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 12

The Virus–Host Interactome

Knowing the Players to Understand the Game

Monika Schneider[1], Jeffery R. Johnson[2], Nevan J. Krogan[2], Sumit K. Chanda[1]

[1]*Sanford-Burnham Medical Research Institute, La Jolla, CA, USA;* [2]*University of California, San Francisco, CA, USA*

Chapter Outline

1. INTRODUCTION

The interactions between viruses and human host cells encompass the activation of immune defenses, viral countermeasures, and viral hijacking of cellular proteins. As discussed in previous chapters, mammalian cells have a robust arsenal of antiviral and innate immune protective mechanisms, and viruses have evolved proteins that specifically target those host responses. It is the molecular interactions between virus and host that ultimately determine pathogenic outcome. Just as knowing the positions and roles of players is essential to understanding a football or basketball game, knowing the complete roster of relevant genes and proteins, and the roles that they play, is essential to understanding viral pathogenesis. This chapter describes the high-throughput screening approaches being used to identify the molecular "players" in viral pathogenesis, how they interact with one another, and the roles that they play in determining pathogenic outcome.

2. GENETIC FUNCTIONAL SCREENS

In the context of cellular infection, not only do viruses depend on their own genes for successful replication, but they also use genes expressed by the host. The cellular processes that are hijacked vary with the stage of infection, and conversely, viral infection induces a cohort of cellular genes that help the cell fend off the infecting virus. Viruses frequently mutate to adapt to the environment or to improve infection efficiency. In contrast, host proteins mutate at a much slower rate, if at all, and therapeutically targeting host proteins that are important for viral replication may be more successful at thwarting viral escape. A variety of high-throughput screens have therefore been developed to identify host proteins that are important for viral replication (Panda and Cherry, 2012).

Host proteins that promote the survival or replication of a virus are referred to as *host factors,* and host proteins that limit viral survival or replication are termed *restriction factors* or antiviral factors. The loss of host restriction factors results in increased viral replication, whereas overexpression of restriction factors reduces viral proliferation. Examples of broad-acting host restriction factors include the products of many interferon-stimulated genes (see Chapter 4, Innate Immunity).

There are several genetic tools available to uncover host and restriction factors (Table 1). Loss-of-function and

Viral Pathogenesis. http://dx.doi.org/10.1016/B978-0-12-800964-2.00012-4

TABLE 1 Methods Used to Probe the Virus–Host Interactome

Method	How It Works	Advantages	Disadvantages
Loss of function	Transient or stable siRNA or shRNA expression to knockdown expression of target genes	Can be used in high-throughput screens; can reveal genes that play a role in viral infection	False negatives and positives due to off-target activity of mismatched siRNAs or irrelevant immune activation
Gain of function	Overexpression of cDNAs	Can be used in high-throughput screens	False positives due to nonphysiological levels of the target gene
RNA-seq	Whole-genome sequencing of mRNA or small RNA at a given time point	Quantifies all transcripts in the cell	Only captures a snapshot of given infection
Yeast two-hybrid	Pairwise introduction of proteins that are fused to complementary fragments of a readout system	Low-cost and scalable	High false-positive and false-negative rates
Protein microarray	Whole cell lysate incubated with proteins or peptides immobilized on a solid surface	Can identify otherwise transient enzyme–substrate interactions	Protein complexes cannot be identified if the bait protein is not a direct interactor; high false-negative rate
Affinity purification and mass spectrometry	Protein complexes co-purified with tagged proteins then identified through mass spectrometry	Identifies indirect interactions; complexes are formed in vivo, so more physiologically relevant	Direct interactions are difficult to identify

gain-of-function screens probe individual genes to determine how each affects virus viability, whereas RNA-seq technologies measure the quantity of all transcripts induced or reduced by infection at a given time. The best tool to use depends upon the desired information, and each assay can be adjusted for optimal results. The details of genetic functional screens are discussed in the remainder of this section.

2.1 Loss-of-Function Screening Using RNA Interference

RNA interference (RNAi) screens are commonly used to globally identify the proteins that are involved in a cellular phenotype of interest. This method makes use of the cellular RNAi pathway, which naturally produces regulatory microRNAs (miRNAs) and small interfering RNAs (siRNAs) to specifically silence target genes. In the RNAi pathway, noncoding RNAs are processed into shorter stem-loop structures called pre-miRNA, which are then further cleaved into 20–22 base-pair siRNAs. One strand of the double-stranded siRNA, termed the guide strand, is then incorporated into the RNA-induced silencing complex for targeting to the corresponding mRNA. siRNAs and miRNAs are both processed through this pathway and bind to target mRNA through an eight-nucleotide seed region, but with different end results.

siRNAs bind to their target mRNA and induce degradation of the transcript. In contrast, miRNAs bind to their target mRNA and inhibit translation without leading to transcript degradation, in part due to their ability to bind to transcripts without perfect complementarity. The RNAi pathway can be experimentally induced through the introduction of synthetic siRNAs or short hairpin RNAs (shRNAs) in a transient or stable manner. Algorithms and mathematical models have been developed to design and predict targeting efficiency on the basis of sequence characteristics, such as GC content and whether hairpin structures form within the siRNA.

shRNAs intersect the RNAi pathway at an earlier step than siRNAs. shRNAs are approximately 70 nucleotides in length and are engineered to form a hairpin structure, similar to pre-miRNAs. The hairpin structure is recognized by endogenous RNAi proteins, which then process the shRNA into functional siRNAs. shRNAs can be delivered to the cell using an exogenous vector, or they can be incorporated into the genome for stable silencing via a lentiviral vector.

