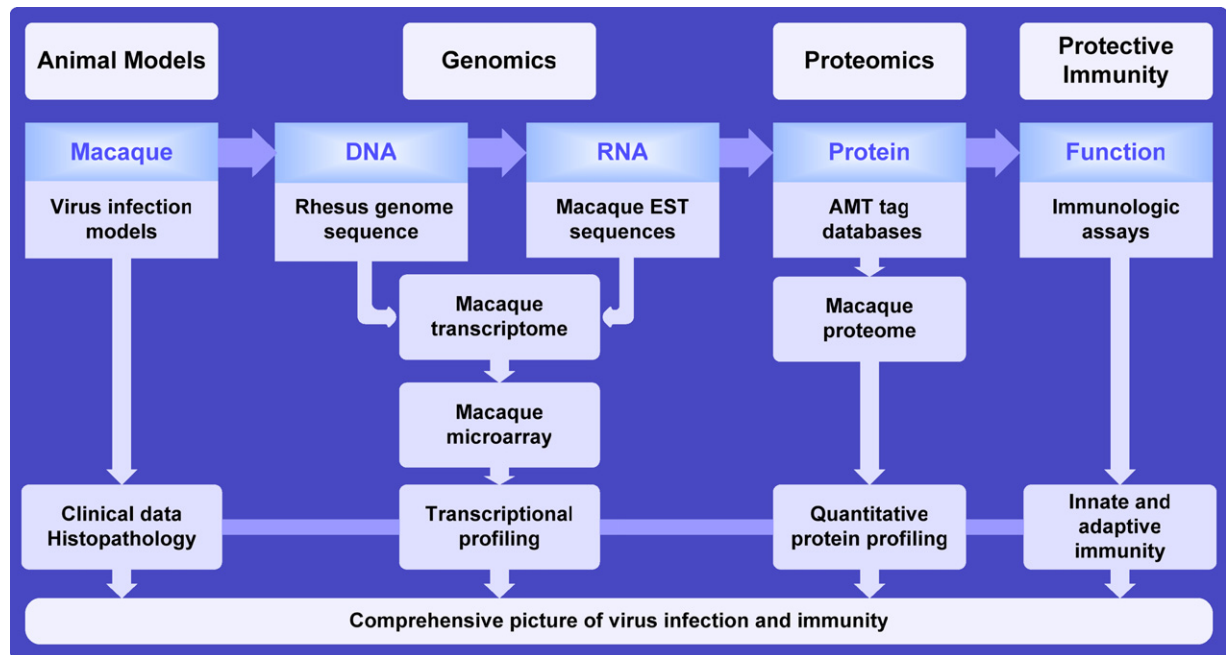


Fig. 2. The Division of Functional Genomics and Infectious Disease provides the virus infection models and the clinical, genomic, proteomic, and immunologic resources needed to perform comprehensive analyses on nonhuman primates. This allows us to analyze the response to virus infection at multiple points along the flow of biological information. EST: expressed sequence tag; AMT: accurate mass and time.

## 6. A vision for combining genomics with animal models of influenza and AIDS

Our interests in combining genomics and macaque infection models have led to the establishment of the Division of Functional Genomics and Infectious Disease at the Washington National Primate Research Center. Our laboratory is the central component of this Division, which is devoted to developing genomic and proteomic technologies (along with Richard Smith's group at PNNL), immunologic resources (in collaboration with Edward Clark and Murali-Krishna Kaja), and advanced computing and bioinformatics applications to enhance the use of nonhuman primates as models for influenza virus infection and AIDS. Our goal is to develop and apply the tools needed to perform comprehensive and integrated analyses on nonhuman primates and to analyze the response to virus infection at multiple points along the flow of biological information: from the whole animal to DNA to RNA to protein to biological function (Fig. 2).

By integrating these diverse types of data, we have the opportunity to better understand the dynamics of the host response to infection and the molecular mechanisms underlying the progression to virus-mediated disease, immunopathology, or the development of protective immunity. We also have the opportunity to better understand how gene expression changes in response to infection translate into changes in protein abundance and function, and how these changes correlate with clinical outcome. Moreover, we have the opportunity to assess how changes in gene expression and protein abundance affect immune

cell function, and how the innate immune response develops and its link to adaptive immunity. Ultimately, we believe this integrated approach will translate into molecular signatures that predict protective immunity or pathology, biomarkers for diagnostic or prognostic assays, and a rational base for improvements to antiviral therapies or vaccine strategies.

