SCIENTIFIC DATA

A comprehensive collection of systems biology data characterizing the host response to viral infection

Brian D. Aevermann[1] *et al.* *

The Systems Biology for Infectious Diseases Research program was established by the U.S. National Institute of Allergy and Infectious Diseases to investigate host-pathogen interactions at a systems level. This program generated 47 transcriptomic and proteomic datasets from 30 studies that investigate *in vivo* and *in vitro* host responses to viral infections. Human pathogens in the *Orthomyxoviridae* and *Coronaviridae* families, especially pandemic H1N1 and avian H5N1 influenza A viruses and severe acute respiratory syndrome coronavirus (SARS-CoV), were investigated. Study validation was demonstrated via experimental quality control measures and meta-analysis of independent experiments performed under similar conditions. Primary assay results are archived at the GEO and PeptideAtlas public repositories, while processed statistical results together with standardized metadata are publically available at the Influenza Research Database (www.fludb.org) and the Virus Pathogen Resource (www.viprbrc.org). By comparing data from mutant versus wild-type virus and host strains, RNA versus protein differential expression, and infection with genetically similar strains, these data can be used to further investigate genetic and physiological determinants of host responses to viral infection.

Design Type(s)	time series design • dose response design • strain comparison design • species comparison design • genetic modification design
Measurement Type(s)	protein expression profiling • transcription profiling assay
Technology Type(s)	mass spectometry assay • DNA microarray
Factor Type(s)	genotype • pathogen • dose • timepoint • replicate
Sample Characteristic(s)	Homo sapiens • Mus musculus • lung adenocarcinoma cell line • lung • Calu-3 cell • Calu-3 cell (2B4 clonal cell line) • Human Airway Epithelial (HAE) cell

[1]J. Craig Venter Institute, La Jolla, CA 92037, USA. *A full list of authors and their affiliations appears at the end of the paper.

Correspondence and requests for materials should be addressed to R.S. (email: rscheuermann@jcvi.org)

Background & Summary

With the recognition that host responses to pathogen infection play key roles in disease severity and mortality, virologists have shifted toward integrated systems biology approaches to identify therapeutics that target host pathways[1,2]. To support the cross-disciplinary approaches necessary to address a systems-level analysis, the Division of Microbiology and Infectious Diseases (DMID) at the National Institute of Allergy and Infectious Disease (NIAID) established the Systems Biology for Infectious Diseases Research (SysBio) program that provided support for four centers from 2008–2013: Systems Influenza, Systems Virology, Systems Biology of Enteropathogens, and Mycobacterium tuberculosis Systems Biology[3]. The Systems Influenza and Systems Virology centers focused on elucidating the mechanisms of how viral regulation of the host cellular circuitry contributes to viral replication and disease severity, thereby elucidating host pathways that could serve as potential new therapeutic targets.

Several publications have reported the findings from these virology-focused SysBio centers. Overall characterization of the host responses demonstrated that pro-inflammatory interferon (IFN) signalling pathways were enriched following infection with either influenza A or Betacoronavirus infection *in vitro*[4,5] and *in vivo*[6,7]. These studies characterized several aspects of viral-host interactions that influence disease severity. Experiments designed to tease out the contribution of the virus and the host to the severity of host response reported that specific mutations in the H5N1 viral genome affected the kinetics and magnitude of the host response, rather than changes in specific host response factors[8]. Another study found that early response signatures to influenza virus infection correlated with the severity of late disease[9]. Lipidomics analysis of mice infected with various influenza A virus strains of varying pathogenicity have also been reported[10]. Similarly, the disease severity of SARS-CoV correlated with the dysregulation of the urokinase pathway and an increase in fibrinolysin pathway activity[7]. An integrated network interrogation approach using both transcriptomic and proteomic data sets predicted a subset of key regulator molecules responding to SARS-CoV and influenza infections that could be used as host targets for therapeutic intervention[11,12]. Genetic modification of non-structural protein 16 in SARS-CoV showed enhanced susceptibility to type I and III interferon responses, making it a good vaccine strain candidate[13].