Although siRNA and shRNA are effective means of reducing the expression of target genes, it is expensive to generate whole-genome siRNA or shRNA libraries. This hurdle can be bypassed through the use of in vitro-generated siRNAs, termed endoribonuclease-prepared siRNAs (esiRNAs) (Yang et al., 2002). This method uses a cDNA library to transcribe each gene; the long dsRNA is then digested by bacterial RNase III. The resulting esiRNAs effectively knock down target gene expression. All gene-targeting strategies using RNAi have their limitations,

and as whole-genome screens are more widely used, reproducibility, consistency between assays, and off-target activities remain a concern (Jackson and Linsley, 2010).

2.2 CRISPR Genome Editing

A new system for gene knockdown, termed CRISPR (for clustered regularly interspaced short palindromic repeats), has recently been developed. CRISPR was originally discovered in bacteria, where it provides immunity against bacterial viruses by disrupting viral transcription (Barrangou et al., 2007). Genome-editing strategies using CRISPR take advantage of bacterial Cas9, an enzyme that catalyzes double-stranded DNA breaks, to specifically target and create deletions in the exons of a desired gene. CRISPR functions through the base pairing of a guide RNA (gRNA) to a specified genomic location. The gRNA directs Cas9 to the target sequence where it induces a double-strand break, leading to a sequence deletion or insertion. Unlike siRNA and shRNA, which can result in an incomplete knockdown of the target gene, CRISPR permanently edits the genome and can result in the complete knockout of a gene.

CRISPR has been used to achieve individual gene knockdown on a genome-wide scale by stably expressing Cas9 in mouse embryonic stem cells or in HeLa cells (Koike-Yusa et al., 2014; Zhou et al., 2014). A genome-wide library of pooled gRNAs is then generated and gRNAs are individually expressed using lentiviral vectors. There are few false positives, and the phenotype is stronger than that observed when the same genes are knocked down with a shRNA.

Although CRISPR offers several advantages to traditional genome editing tools, some studies have indicated that gRNAs can cause off-target gene mutation. In addition, genes that have multiple alleles may require multiple rounds of CRISPR treatment to completely knock out the target gene, and single-cell sorting is required to identify which cells harbor a homozygous or heterozygous knockout. Finally, as with RNAi, gene knockout using CRISPR is dependent on the ability of delivery vectors to enter the cell.

2.3 Gain-of-Function Screening

Genetic gain-of-function screens are the converse of loss-of-function screens since they ectopically express genes, sometimes in excess of physiological expression. There are several whole-genome libraries that can be used for this type of screening depending upon the readout and type of cell used. These include the Mammalian Gene Collection, which is curated by the National Institutes of Health, and the human ORFeome collection, which is curated by the Center for Cancer Systems Biology. Gain-of-function screens can also be run using either transient transfection or through the generation of stable expression cell lines.

There are scenarios in which gain-of-function screens are advantageous. For example, viral infection may cause the down-regulation of particular genes that have antiviral effects. Knocking down repressed or lowly expressed genes is not likely to result in a measureable phenotype. However, overexpression of these genes may overcome viral down-regulation, resulting in reduced viral proliferation. A disadvantage of overexpression is that it can generate more false positives than would be found in a loss-of-function screen because of nonspecific activities due to nonphysiological levels of gene dosage.

2.3.1 *Application of Gain-of-Function Screens to Viral Pathogenesis*

An example of using a gain-of-function screen to identify host antiviral proteins is shown in Figure 1. In this screen, interferon-stimulated genes (ISGs) were identified through published microarray and RNAi screening sets, and these data were used to curate a library of nearly 400 ISGs (Schoggins et al., 2011). The ISGs were then individually overexpressed to test which would have the ability to inhibit viral replication. Multiple viruses were used in the screen, including hepatitis C virus, HIV-1, yellow fever virus, West Nile virus, Venezuelan equine encephalitis virus, and Chikungunya virus. This approach revealed ISGs that were broadly antiviral as well as those that specifically inhibited only one or two viruses.

2.4 Vector Delivery Methods

The efficiency of an RNAi vector is dependent on the cell type and delivery method. There are several well-characterized methods for RNAi delivery, but desired throughput may limit the options available for a particular assay (Table 2). Lipid delivery methods use a phospholipid bilayer to form a vesicle around the siRNA or shRNA, which fuses with the cell membrane. Lipofection typically has a low level of toxicity, but sensitivity is dependent on the type of reagent used and the cell type being transfected. This method can be used for high-throughput screens. Electroporation delivers an electric pulse to the cells, which results in membrane pores, and the charge of the electric pulse helps to move the nucleotide vector into the cell. Electroporation can cause considerable cell death and requires transfer of the cells between the electroporation device and the plating well. Because of this extra step, this delivery method has limited throughput. Finally, viral vectors can be used for RNAi delivery. Retroviral vectors become integrated into the cellular genome, allowing for stable expression of the delivered shRNA. In contrast, adenoviruses do not incorporate into the genome and are therefore only useful for transient expression.

Due to their effectiveness at specifically reducing the expression of target genes, RNAi technologies also have