## 7. Conclusions

The studies outlined in this review provide just a glimpse of how we are using genomic and proteomic technologies to unravel the complexities of virus–host interactions. In addition to work with HCV and influenza, we are using genomic approaches to study a variety of other viruses. Tracey Baas heads collaborative research on SARS coronavirus [36] and herpes simplex virus [37]. When not working on pandemic influenza, John Kash also leads efforts focused on Ebola virus [38] and vaccinia virus [39], and Matthew Thomas and Michael Agy have explored gene expression changes associated with simian immunodeficiency virus infection [40]. We have also worked with Michael Gale on West Nile virus [41], and in our work on P58<sup>IPK</sup> (a cellular PKR inhibitor), we have even explored gene expression patterns associated with pancreatic  $\beta$ -cell depletion in diabetic mice [42]. Jamie Fornek is continuing the work on P58<sup>IPK</sup>, including using P58<sup>IPK</sup> knockout mice and genomics to examine the role of this PKR inhibitor during influenza virus infection. Gregory Zornetzer is developing a targeted proteomics approach to identifying



protein–protein interactions associated with key viral proteins.

The technologies we use continue to improve and data is accumulating at an ever increasing rate. Today's microarrays represent over 18,000 genes (over 10-fold more than when we started) that together with assorted controls are arrayed as an impressive 4 by 44,000 oligonucleotides per slide. Barry Robinson is working to incorporate laser capture microdissection into our studies, which will give us our first opportunity to examine the gene expression changes that occur in specific cell types isolated from complex tissues. Bioinformatics capabilities are becoming evermore sophisticated and Jon Rue and Matthew Harding ensure that we stay atop the most recent advances in data management and analysis. Our Expression Array Manager database now contains information from over 7000 expression profiles representing more than 50 million measurements of differential gene expression.

As we have discussed elsewhere [43], the future of virology depends on making progress in understanding how things work at a systems level, and we must ensure that genomic and proteomic technologies do more than simply generate ever larger quantities of data without providing clues to underlying function. Nevertheless, we are confident that as technologies and experimental systems continue to progress, and data integration strategies become more mature [44], functional genomics will make ever greater contributions to virology. As an example, we are continuing to take our use of genomics in new directions and Robert Palermo is working together with Marjorie Robert-Guroff (National Cancer Institute) to pursue genomic approaches to improving AIDS vaccine development. Genomic analyses during vaccine trials may reveal gene expression markers of protective immunity or gene expression changes that are indicative of a predisposition to a particular response to immunization and subsequent challenge. All told, the past eight years have been an exciting journey full of discoveries, and we look forward to new breakthroughs in the future. Indeed, life after the Milstein award has not been too bad.

### Acknowledgements

We thank William Morton and David Anderson, past and present Directors of the Washington National Primate Research Center, respectively, for sharing the vision of exploiting genomic technologies to maximize the benefits gained from research using animal models and for their unwavering support of the Division of Functional Genomics and Infectious Disease. We also thank Marlene Wambach, who has been with the laboratory since day one, Cynthia Baker, Rosalind Billharz, Cecilia Boyer, Victoria Carter, Eric Chan, Alan Goodman, Kara Jensen, Yu Li, Socorro Ortiz, Grace Park, Tom Teal, and other colleagues past and

present that have contributed to the laboratory's success. We apologize for not being able to list them all by name. Finally, we are grateful to Sean Proll for Fig. preparation. Funding for genomic and proteomic studies in our laboratory is provided by Public Health Service grants R01AI022646, R01HL080621, R21AI017892, R21AI063436, R24RR-016354, P01AI052106, P01AI058113, P30DA015625, and P51RR000166 from the National Institutes of Health.