To support the broad dissemination of research datasets from these SysBio centers and other supported programs, DMID established the Bioinformatics Resource Centers for Infectious Diseases (BRC) program to provide public database and analysis resources for the infectious disease research community. A data dissemination working group, consisting of representatives from NIAID, the SysBio centers, and the BRCs was established to develop a data management and release approach that would maximize the reuse potential of the generated data. The working group leveraged existing public archiving strategies to store the primary assay results and devised metadata standards to represent the experiment descriptions and interpreted results in a consistent and standardized way. The final derived host response data and associated structured standardized metadata were made publically available at the BRC appropriate for the system under study. As the virus-centric BRCs, the Influenza Research Database (IRD, www.fludb.org)[14] and the Virus Pathogen Database and Analysis Resource (ViPR, www.viprbrc.org)[15] host data produced by the Systems Influenza and Systems Virology centers.

Here we describe 47 transcriptomic and proteomic datasets from 30 studies generated by these SysBio centers and made available through public resources. These studies characterized the host response to infection by members of the *Orthomyxoviridae* and *Coronaviridae* virus families, including pandemic influenza A H1N1 virus, highly pathogenic H5N1 avian influenza (HPAI) virus, severe acute respiratory syndrome coronavirus (SARS-CoV), and Middle East respiratory syndrome coronavirus (MERS-CoV); all of which are capable of causing severe respiratory infections and pose significant threats to humans on a global scale[4,16]. Host responses were evaluated in either *in vivo* (mouse) or *in vitro* (human cell line) model systems using three major experimental designs: longitudinal time-course, dose response, and genetic modification (involving both genetically manipulated viruses and hosts) comparisons. Host responses to these manipulated variables were assessed using transcriptomic (gene expression microarray) and proteomic (mass spectrometry) assays. Together, this collection represents the first coordinated effort to create a systems level description of host-pathogen interactions using multiple viral strains, host models, and '-omics' technologies.

Methods

The experimental designs reported here include longitudinal (i.e. time course), dose response, and genetic modification of both virus and host strain comparisons. The intent of these designs was to assemble a unified view of the virulence mechanisms and replication strategies resulting from viral regulation of host cellular processes. For the present experiments, either transcriptomic (using gene expression microarrays) or proteomic (using liquid chromatography coupled with mass spectrometry) methodologies were used to collect the reported data. The materials used, sample preparation protocols, validation procedures, data processing, and hypothesis testing that were performed are described below. Throughout the Methods section, Study IDs are appended to indicate which studies used a given method. A summary of the overall experiment workflow and the experiment factor conditions that were compared is presented in Fig. 1. The relationship between the experiment workflow and the experimental metadata, analysis metadata, primary assay results, and derived data is also presented. Study designs for all experiments, including relevant repository identifiers, are summarized in Table 1 (available online only), with individual RNA
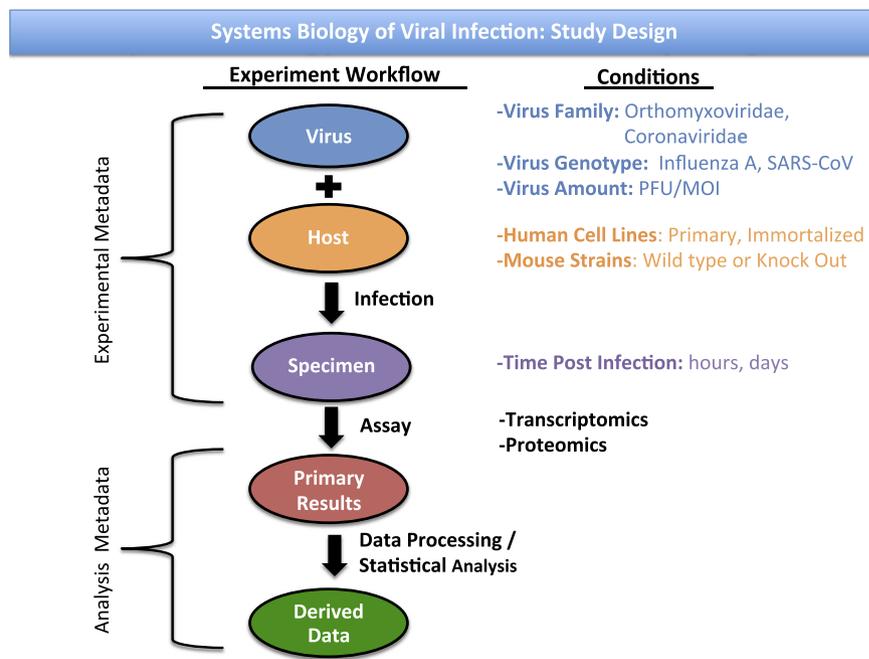
Figure 1. Overall study design for the study of the systems biology of viral infection. The ovals down the center represent steps in the experimental workflow common to all datasets. Text on the left gives the type of metadata required to describe the components of the workflow. On the right lists the experiment conditions investigated at each step throughout these datasets.