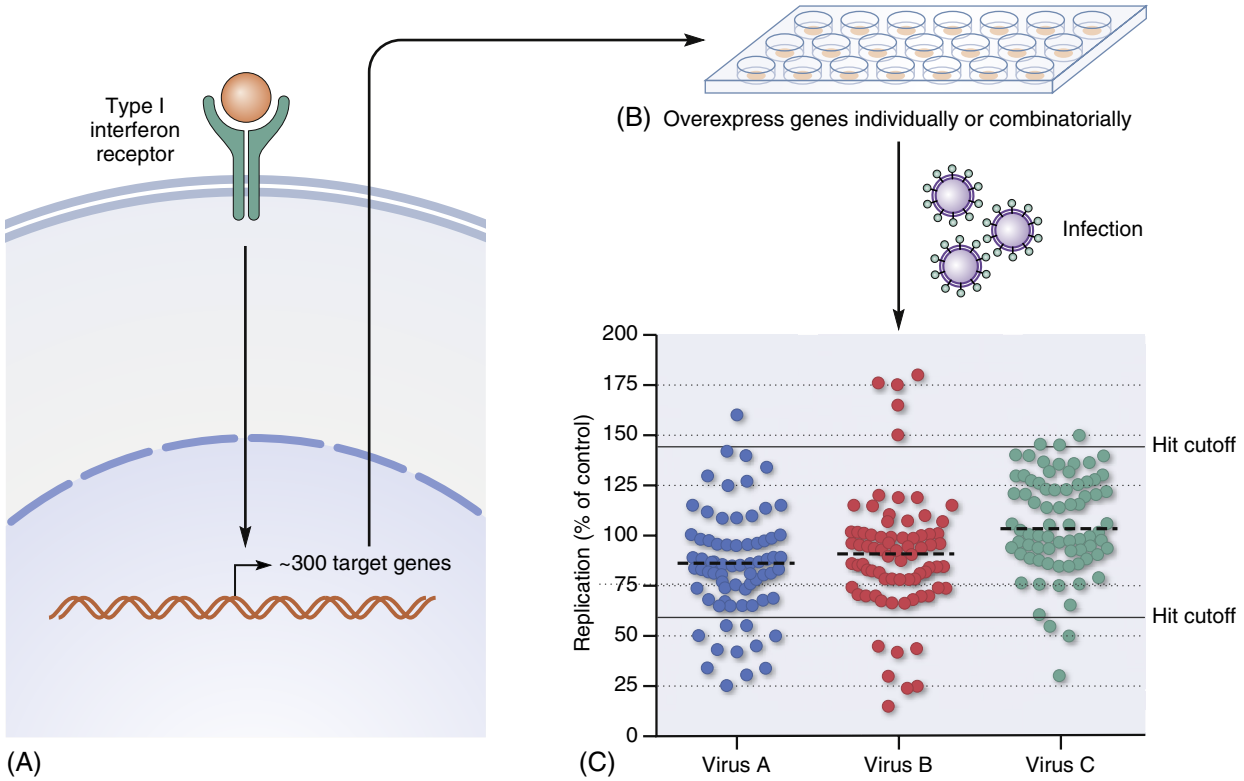

FIGURE 1 Gain-of-function screen to identify host antiviral proteins. (A) A library of genes that are induced after type I IFN treatment (ISGs) are subcloned into a lentiviral expression vector. (B) The genes are overexpressed and individual sets of cells are infected with a panel of different viruses. (C) Replication efficiency is measured, and those genes that significantly inhibit viral replication (lower cutoff) are further investigated for antiviral effects. Hits above the upper cutoff represent ISGs that enhance viral replication.

TABLE 2 Methods to Deliver Interfering RNA to Cells

Entry Method	How It Works	Potential Off-target Effects	Throughput
Lipid vectors	Lipid vesicle merges with the cell and releases through endosome	Cell death; irrelevant immune activation	High
Electroporation	Electric charge forms pores in the cell membrane, RNAi vectors enter into cytoplasm	Cell death	Low
Viral vectors	Fusion with cell membrane	Irrelevant immune activation; targeting wrong gene	Medium

great potential for therapeutic use. However, the delivery of RNAi vectors in vivo is a much greater challenge than delivery to cell lines. Uptake of siRNA or shRNA by target cells in vivo is poor, and the RNAi vectors themselves are often rapidly degraded once they enter the bloodstream. There has been some success in targeting specific organs, such as the liver, but systemic delivery poses additional challenges. There are, however, promising avenues for this type of delivery, including chemical modification of the siRNA or shRNA, cholesterol conjugation, viral vectors, and various polymers (Kanasty et al., 2013).

2.4.1 *Applications of RNAi Vector Delivery to Viral Pathogenesis*

Despite therapeutic delivery challenges, methods have been developed for screening siRNAs in vivo. For example, tumor suppressors have been identified by introducing shRNAs—targeting the mouse orthologs of genes deleted in human liver cancer—into premalignant cancer cells and then transplanting these cells into mice and testing for their ability to promote the formation of tumors (Zender et al., 2008). Virus–host interactions have also been probed by

incorporating miR-30-based hairpins into Sindbis virus. The shRNAs that conferred a survival advantage for the virus were then identified by using mRNA-seq. This method identified novel host factors required for antiviral gene transcription as well as many known ISGs (Varble et al., 2013).

2.5 Screening Formats and Readouts

Screening "hits" are defined as those genes that fall above or below a certain threshold, which is set by the control values. There is no uniform way to determine hits, and each research group has unique ways of identifying genes of value. Generally, the quality of a screen is determined through the calculation of a Z prime, which uses the means and standard deviations of the control samples to estimate the intrinsic value of the experimental samples. The Z prime cannot exceed 1, which indicates a perfect assay, and a good separation between negative and positive controls yields a Z prime between 0.5 and 1. Different methods of determining hits may account for some of the variability seen in screens that have been run using similar parameters.

Loss-of-function assays can be run as whole-genomic or targeted subgenomic screens, and siRNAs or shRNAs can be delivered in gene-based pooled or in arrayed formats. In pooled libraries, two to four siRNAs for the same gene are delivered simultaneously, which increases the likelihood of achieving significant gene knockdown. However, if one of the siRNAs is toxic to the cell, then that gene will be eliminated from the screen regardless of the effects of the other siRNAs. Following screening using a pooled library, the pools need to be deconvoluted through confirmation of phenotype and knockdown using each siRNA individually.

In an arrayed format, each siRNA or shRNA is plated in an individual well. Because multiple siRNAs or shRNAs may be used for each gene, this method greatly increases the size of the initial screen. However, screening in this way allows for the identification of toxic siRNAs and eliminates the need for a deconvolution step. With this method, target genes that have two or more effective siRNAs are apparent immediately, and thus off-target effects can be minimized.