### References

- [1] Katze MG. Interferon, PKR, virology, and genomics: what is past and what is next in the new millennium? *J Interferon Cytokine Res* 2002;22:283–6.
- [2] Geiss GK, Bumgarner RE, An MC, Agy MB, van't Wout AB, Hammersmark E, et al. Large-scale monitoring of host cell gene expression during HIV-1 infection using cDNA microarrays. *Virology* 2000;266(1):8–16.
- [3] Geiss GK, An MC, Bumgarner RE, Hammersmark E, Cunningham D, Katze MG. Global impact of influenza virus on cellular pathways is mediated by both replication-dependent and -independent events. *J Virol* 2001;75:4321–31.
- [4] Geiss G, Ge J, Guo J, Bumgarner R, Katze MG, Sen GC. A comprehensive view of gene expression by double-stranded RNA-mediated cell signaling. *J Biol Chem* 2001;276:30178–82.
- [5] Kash JC, Cunningham DM, Smit MW, Park Y-W, Fritz D, Wilusz J, et al. Selective translation of eukaryotic mRNAs: functional molecular analysis of GRSF-1, a positive regulator of influenza virus protein synthesis. *J Virol* 2002;76:10417–26.
- [6] Park Y-W, Wilusz J, Katze MG. Regulation of eukaryotic protein synthesis: selective influenza viral mRNA translation is mediated by the cellular RNA-binding protein GRSF-1. *Proc Natl Acad Sci USA* 1999;96:6694–9.
- [7] Korth MJ, Katze MG. Unlocking the mysteries of virus–host interactions: does functional genomics hold the key? *Ann N Y Acad Sci* 2002;975:160–8.
- [8] Gale Jr M, Korth MJ, Tang NM, Tan S-L, Hopkins DA, Dever TE, et al. Evidence that hepatitis C virus resistance to interferon is mediated through repression of the PKR protein kinase by the nonstructural 5A protein. *Virology* 1997;230:217–27.
- [9] Geiss GK, Carter VS, He Y, Kwieciszewski BK, Holzman T, Korth MJ, et al. Gene expression profiling of the cellular transcriptional network regulated by alpha/beta interferon and its partial attenuation by the hepatitis C virus nonstructural 5A protein. *J Virol* 2003;77:6367–75.
- [10] Smith MW, Yue ZN, Geiss GK, Sadovnikova NY, Carter VS, Boix L, et al. Identification of novel tumor markers in hepatitis C virus-associated hepatocellular carcinoma. *Cancer Res* 2003;63:859–64.
- [11] Smith MW, Yue ZN, Korth MJ, Do HA, Boix L, Fausto N, et al. Hepatitis C virus and liver disease: global transcriptional profiling and identification of potential markers. *Hepatology* 2003;38:1458–67.
- [12] Smith MW, Walters K-A, Korth MJ, Fitzgibbon M, Proll SC, Thompson JC, et al. Gene expression patterns that correlate with hepatitis C and early progression to fibrosis in liver transplant recipients. *Gastroenterol* 2006;130:179–87.
- [13] Lederer SL, Walters KA, Proll S, Paepfer B, Robinson S, Boix L, et al. Distinct cellular responses differentiating alcohol- and hepatitis C virus-induced liver cirrhosis. *Virol J* 2006;3:98.
- [14] Walters KA, Smith MW, Pal S, Thompson JC, Thomas MJ, Yeh MM, et al. Identification of a specific gene expression pattern associated