experiment samples detailed in Table 2 (available online only) and individual protein experiment samples detailed in Table 3 (available online only). Sample tracking from animal subjects to experiment samples is given in Table 4 (available online only).

Viruses

Within the studies reported here, the host response to viral agents from two major viral families, Orthomyxoviridae and Coronaviridae, are characterized. From Orthomyxoviridae, two influenza A subtypes H5N1 and H1N1 were chosen in order to investigate influenza viruses with higher and lower reported pathogenicity, respectively. Highly Pathogenic Avian Influenza H5N1 (HPAI) has been commonly referred to as 'Avian Flu' as it is endemic in bird populations. HPAI is known to exhibit 'stuttering transmission' in humans, with a relatively high mortality rates (>50%), but to-date has not acquired efficient human-to-human transmission. This high mortality rate together with its pathological similarities to the 1918 H1N1 'Spanish Flu' make it an important model to study and prevent future lethal pandemics. On the other hand, the more recent seasonal and pandemic H1N1 strains are considered to have lower pathogenicity, rarely causing severe disease or death. Since its emergence, pandemic H1N1 has been found to recur in the human population as a seasonal influenza. H1N1 influenza isolates used in these studies include recent pandemic strains, previous seasonal strains, and a reconstructed 1918 strain.

Virus preparations for the Influenza A/Vietnam/1203/2004 (H5N1) (abbreviated as VN1203) isolate were derived from a plasmid-based reverse-genetic system as previously described[17,18]. Four A/Viet Nam/1203/2004 mutant viruses that contain modified pathogenicity determinants were similarly generated by reverse genetics coupled with site-directed mutagenesis[17–19], including H5N1 VN1203-HAavir in which the hemagglutinin (HA) surface protein poly-basic cleavage site was mutated[20]; H5N1 VN1203-PB2627E which possesses a K to E amino acid substitution at residue 627 of the PB2 protein[21]; H5N1 VN1203-NS1trunc, in which a stop codon was introduced at amino acid 124 of the NS1 protein open reading frame[8]; and H5N1 VN1203-PB1F2del, in which the PB1-F2 open reading frame was eliminated[8]. Additional natural influenza strains used in these studies include: A/California/04/2009 (H1N1) (abbreviated as CA04), A/Netherlands/602/2009 (H1N1) (abbreviated as NL602; kindly provided by Ron Fouchier, Erasmus Medical Center, Rotterdam, Netherlands)[22], A/Mexico/4482/09 (H1N1) (Mex/4482), A/New Jersey/8/1976 (H1N1) (NJ/8/76), and A/Brisbane/59/07 (H1N1) (Brisbane/59/07)[6]. The reconstructed r1918 strain was also included in one study[23]. All virus strains were propagated and amplified in Madin-Darby canine kidney (MDCK) cells at 37° C in an atmosphere of 5% CO_2, except NJ/8/76, which was propagated in 10-day-old embryonated hens' eggs[6]. (CA04M001, ICL004, ICL006, ICL010, ICL011, ICL012, IM001, IM002, IM004, IM005, IM006A, IM006B, IM007, IM009, IM010, SBRI_AA_E1).