2.5.1 *Viral Replication Readouts and Their Applications to Viral Pathogenesis*

There are several types of readouts that can be used to measure viral replication. A luciferase reporter can be incorporated into the virus, such that is it activated when the virus replicates. Lower values will be reported if the virus fails to enter the cell or if it is unable to subsequently propagate. This method was used in a screen to identify host factors involved in influenza virus replication (Konig et al., 2010). The influenza virus used had the viral HA gene replaced with a Renilla luciferase gene, which allowed luciferase activity to be measured following infection. Almost 295 genes were identified as being important for early replication steps, including those involved in endosomal processes, intracellular trafficking, and ubiquitination.

Another method to measure viral replication is through tracking the expression of a viral protein, or by replacing a viral protein with a tagged or fluorescent protein, such as green fluorescent protein. After infection, the cells can then be stained for the endogenous or tagged protein, and the amount of virus in each cell can be measured through high-content imaging, which uses robotics to image proteins or process cells through fluorescent microscopy. This method was also used to identify host factors required for influenza virus replication (Brass et al., 2009). Cells were transfected with siRNA pools, and after influenza virus infection, the cells were immunostained for the presence of HA, which was used as a surrogate for viral replication (Figure 2). Nearly 150 host antiviral genes were identified as potential targets after validation assays. This loss-of-function screen identified an important family of host restriction factors, call the IFITM (interferon-inducible transmembrane) proteins, which potently inhibit the replication of influenza virus, West Nile virus, and Dengue virus. IFITM3 was later confirmed as a broad-acting restriction factor through a gain-of-function screen using a curated list of ISGs (Schoggins et al., 2011).

Finally, spreading replication can be measured by quantifying the amount of new virus produced from the originally infected cells. Host factors necessary for late influenza virus replication were identified by either measuring the presence of a virus-specific protein 24 h after infection, or by transferring the supernatant of the infected cells onto an uninfected reporter cell line. By using this approach, nearly 300 host genes were found to positively influence influenza virus proliferation (Karlas et al., 2010).

2.5.2 *Host Response Readouts and Their Applications to Viral Pathogenesis*

Measurement of the host response often focuses on the inflammatory pathways that are activated following viral infection. Generally, the focus is on the Toll-like receptor, RIG-I, and interferon (IFN) signaling pathways, as these are well-defined viral innate response pathways (see Chapter 4, Innate Immunity). These pathways converge on NF-κB, IRF3, and IFN-β transcription, and binding sites for these transcription factors can be attached to a luciferase reporter gene so that activation of the signaling pathway can be quantified. Additionally, the translocation of these transcription factors from the cytoplasm to the nucleus, which also signals activation of the pathway, can be measured through high-content imaging. Further, the host response can be measured by quantification of protein production, gene expression, or receptor up-regulation.

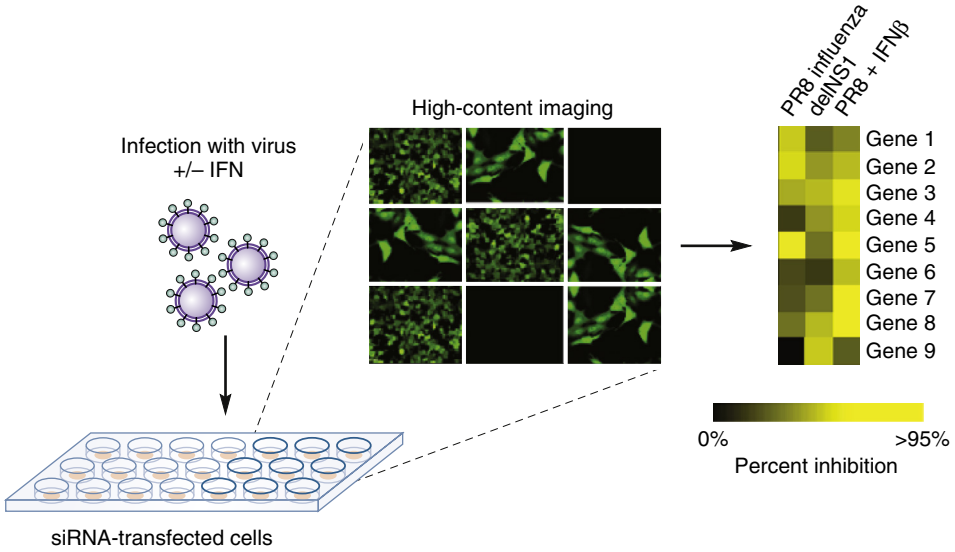

FIGURE 2 siRNA screen to measure the effects of ISGs on influenza virus replication. Cells are transfected with siRNA pools, and after infection with wild-type influenza virus (PR8), influenza virus with a deletion in the host immune inhibiting NS1 gene, or wild-type virus in the presence of IFN-β, cells are immunostained for hemagglutinin (HA). HA, which is an influenza-encoded protein that can be used as a surrogate measure of viral replication levels, is quantified using imaging technologies after RNAi treatments. Inhibition of viral replication after gene depletion reveals host factors that are required by the virus for replication. Prior stimulation with interferon enables an understanding of the virus–host interaction in cells that are in a preexisting antiviral state. The heat map reflects potential effects of knockdown of individual genes on the replication levels of wild-type influenza virus; influenza virus with a deletion in the NS1 gene; or wild-type influenza in cells pretreated with IFN-β.

As an example, in a search for genes needed for the induction of RIG-I, IFN-β signaling was used as a readout for the host response to Sendai virus (SeV) infection (Baril et al., 2013). This screen incorporated a genome-wide lentiviral-based shRNA library and cells expressing a luciferase reporter gene under the control of an IFN-β promoter. After SeV infection, those genes that either increased or decreased IFN-β activation were identified and were further mapped to the RIG-I signaling pathway using follow-up assays. In this way, WNT family members were identified as novel negative regulators of RIG-I signaling.

