Pearls

Identifying Host Factors That Regulate Viral Infection

Tien-Huei Hsu, Katherine R. Spindler*

Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan, United States of America

The Host Side of Viral Infection

One goal of virology research is to identify viral and host factors involved in infection, in order to develop antiviral therapies. Drugs targeting viral proteins have certain key disadvantages. They often affect only a specific viral species or subtype. Also, the low-fidelity polymerases of many medically important viruses, including HIV and influenza, make them prone to rapid mutations, leading to development of drug resistance. In addition, viruses encode few proteins, limiting the number of available targets.

Targeting host proteins is a practical alternative. Viruses use host proteins at multiple stages of their life cycles. Identifying host functions subverted by viruses will further our understanding of viral life cycles and may provide a catalog of novel drug targets that are unlikely to mutate following therapy. Furthermore, targeting the host may result in therapies with a broader range than traditional antivirals. Exciting progress has been made in recent years in this field; the development of new genomic and proteomic tools enables identification of interacting host factors at an unprecedented scale and level of detail. Together with the use of bioinformatics, these approaches hold promise for accelerating our understanding of virus–host interactions.

Genomics Techniques to Identify Host Factors

Host genetic background can significantly influence the outcome of viral infection. Genetic studies identify host factors required for successful viral infection through phenotypic effects such as susceptibility. The ability to manipulate experimental animals has expanded our knowledge of host factors involved in infection. For example, inbred mice that exhibit inherent phenotypic differences in their susceptibility profiles can be bred to generate progeny whose genotypes and phenotypes can be determined. Linkage analysis tools can then be used to identify a candidate region, and potential disease susceptibility genes can be prioritized for positional cloning.

Through genetic mapping, mouse cytomegalovirus (MCMV) susceptibility was determined to be associated with the loss of an activating natural killer cell receptor [1]. A genetic approach was also used to identify the *Flv* gene, subsequently identified as *Oas1b*, a member of the OAS/RNASEL innate immune system, which is responsible for controlling resistance to West Nile virus infection in mice [2]. A quantitative trait locus (QTL) strongly linked to susceptibility to mouse adenovirus type 1 was identified and reduced rapidly from an 18-Mb region to only 0.75 Mb through positional cloning involving backcross mice, polymorphic markers, and single nucleotide polymorphism haplotype identity [3]. Each of these studies began with the identification of mouse strains with differing susceptibilities to infection. However, due to the small number of inbred mouse strains and the limited genetic diversity of currently available strains, researchers are not able to achieve strong mapping resolution initially and must use additional methods to identify candidate genes. Optimally, researchers will be able to map genetic loci at a resolution that allows identification of individual genes, eliminating the steps of candidate gene prioritization.

To develop a genetically diverse panel of inbred mouse strains to increase mapping resolution, a community effort has created the Collaborative Cross (CC) [4]. In a recent study, 44 pre-CC mouse strains were used to identify 21 QTLs associated with regulation of host response to influenza infection [5]. Pre-CC mice are in the process of becoming inbred CC strains; this study clearly demonstrates that CC mice have greater phenotypic diversity than standard inbred mouse strains. Pre-CC mice were also used to create Diversity Outbred (DO) mice [6]. DO mice are maintained through outcrossing to maintain allelic diversity; CC mice are inbred to generate stable clones. Complementary use of CC and DO mice will allow researchers to identify genes important in complex traits such as susceptibility to viral infection. However, these strategies rely on identifying pre-existing variants in host susceptibility genes.

In contrast, novel germline mutations can be created using mutagens, such as *N*-ethyl-*N*-nitrosourea [7]. MCMV-resistant mice were mutagenized and selected for susceptibility to MCMV. Genes associated with resistance were then identified through positional cloning and sequencing. This same approach was recently used to identify a mouse gene, *Eif2ak4* (encoding GCN2), involved in susceptibility to MCMV and human adenovirus [8]. The advantage of this approach in identifying new host factors and pathways is that it is unbiased and does not make assumptions of the genes involved.

Efforts to determine human homologs of susceptibility genes identified in mouse models are underway to translate these findings to human disease. Mouse studies are an important starting point for uncovering virus–host interactions, especially when orthologous human genes are present. However, human populations are outbred, and variations in response to viral infection are expected, resulting in less than clear interpretation of results. In humans, genome-wide linkage analysis studies have been limited to chronic infectious diseases, due to the difficulty in recruiting families with multi-case acute viral infections. A whole genome scan conducted with Gambian families identified a major susceptibility locus to chronic hepatitis B infection that contains a cluster of cytokine receptor genes [9]. Family-based linkage studies have the additional disadvantage of having low power in identifying genes involved in susceptibility to viral diseases that involve the complex interaction of multiple genes. In addition,

family members share many genes, making it difficult to identify the relevant genes involved in viral susceptibility.