The PR8, X31, and VN1203(6+2) influenza A viruses were generated using the 8-plasmid influenza reverse genetics system[24]. All three viruses are based on PR8 (A/Puerto Rico/8/34), a mouse-adapted H1N1 strain originally derived from a human isolate. X31 is a mouse adapted H3N2 strain with the 6 internal genes of PR8 and the HA and NA derived genes from A/Aichi/2/1968. rgVN1203(6+2) (VN) contains the 6 internal genes of PR8 and the H5N1 HA and NA genes from A/Vietnam/1203/2004 with the polybasic cleavage site of the HA modified to restrict its replication to the airway[25,26]. (SBRI_AA_E1).

In addition to the influenza A strains, the studies reported here investigate recently characterized Coronaviridae strains. These strains are all members of the betacoronavirus genus including Severe Acute Respiratory Syndrome (SARS-CoV) and Middle Eastern Respiratory Syndrome (MERS-CoV) Coronaviruses. Both viruses cause severe acute respiratory illness, with SARS-CoV and MERS-CoV infection resulting in an overall mortality rate of ~10% and 38%, respectively. SARS-CoV was marked by its ease of transmissibility between humans, while MERS-CoV appears to be more limited in its human transmissibility. These viruses represent emerging pathogens not formerly recognized as a threat to human health, which adds an aspect of urgency to their research and characterization.

The SARS-CoV virus preparations used in these studies include a wild-type infectious clone derived from SARS-CoV (abbreviated icSARS-CoV), icSARS-CoV dORF6 in which accessory ORF6 was deleted, and Bat-SRBD, a reconstruction from a Bat-SCoV consensus genome with SARS-CoV receptor-binding domain (RBD)[27]. These strains were obtained from the Baric laboratory's existing infectious clone constructs as previously described[28–30]. Recombinant mouse-adapted SARS-CoV (MA15)[31], SARS-CoV dNSP16 (ref. 13), and human coronavirus Middle East Respiratory Syndrome Coronavirus strain EMC 2012 (abbreviated MERS-CoV) (from Bart L. Haagmans Erasmus Medical Center, Rotterdam, Netherlands) were also evaluated. All strains belonging to the *Betacoronavirus* genus were propagated in Vero E6 cells. For more details, see virus strain descriptions given in Table 5 (available online only). (ECL001, SCL005, SCL006, SHAE002, SHAE003, SHAE004, SM001, SM003, SM004, SM007, SM009, SM012, SM014, SM015, SM019, SM020).

Human cell lines, infections, and extractions

Most cell line studies involving infection with influenza A viruses were conducted in Calu-3 cells, a human lung adenocarcinoma cell line (kindly provided by Raymond Pickles; University of North Carolina, Chapel Hill, NC, USA). Cells were maintained in a 1:1 mixture of Dulbecco's modified Eagle's medium and Ham's F12 nutrient medium (DF12; Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum. For infection, Calu-3 cells were plated in 6 well plates at 1×10^6 cells per well. Cells were washed after 24 hours and infected 48 hours after plating. Infections were performed by removing culture medium, washing cells twice with 1x PBS, and inoculating each well with 300 μl of DF12 supplemented with 0.3% bovine serum albumin (DF12-BSA) containing virus at a multiplicity of infection of 1 to 3 plaque forming units per cell. Cultures were mock-infected by inoculating with DF12-BSA lacking virus. Cells were then incubated for 40 min at 37° C, and after the inoculum was removed, cells were washed with 1x PBS and cultured in 2 mL of DF12-BSA containing L-(tosylamido-2-phenyl) ethyl chloromethyl ketone (TPCK)-treated trypsin. At each sample collection time point, aliquots of medium were reserved for plaque assays (see below). Time zero samples were immediately harvested using the appropriate procedure depending on the assay to be performed.