2.6 Integrating Data from Multiple Screens

RNAi screens usually generate an abundance of data and hits. Perhaps surprisingly, when similar screens are performed by different groups, there is often little overlap in the hits from each screen. It can therefore be difficult to determine genes that may be valuable for follow-up studies or therapeutic targeting. In the case of the three RNAi screens described above—which together identified over 700 genes that potentially impact influenza virus replication—there was only *one* common gene hit. The differences in the outcomes of these screens are likely due to variability in the type of cells used, the readout used, the type of virus, and the time points used for the assays.

Although these results may at first appear discouraging, a closer look at the data reveals that certain cell signaling pathways and protein complexes are overrepresented by the hits in these screens. A *meta*-analysis uncovered several common groups of host factors that are involved in processes crucial to influenza proliferation (Stertz and Shaw, 2011). Most of these overlaps occur at the gene pathway and protein complex level rather than at the level of individual genes.

2.7 Transcriptional Profiling

Viral infection has a dramatic effect on the expression, splicing, and turnover of cellular mRNAs. The full extent of these changes is beginning to be elucidated through RNA-seq, which is a method that builds upon the advances made by next-generation sequencing (see Chapter 11, Systems Virology). RNA-seq provides a snapshot of the global cellular transcriptome at the time of sample collection.

As described above, RNAi screens can illuminate the signaling pathways that are hijacked by a virus. In contrast, RNA-seq reveals the impact of infection on cellular transcription. Historically, DNA microarrays have served this purpose; however, the advantage of using RNA-seq is that all RNA within a cell population can be quantified. RNA-seq can therefore decipher splice variations and noncoding transcripts, which exponentially increases the information provided by each sample. However, sampling at multiple time points is usually necessary to provide the desired genomic information, and detailed kinetic studies can quickly become expensive.

RNA-seq can be used to simultaneously analyze both host and viral transcription, which is particularly informative

for DNA viruses with large genomes. For example, this approach was used to analyze the expression patterns of both murine cytomegalovirus and infected mouse embryonic fibroblasts (Juranic Lisnic et al., 2013). The sequencing of samples from nine different time points after infection revealed many novel viral transcripts, including antisense transcripts, spliced transcripts, and transcripts that overlapped multiple annotated genes. In fact, the most highly abundant transcripts did not have known functions. On the host side, many of the most up-regulated genes were associated with transcription and cellular defense. Down-regulated genes not previously known to have a role in infection were also identified.

Because RNA-seq can be used to analyze noncoding RNAs, it is beginning to be used to investigate the roll of such RNAs in viral infection. Certain miRNAs have already been demonstrated to play roles in viral persistence or host defense; however, a comprehensive picture of the miRNAs that are present or induced during viral infection is lacking. Further, most miRNAs still have unknown significance in the context of infection. Deep sequencing of RNA isolated from the lungs of four different mouse strains that had been infected with SARS coronavirus or influenza virus revealed extensive differential expression of diverse classes of short noncoding RNAs (Peng et al., 2011). Such studies are providing a greater understanding of overall transcriptional changes due to viral infection and provide an important resource for determining the functional role of noncoding RNAs in the antiviral response.

3. PHYSICAL PROTEIN INTERACTION SCREENS

Proteins typically act in complexes that direct their specificity, activity, localization, and interactions with other protein complexes and cellular machinery. It is therefore desirable to comprehensively identify members of protein complexes in order to enhance the functional and mechanistic knowledge of individual proteins. For uncharacterized proteins lacking functional annotation, identifying interacting proteins can be used to assign functions through testable "guilt-by-association" hypotheses. For well-characterized proteins, an unbiased interactome characterization may uncover previously unknown functions.

Of particular relevance to viral pathogenesis, the identification of physical interactions between virus and host proteins provides putative targets for therapeutic intervention. Furthermore, host proteins that interact with viruses are frequently under strong selective pressure and thus may be unable to escape antiviral drugs, in contrast to rapidly evolving viral proteins. In any case, the identification of virus–host protein–protein interactions identifies viral vulnerabilities and dependencies on the host cell.

Screens to identify protein–protein interactions can be divided into two categories: those that identify direct, physical interactions, and those that identify components of a protein complex. Screens identifying direct, physical interactions frequently use yeast two-hybrid methods or other complementation assays. Protein arrays may also identify direct, physical interactions as well as enzyme–substrate relationships. Screens that identify components of a protein complex frequently rely on protein co-purification using co-immunoprecipitation or affinity purification. Each of these types of screens is discussed in the remainder of this chapter.

3.1 Two-Hybrid Screening

Two-hybrid screening approaches are high-throughput complementation assays that test for protein–protein or protein–DNA interactions. The assay is typically performed by introducing proteins of interest pairwise into yeast, with each protein fused to a transcription factor that has been split into two complementary fragments. Conventionally, the protein fused to the N-terminal DNA-binding domain of the transcription factor is referred to as the "bait" and the protein fused to the C-terminal activation domain as the "prey." When brought into close proximity to one another through interaction between the bait and prey proteins, the binding and activation domains of the transcription factor function to activate transcription of a reporter gene. The reporter gene may encode for antibiotic resistance, such that interacting clones can be selected by applying antibiotic pressure. Alternatively, the reporter gene may code for a lethal gene, such that a physical interaction results in a reduction in colony size. Although two-hybrid approaches are typically performed using yeast, these assays have also been adapted to bacterial and mammalian systems (Joung et al., 2000). The disadvantages of two-hybrid approaches are that they often have high false-positive and false-negative rates. Producing proteins at far higher abundance than is biologically relevant can lead to spurious interactions that elevate the false-positive rate. False negatives may occur if N- or C-terminal fusions disrupt interaction interfaces, or if proper protein folding, processing, or posttranslational modifications cannot be recapitulated.