Genome-wide association studies (GWAS) have been used to identify human susceptibility loci. Whole genomes of a large human population can be scanned to identify genetic variations frequently associated with susceptibility to infection by a particular pathogen or with severity of disease. The HLA-viral peptide interaction was identified through GWAS as a major genetic factor responsible for HIV control [10]. GWAS require a large sample size and can suffer from sample-selection biases of cases and controls. These studies also have limited ability to detect variants with small effect or low frequencies. However, next generation sequencing will facilitate identification of rare mutations associated with host susceptibility.

Direct Protein-Based Techniques to Identify Host Factors

Many methods can be used to identify physical interactions between viral and host proteins. One of the earliest of these was co-immunoprecipitation of viral and cellular protein complexes with specific antisera to viral and host proteins. The tumor suppressor protein p53 was first identified by co-immunoprecipitation in complexes with adenovirus E1B 55 kDa protein and in complexes with SV40 large T antigen [11]. The tumor suppressor protein Rb co-immunoprecipitates with adenovirus E1A protein [12]. These findings provided critical evidence that oncogenic viruses promote tumorigenesis by inactivating tumor suppressor proteins. However, co-immunoprecipitation is performed in vitro and may not accurately represent the interaction of proteins in vivo. In addition, weak or less stable interactions may be overlooked.

Additional techniques used to detect interactions of viral and host proteins include yeast-two-hybrid (Y2H), tandem affinity purification, virus overlay protein binding assay (VOPBA), glutathione S-transferase protein purification, and co-immunoprecipitation followed by mass spectrometry analysis. Y2H is amenable to high-throughput screening, and genome-scale Y2H studies have identified host–viral protein interactions for a variety of viruses, including Epstein-Barr virus, HIV, influenza virus, vaccinia virus, Moloney murine leukemia virus, and hepatitis C virus [13]. When the Y2H approach is adapted to high-throughput format, a single "bait" can be tested against multiple "preys" for physical interaction. VOPBA is a screen for interacting proteins using electrophoresis of cellular contents, followed by blotting to a membrane and "probing" with virus. VOPBA has been used to identify virus receptors for human adenovirus [14], respiratory syncytial virus [15], lymphocytic choriomeningitis virus, and Lassa fever virus [16]. Results from Y2H and VOPBA can be validated by co-immunoprecipitations of co-transfected proteins, but the techniques are limited to direct protein–protein interactions.

Gene silencing techniques can assist in defining effects of cellular factors on viral infection that are both direct and indirect. These include host factors (i) that interact directly with viral proteins, (ii) that are present in viral–host protein complexes, (iii) that bind to non-protein components of viruses, and (iv) that are involved in signaling pathways, other cellular processes involved in viral infection, and host immunity. Genome-scale RNA interference (RNAi) screening is a high-throughput method used to investigate diverse biological processes, including host factors involved in viral pathogenesis. Large-scale RNAi screens have been used to identify host factors for a number of important human viruses, including HIV, hepatitis C virus, influenza virus, West Nile virus, and dengue virus [13]. However, because it is technically challenging to develop complete RNAi libraries of the human genome, important candidates may be missed [17]. RNAi screens are highly sensitive to experimental variation, and the overlap of positive hits between similar studies can vary [18]. Also, because RNAi screens are resource intensive, often few time points are examined, limiting knowledge of dynamic changes during viral infection.

Molecular imaging techniques are increasingly being used to visualize transient or dynamic interactions. Live cell imaging microscopy techniques have advanced significantly, allowing detection of single molecules in the absence of artifacts caused by fixation methods. Events of influenza entry have been dissected using real-time microscopy, providing new insights into cellular endocytic pathways [19]. Two different host proteins that interact with the Sindbis virus at different stages of infection were identified using a GFP-tagged viral protein, further demonstrating the usefulness of imaging approaches [20]. Major considerations of this technique are the maintenance of constant physiological conditions (e.g., temperature and pH), and the prevention of photobleaching of dyes. These may prove challenging following extended imaging. However, the ability to monitor rapidly changing interactions may provide critical insight into viral processes that are not readily measured using other methods.

Data Repositories

Data generated from high-throughput techniques have furthered our understanding of the virus–host interface, and efforts are being made to identify and analyze candidate drug targets. To maximize the benefits of these screens, data need to be accessibly stored and modeled into networks. Several online repositories, including VirHostNet [21], VirusMINT [22], and BiologicalNetworks [23], enable modeling of current data to gain broad understanding of protein and gene networks involved in viral infection. Multi-scale data integration approaches allow for simultaneous analysis of different datasets, such as phylogeny, literature searches, virulence, and epidemiological data. However, there has been no standardization of where data should be deposited, and participation is voluntary.