- with HCV-induced pathogenesis in HCV- and HCV/HIV-infected individuals. *Virology* 2006;350:453–64.
- [15] Mercer DF, Schiller DE, Elliott JF, Douglas DN, Hao C, Rinfret A, et al. Hepatitis C virus replication in mice with chimeric human livers. *Nat Med* 2001;7(8):927–33.
- [16] Walters KA, Joyce MA, Thompson JC, Proll S, Wallace J, Smith MW, et al. Application of functional genomics to the chimeric mouse model of HCV infection: optimization of microarray protocols and genomics analysis. *Virology* 2006;337:37.
- [17] Walters KA, Joyce MA, Thompson JC, Smith MW, Yeh MM, Proll S, et al. Host-specific response to HCV infection in the chimeric SCID-beige/Alb-uPA mouse model: role of the innate antiviral immune response. *PLOS Pathogens* 2006;2(6):591–601.
- [18] Yan W, Lee H, Deutsch EW, Lazaro CA, Tang W, Chen E, et al. A dataset of human liver proteins identified by protein profiling via isotope coded affinity tag (ICAT) and tandem mass spectrometry. *Mol Cell Proteomics* 2004;3:1039–41.
- [19] Yan W, Lee H, Yi EC, Reiss D, Shannon P, Kwieciszewski BK, et al. System-based proteomic analysis of the interferon response in human liver cells. *Genome Biol* 2004;5(8):R54.
- [20] Desiere F, Deutsch EW, Nesvizhskii AI, Mallick P, King NL, Eng JK, et al. Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. *Genome Biol* 2005;6(1):R9.
- [21] Bogdanov B, Smith RD. Proteomics by FTICR mass spectrometry: top down and bottom up. *Mass Spectrom Rev* 2005;24(2):168–200.
- [22] Jacobs JM, Diamond DL, Chan EY, Gritsenko MA, Qian W, Stastna M, et al. Proteome analysis of liver cells expressing a full-length hepatitis C virus (HCV) replicon and biopsy specimens of posttransplantation liver from HCV-infected patients. *J Virol* 2005;79(12):7558–69.
- [23] Diamond DL, Proll SC, Jacobs JM, Chan EY, Camp DG, Smith RD, et al. HepatoProteomics: applying proteomic technologies to the study of liver function and disease. *Hepatology* 2006;44(2):299–308.
- [24] Fauci AS. Pandemic influenza threat and preparedness. *Emerg Infect Dis* 2006;12(1):73–7.
- [25] Baskin CR, García-Sastre A, Tumpey TM, Bielefeldt-Ohmann H, Carter VS, Nistal-Villán E, et al. Integration of clinical data, pathology, and cDNA microarrays in influenza virus-infected pigtailed macaques (*Macaca nemestrina*). *J Virol* 2004;78:10420–32.
- [26] Magness CL, Fellin PC, Thomas MJ, Korth MJ, Agy MB, Proll SC, et al. Analysis of the *Macaca mulatta* transcriptome and the sequence divergence between *Macaca* and human. *Genome Biol* 2005;6(7):R60.
- [27] Rhesus macaque genome sequencing and analysis consortium. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 2007;316:222–234.
- [28] Wallace JC, Korth MJ, Paepfer BW, Proll SC, Thomas MJ, Magness CL, et al. High-density rhesus macaque oligonucleotide microarray design using early-stage rhesus genome sequence information and human genome annotations. *BMC Genomics* 2007;8:28.
- [29] Baas T, Baskin CR, Diamond DL, García-Sastre A, Bielefeldt-Ohmann H, Tumpey TM, et al. Integrated molecular signature of disease: analysis of influenza virus-infected macaques through functional genomics and proteomics. *J Virol* 2006;80:10813–28.
- [30] Taubenberger JK, Morens DM. 1918 Influenza: the mother of all pandemics. *Emerg Infect Dis* 2006;12(1):15–22.
- [31] Geiss GK, Salvatore M, Tumpey TM, Carter VS, Wang X, Basler CF, et al. Cellular transcriptional profiling in influenza A virus-infected lung epithelial cells: the role of the nonstructural NS1 protein in the evasion of the host innate defense and its potential contribution to pandemic influenza. *Proc Natl Acad Sci USA* 2002;99:10736–41.
- [32] Kash JC, Basler CF, García-Sastre A, Carter V, Billharz R, Swayne DE, et al. Global host immune response: pathogenesis and transcriptional profiling of type A influenza viruses expressing the hemagglutinin and neuraminidase genes from the 1918 pandemic virus. *J Virol* 2004;78:9499–511.
- [33] Tumpey TM, Basler CF, Aguilar PV, Zeng H, Solorzano A, Swayne DE, et al. Characterization of the reconstructed 1918 Spanish influenza pandemic virus. *Science* 2005;310(5745):77–80.
- [34] Kash JC, Tumpey TM, Proll SC, Carter V, Perwitasari O, Thomas MJ, et al. Genomic analysis of increased host immune and cell death responses induced by 1918 influenza virus. *Nature* 2006;443:578–81.
- [35] Kobasa D, Jones SM, Shinya K, Kash JC, Copps J, Ebihara H, et al. Aberrant innate immune response in lethal infection of macaques with the 1918 influenza virus. *Nature* 2007;445:319–23.
- [36] Baas T, Taubenberger J, Chong PY, Chui P, Katze MG. SARS-CoV virus–host interactions and comparative etiologies of acute respiratory distress syndrome as determined by transcriptional and cytokine profiling of formalin-fixed paraffin-embedded tissues. *J Interferon Cytokine Res* 2006;26:309–17.
- [37] Pasiacka TJ, Baas T, Carter VS, Proll SC, Leib DA, Katze MG. Functional genomics analysis of herpes simplex virus type 1 counteraction of the host innate response. *J Virol* 2006;80:7600–12.
- [38] Kash JC, Mühlberger E, Carter V, Grosch O, Perwitasari O, Proll SC, et al. Global suppression of the host antiviral response by Ebola- and Marburgviruses: increased antagonism of the type 1 interferon response is associated with enhanced virulence. *J Virol* 2006;80:3009–20.
- [39] Langland JO, Kash JC, Carter V, Thomas MJ, Katze MG, Jacobs BL. Suppression of proinflammatory signal transduction and gene expression by the dual nucleic acid binding domains of the vaccinia virus E3L proteins. *J Virol* 2006;80(20):10083–95.
- [40] Thomas MJ, Agy MB, Proll SC, Paepfer BW, Li Y, Jensen KL, et al. Functional gene analysis of individual response to challenge with simian immunodeficiency virus SIVmac239 in *Macaca mulatta* peripheral blood mononuclear cell culture. *Virology* 2006;348:242–52.
- [41] Fredericksen BL, Smith M, Katze MG, Shi PY, Gale Jr M. The host response to West Nile Virus infection limits viral spread through the activation of the interferon regulatory factor 3 pathway. *J Virol* 2004;78(14):7737–47.
- [42] Ladiges WC, Knoblauch SE, Morton JF, Korth MJ, Sopher BL, Baskin CR, et al. Pancreatic beta-cell failure and diabetes in mice with a deletion mutation of the endoplasmic reticulum molecular chaperone gene P58<sup>IPK</sup>. *Diabetes* 2005;54(4):1074–81.
- [43] Wallace JC, Korth MJ, Diamond DL, Proll SC, Katze MG. Virology in the 21st century: finding function with functional genomics. *Future Virol* 2006;1:47–53.
- [44] Reiss DJ, Avila-Campillo I, Thorsson V, Schwikowski B, Galitski T. Tools enabling the elucidation of molecular pathways active in human disease: application to Hepatitis C virus infection. *BMC Bioinformatics* 2005;6:154.