Two-hybrid approaches have been used to comprehensively characterize interactions between host proteins and proteins derived from a variety of viruses, including Kaposi sarcoma-associated herpesvirus, varicella-zoster virus, murine γ-herpesvirus 68 (MHV-68), vaccinia virus, SARS coronavirus, influenza virus (Friedel and Haas, 2011). In the case of influenza virus, an integrated approach was used to identify and validate interactions between viral and human proteins by complementing a comprehensive yeast two-hybrid assay with additional large-scale experiments (Shapira et al., 2009). This included the measurement of

cellular transcriptional responses following transfection with influenza viral RNA, IFN-β treatment, and infection with an influenza strain lacking the NS1 gene (responsible for inhibiting the innate immune sensing of viral RNA and downstream IFN production). A set of genes found to be regulated in either the two-hybrid screen or the gene expression screens were also tested in siRNA knockdown screens measuring influenza replication and IFN-β production. Integrating the data resulting from these various assays revealed that viral polymerase subunits were enriched for interactions resulting in the positive regulation of IFN production, suggesting that the viral polymerase, in addition to NS1, plays a role in inhibiting the IFN response.

A similar integrated approach was used to characterize virus–host interactions for murine γ-herpesvirus MHV-68 (Lee et al., 2011). Using a yeast two-hybrid approach, a library of 84 MHV-68 genes was screened against each other to identify 23 intraviral interactions. The library was also screened against a cDNA library derived from human liver cells to identify 243 virus–host interactions. An affinity purification approach validated 70% of the intraviral interactions, giving an estimate of the false-positive rate of the yeast two-hybrid screen. Network analyses indicated that cellular proteins targeted by MHV-68 had more partners in a cellular protein–protein interaction network than expected by chance. This integrated screening and validation approach therefore yielded viral–viral and viral–host protein interaction networks.

3.2 Protein Microarray Screening

Protein microarrays are constructed by immobilizing proteins at high density on a solid surface. The proteins may be individually purified, or they may be synthetic peptides generated using chemical peptide synthesis. The task of cloning and purifying thousands of native, full-length proteins to immobilize on an array may seem an insurmountable task, but arrays containing an impressive 17,000 full-length human proteins have been constructed (Hu et al., 2012). Proteins immobilized at addressable locations on a microarray can be used to identify not only protein–protein interactions, but also interactions with nucleic acids, antibodies, and small molecules.

The most cited limitation of protein microarrays relates to the comprehensiveness of the protein libraries available on the chip. In addition, proteins immobilized on microarrays are typically produced in bacteria or yeast and are therefore prone to false-negative interactions due to incorrect folding or lack of posttranslational modifications (similar to that observed using two-hybrid assays).

Protein microarrays have been used to identify conserved substrates for viral kinases that may represent targets for antiviral drugs. Herpes simplex virus, human cytomegalovirus, Epstein–Barr virus, and Kaposi's sarcoma-associated herpesvirus each encode a serine/threonine kinase that is necessary for viral replication and spread. To determine the extent to which substrates for these kinases are conserved, a human protein microarray was used to identify the cellular substrates of each kinase (Li et al., 2011). This approach resulted in the identification of 643 nonredundant substrates, 110 of which are shared by at least 3 kinases. The shared substrates were then mapped onto a network of existing data for protein–protein interactions, enzyme–substrate relationships, and gene ontology functional classes resulting in the identification of a highly connected cluster of DNA damage response proteins.

3.3 Affinity Purification and Mass Spectrometry Screening

Modern mass spectrometry platforms are increasingly sensitive and capable of characterizing complex protein mixtures at unprecedented depth. These platforms can be used to identify protein complexes that are formed in vivo and purified intact, thus the approach can identify protein interactions as they occur in a natural biological system. A major advantage of co-purification approaches is that they identify many proteins contained within a complex, making the process of identifying complexes and pathways associated with a protein of interest a much simpler task than with two-hybrid or protein microarray approaches (which rely on bioinformatics approaches and public databases to identify complexes and prioritize interactions). Affinity purification coupled with mass spectrometry, termed AP-MS, has been used to characterize a wide range of biological systems, including virus–host interactions.

Although many biological systems are amenable to AP-MS analysis, the use of affinity-tagged proteins introduces an inherently synthetic aspect to the assay. The most common method of introducing a tagged protein into a cell is by transfection, a method that does not work well for many types of primary cells. Furthermore, AP-MS requires a fairly large amount of starting material, which is not scalable for many cell types and primary cell systems, and AP-MS screens do not identify direct, physical interactions. Entire complexes are co-purified such that the specific interactions between members of a complex may be obscured. Cross-linking approaches combined with AP-MS can overcome this problem by providing distance constraints between proteins within a complex. Cross-linking AP-MS approaches are still highly specialized, particularly with respect to the bioinformatics interpretation of the spectra of cross-linked peptides.

AP-MS has been used to identify HIV–human protein interactions. For this approach, all the genes associated with the HIV genome, as well as unprocessed polyproteins, were cloned into a vector that contained a dual affinity tag fused