Concluding Remarks

Various techniques have facilitated identification of host factors involved in viral infection. Verification of these candidates through biochemical, genetic, and immunological methods may progressively become the rate-limiting step. Virologists will increasingly need to collaborate with other scientists to realize the full potential of the collected data. Use of simulations and models will enable better depiction of infection events. Structural biology can also be used to visualize protein interfaces at high resolution. The identification of host proteins through the many approaches described in this review is only a starting point for exploring function and mechanism, with the aim of uncovering cellular pathways affecting viral replication that can be targeted for drug development.

Acknowledgments

References

1. Webb JR, Lee SH, Vidal SM (2002) Genetic control of innate immune responses against cytomegalovirus: MCMV meets its match. Genes Immun 3: 250–262.

2. Samuel CE (2002) Host genetic variability and West Nile virus susceptibility. Proc Natl Acad Sci U S A 99: 11555–11557.

3. Welton AR, Chesler EJ, Sturkie C, Jackson AU, Hirsch GN, et al. (2005) Identification of quantitative trait loci for susceptibility to mouse adenovirus type 1. J Virol 79: 11517–11522.

4. Collaborative Cross Consortium (2012) The genome architecture of the collaborative cross mouse genetic reference population. Genetics 190: 389–401.

5. Bottomly D, Ferris MT, Aicher LD, Rosenzweig E, Whitmore A, et al. (2012) Expression quantitative trait Loci for extreme host response to influenza a in pre-collaborative cross mice. G3 (Bethesda) 2: 213–221.

6. Svenson KL, Gatti DM, Valdar W, Welsh CE, Cheng R, et al. (2012) High-resolution genetic mapping using the mouse diversity outbred population. Genetics 190: 437–447.

7. Crozat K, Georgel P, Rutschmann S, Mann N, Du X, et al. (2006) Analysis of the MCMV resistome by ENU mutagenesis. Mamm Genome 17: 398–406.

8. Won S, Eidenschenk C, Arnold CN, Siggs OM, Sun L, et al. (2012) Increased susceptibility to DNA virus infection in mice with a GCN2 mutation. J Virol 86: 1802–1808.

9. Frodsham AJ, Zhang L, Dumpis U, Taib NA, Best S, et al. (2006) Class II cytokine receptor gene cluster is a major locus for hepatitis B persistence. Proc Natl Acad Sci U S A 103: 9148–9153.

10. Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, et al. (2010) The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science 330: 1551–1557.

11. Sarnow P, Ho YS, Williams J, Levine AJ (1982) Adenovirus E1b-58kd tumor antigen and SV40 large tumor antigen are physically associated with the same 54 kd cellular protein in transformed cells. Cell 28: 387–394.

12. Harlow E, Whyte P, Franza BR, Jr., Schley C (1986) Association of adenovirus early-region 1A proteins with cellular polypeptides. Mol Cell Biol 6: 1579–1589.

13. Friedel CC, Haas J (2011) Virus-host interactomes and global models of virus-infected cells. Trends Microbiol 19: 501–508.

14. Wu E, Fernandez J, Fleck SK, Von Seggern DJ, Huang S, et al. (2001) A 50-kDa membrane protein mediates sialic acid-independent binding and infection of conjunctival cells by adenovirus type 37. Virology 279: 78–89.

15. Tayyari F, Marchant D, Moraes TJ, Duan W, Mastrangelo P, et al. (2011) Identification of nucleolin as a cellular receptor for human respiratory syncytial virus. Nat Med 17: 1132–1135.

16. Cao W, Henry MD, Borrow P, Yamada H, Elder JH, et al. (1998) Identification of alpha-dystroglycan as a receptor for lymphocytic choriomeningitis virus and Lassa fever virus. Science 282: 2079–2081.

17. Shan G (2010) RNA interference as a gene knockdown technique. Int J Biochem Cell Biol 42: 1243–1251.

18. Goff SP (2008) Knockdown screens to knockout HIV-1. Cell 135: 417–420.

19. Lakadamyali M, Rust MJ, Babcock HP, Zhuang X (2003) Visualizing infection of individual influenza viruses. Proc Natl Acad Sci U S A 100: 9280–9285.

20. Cristea IM, Carroll JW, Rout MP, Rice CM, Chait BT, et al. (2006) Tracking and elucidating alphavirus-host protein interactions. J Biol Chem 281: 30269–30278.

21. Navratil V, de Chassey B, Meyniel L, Delmotte S, Gautier C, et al. (2009) VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks. Nucleic Acids Res 37: D661–D668.

22. Chatr-aryamontri A, Ceol A, Peluso D, Nardozza A, Panni S, et al. (2009) VirusMINT: a viral protein interaction database. Nucleic Acids Res 37: D669–673.

23. Baitaluk M, Sedova M, Ray A, Gupta A (2006) BiologicalNetworks: visualization and analysis tool for systems biology. Nucleic Acids Res 34: W466–W471.