**Michael G. Katze** is Professor of Microbiology and Associate Director and Core Staff Scientist at the Washington National Primate Research Center. He received his Ph.D. from Hahnemann Medical College and was a postdoctoral fellow at the University of Uppsala in Sweden as part of a fellowship with the European Molecular Biology Organization. Prior to joining the faculty at the University of Washington, Dr. Katze conducted research in molecular biology and virology at the Memorial Sloan-Kettering Cancer Center in New York City. Dr. Katze has studied virus–host interactions for more than 25 years and is a recognized leader in applying genomic and proteomic technologies to the study of virus–host interactions and the interferon response. He is an author of over 160 papers and reviews, and in September 1999, received the Milstein award from the International Society of Interferon and Cytokine Research for his contributions to the interferon field. In February 2006, Dr. Katze was awarded the prestigious Dozor Scholar Award by the Israeli Microbiology Society for his research

accomplishments. He heads a laboratory of over 30 individuals, including graduate students, postdoctoral fellows, research scientists, technologists, software engineers, and bioinformatics and information technology specialists.

**Marcus J. Korth** is a Senior Research Scientist in Dr. Katze's laboratory. He received Bachelor's degrees in Psychology from the University of Wisconsin-Eau Claire and in Microbiology and Medical Technology from the University of Montana. He received his Ph.D. in Microbiology from the University of Washington in 1993 for work on the role of fimbrial adhesins in *Escherichia coli* pathogenesis. After completing a postdoctoral fellowship with Dr. Katze, where he focused on characterizing the PKR inhibitor and endoplasmic reticulum molecular chaperone P58<sup>IPK</sup>, he has remained in the laboratory where he is interested in applying high-throughput technologies to the study of virus–host interactions.