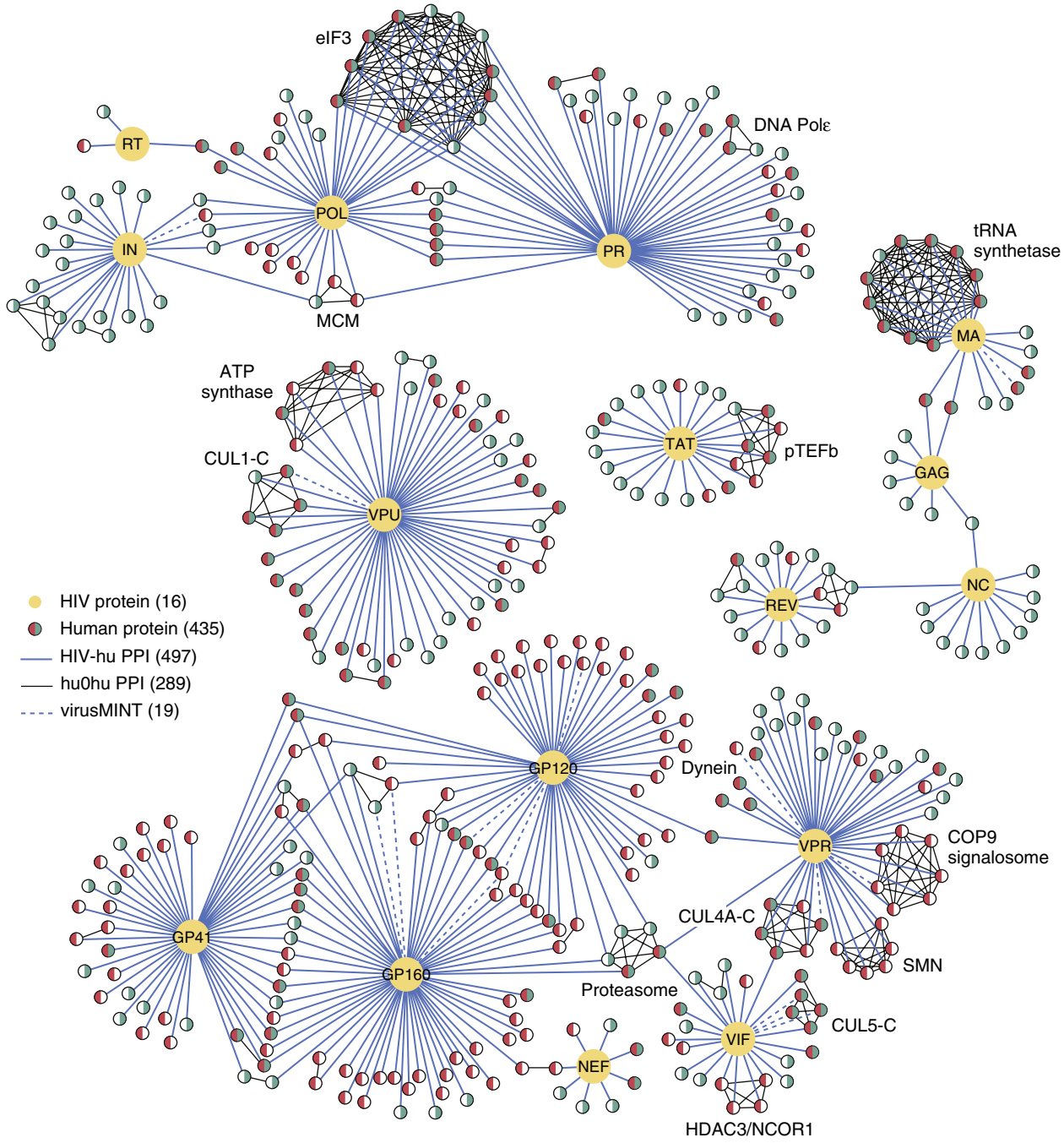

FIGURE 3 Network representations of HIV–human protein interactions. Yellow nodes correspond to viral proteins and all other nodes represent human factors (red indicates the interaction was identified in Jurkat cells and blue indicates the interaction was identified in HEK293 cells). Blue edges between nodes correspond to the HIV–human protein interactions that were identified whereas black edges represent connection between human proteins derived from several protein–protein interaction databases. *Adapted from Jäger et al. (2011), with permission.*

at the C-terminal end of the proteins (Jäger et al., 2011). Each clone was transiently transfected into HEK293 cells and was also used to generate stably expressed, tetracycline-inducible versions in Jurkat cells. HIV–human protein complexes were purified by affinity purification and the resulting complexes were analyzed by mass spectrometry. An unsupervised scoring system was then used to identify host–pathogen protein–protein interactions and to separate nonspecific from specific interactions. Using this score, a high-confidence interaction map was generated and overlaid with human–human protein interactions and HIV–human genomics data (Figure 3).

4. DATA INTEGRATION AND NETWORK ANALYSIS

Data integration is crucial to recognizing signaling pathways and nodes of activity that mediate the interactions between virus and host. This integration can be performed through side-by-side analysis of multiple loss-of-function screens, or it can be done by integrating orthogonal datasets. Protein–protein interaction studies provide clear evidence of which individual proteins interact with each other. Network representations then allow host–pathogen interactions to be visualized.

Protein–protein interaction data can also be applied to genetic screens to better understand the functional significance of each gene (Shapira et al., 2009). However, individual gene arrays make it more difficult to determine genetic interactions, or how the absence of two or more genes will affect the virus–host interactome. Genetic interactions can further reveal whether two genes function independently or in conjunction with each other. For example, if knocking down two genes independently results in reduced viral replication, there is no easy way to determine whether it is because both genes are in the same pathway or because the genes are in pathways that act in parallel. If both genes are knocked down at the same time, and the reduction in viral replication is additive, then it is likely that these genes are in independent signaling pathways. If viral replication is reduced to a similar level as either gene alone, then the genes are likely in the same pathway. Predictions can be made based on protein–protein interactions, but designing studies that systematically test combinations of genes is the most effective method to determine the genetic and functional relationships between these target genes.

Although data integration and analysis is challenging, there have been many efforts to make high-throughput data more accessible and universal (Masseroli et al., 2014). Close interactions between virologists and biostatisticians are necessary to develop proper analysis tools. The quest to elucidate the virus–host interactome has already benefited greatly from open-source databases. For example, Gene Ontology (GO) is a bioinformatics database that allows researchers to categorize their gene lists into functional groups, and GO data can be complemented with commercial (e.g., Ingenuity Pathway Analysis) or open-source (e.g., Cytoscape) network analysis programs. These programs are used to visualize clusters of activities and provide an idea about how the genes are interrelated. Using screening data, GO, and known protein interactions, a predicted interaction network can be generated. Importantly, the cumulative results from genome-wide screens provide a valuable resource for the field and are being used to create a clearer picture of virus–host interactions.

5. REPRISE

Identifying all of the viral and cellular factors that impact viral infection, replication, and pathogenesis is a monumental task. Although a truly comprehensive determination of every such factor is currently beyond reach, a variety of experimental approaches are being used to work toward this goal. Gene-based approaches include loss-of-function screening using siRNAs or shRNAs, and screens of this type have identified hundreds of human host factors required for influenza virus replication. Newer CRISPR-based approaches, which provide improved targeting and stable and complete gene knockouts, are also being developed for use in high-throughput screens. Gain-of-function screens complement these approaches by providing the ability to identify host antiviral genes that may be down-regulated by the infecting virus. RNA-seq, which can identify and quantify entire viral and cellular transcriptomes, is yielding new views into the complexity of transcription and the incredible diversity of coding and noncoding RNA transcripts. Many RNAs that no one had ever before thought to look for are likely to have roles in the virus–host interactome.

Protein-based approaches provide information on how viral and host factors interact with one another and how such interactions contribute to infection outcome. Two-hybrid screens and protein microarrays are used to identify direct, physical interactions between proteins, or between proteins and nucleic acids, or other small molecules. Affinity purification and mass spectrometry complement these approaches by identifying the members of protein complexes. Together, these approaches are being used to identify cellular substrates for viral enzymes and to construct detailed network models of the interactions between viral and host proteins.

In this chapter, the focus has been on identifying the genes and proteins, and the protein–protein interactions, which contribute to the virus–host interactome. As we will see in the following chapters, even the most comprehensive lists of genes and proteins will be only part of the overall roster of factors involved in viral pathogenesis. Viruses also impact host metabolic processes and interact with host metabolites. In addition, players beyond the virus and host—the components of the host microbiome—contribute to infection outcome. In the following chapter, we take up the contribution of host genetics to virus–host interactions and see how components of the virus–host interactome may therefore differ between individuals, with sometimes dramatic consequences.

REFERENCES

Baril M, Es-Saad S, Chatel-Chaix L, et al. Genome-wide RNAi screen reveals a new role of a WNT/CTNNB1 signaling pathway as negative regulator of virus-induced innate immune responses. PLoS Pathog 2013;9:e1003416.

Barrangou R, Fremaux C, Deveau H, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science 2007;315:1709–12.

Brass AL, Huang IC, Benita Y, et al. The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. Cell 2009;139:1243–54.

Friedel CC, Haas J. Virus-host interactomes and global models of virus-infected cells. Trends Microbiol 2011;19:501–8.

Hu CJ, Song G, Huang W, et al. Identification of new autoantigens for primary biliary cirrhosis using human proteome microarrays. Mol Cell Proteomics 2012;11:669–80.

Jackson AL, Linsley PS. Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. Nat Rev Drug Discov 2010;9:57–67.

Jäger S, Cimermancic P, Gulbahce N, et al. Global landscape of HIV-human protein complexes. Nature 2011;481:365–70.

Joung JK, Ramm EI, Pabo CO. A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. Proc Natl Acad Sci USA 2000;97:7382–7.

Juranic Lisnic V, Babic Cac M, Lisnic B, et al. Dual analysis of the murine cytomegalovirus and host cell transcriptomes reveal new aspects of the virus-host cell interface. PLoS Pathog 2013;9:e1003611.

Kanasty R, Dorkin JR, Vegas A, et al. Delivery materials for siRNA therapeutics. Nat Mater 2013;12:967–77.

Karlas A, Machuy N, Shin Y, et al. Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. Nature 2010;463:818–22.

Koike-Yusa H, Li Y, Tan EP, et al. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nat Biotechnol 2014;32:267–73.

Konig R, Stertz S, Zhou Y, et al. Human host factors required for influenza virus replication. Nature 2010;463:813–7.

Lee S, Salwinski L, Zhang C, et al. An integrated approach to elucidate the intra-viral and viral-cellular protein interaction networks of a gamma-herpesvirus. PLoS Pathog 2011;7:e1002297.

Li R, Zhu J, Xie Z, et al. Conserved herpesvirus kinases target the DNA damage response pathway and TIP60 histone acetyltransferase to promote virus replication. Cell Host Microbe 2011;10:390–400.

Masseroli M, Mons B, Bongcam-Rudloff E, et al. Integrated bio-search: challenges and trends for the integration, search and comprehensive processing of biological information. BMC Bioinformatics 2014;15(Suppl. 1):S2.

Panda D, Cherry S. Cell-based genomic screening: elucidating virus-host interactions. Curr Opin Virol 2012;2:784–92.

Peng X, Gralinski L, Ferris MT, et al. Integrative deep sequencing of the mouse lung transcriptome reveals differential expression of diverse classes of small RNAs in response to respiratory virus infection. mBio 2011;2:00198–211.

Schoggins JW, Wilson SJ, Panis M, et al. A diverse range of gene products are effectors of the type I interferon antiviral response. Nature 2011;472:481–5.

Shapira SD, Gat-Viks I, Shum BO, et al. A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. Cell 2009;139:1255–67.

Stertz S, Shaw ML. Uncovering the global host cell requirements for influenza virus replication via RNAi screening. Microbes Infect 2011;13:516–25.

Varble A, Benitez AA, Schmid S, et al. An in vivo RNAi screening approach to identify host determinants of virus replication. Cell Host Microbe 2013;14:346–56.

Yang D, Buchholz F, Huang Z, et al. Short RNA duplexes produced by hydrolysis with *Escherichia coli* RNase III mediate effective RNA interference in mammalian cells. Proc Natl Acad Sci USA 2002;99:9942–7.

Zender L, Xue W, Zuber J, et al. An oncogenomics-based in vivo RNAi screen identifies tumor suppressors in liver cancer. Cell 2008;135:852–64.

Zhou Y, Zhu S, Cai C, et al. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. Nature 2014;509:487–91.