Studies examining members of the *Betacoronavirus* genus were performed in a clonal population of Calu-3 cells sorted for high levels of expression of the SARS-CoV cellular receptor angiotensin-converting enzyme 2 (ACE2), referred to as Calu-3 2B4 cells (kindly provided by Chien K. Tseng, University of Texas Medical Branch, Galveston, TX)[32]. Cells were maintained in 1x MEM (Gibco catalog number 11095) supplemented with defined Fetal Bovine Serum (HyClone catalog number SH30070.03) and antibiotic/antimycotic (100x Gibco catalog number 15240). Infections of Calu-3 2B4 cells were performed in a similar manner to the above Calu-3 cells. In the case of Calu-3 2B4 cells, 1.5×10^6 cells per well were seeded with 3 mL of complete culture medium. Culture medium was refreshed after 24 hours and infections administered after 48 hours as above. (SCL005, SCL006).

A subset of studies involving infection with influenza A virus and members of the *Betacoronavirus* genus were performed in human airway epithelial (HAE) cells, which were obtained from airway specimens resected from patients undergoing surgery under University of North Carolina Institutional Review Board, with approved protocols by the Cystic Fibrosis Center Tissue Culture Core. Primary cells were plated at a density of 2.5×10^5 cells per well on permeable Transwell-COL (12-mm-diameter) supports. HAE cultures were generated over 6 to 8 weeks, forming well-differentiated polarized cultures that resembled *in vivo* pseudostratified mucociliary epithelium[33]. Infection of HAE cultures began by washing with 1x PBS followed by challenge with 200 μL of either viral agent or mock to the apical surface. Cultures were incubated at 37° C for 2 hours and the inoculum and unbound viruses were removed by washing three times with 1x PBS. Apical wash samples were harvested to analyze viral growth kinetics and titers by plaque assay in Vero E6 cells (see below). (SHAE002, SHAE003, SHAE004).

Isolation of RNA from cell cultures was performed as follows: cell monolayers were first washed with 5 mL of cold PBS and then lysed directly in the culture dish by adding 1 mL of TRIZOL reagent. Cells were pipetted up and down until cells had a uniform color and consistency. They were then transferred to a 2 ml polypropylene tube (O-ring, screw cap), vortexed thoroughly, and incubated for 5 minutes at room

temperature. Lysates were flash frozen on dry ice and stored at −80° C until shipping on dry ice. (ICL004, ICL006, ICL010, ICL011, ICL012, ECL001, SCL005, SCL006, SHAE002, SHAE003, SHAE004).

Isolation of protein from cell cultures was performed as follows: culture media was removed and cells were washed with 1 mL cold 150 mM (pH 8.5) ammonium bicarbonate buffer. This process was repeated for a total of 3 washes. 300 μL of cold 8 M urea dissolved in 50 mM ammonium bicarbonate buffer [pH 7.8] was then added directly to the well. Cells were scraped into a pre-sterilized 2 ml siliconized tube and incubated at room temperature for 15 minutes minutes to inactive viruses. Samples were then frozen in liquid nitrogen or dry ice/ethanol bath and stored at −80 °C until further processing. (ICL004, ICL006, ICL010, ICL011, ICL012, SCL005, SCL006).

Animal strains, infections, and extractions

The following mouse strains were obtained from the Jackson Laboratory (Bar Harbor, ME, USA): C57BL/6 J (stock no. 000664), Serpine1$^{-/-}$ (stock no. 002507), PLAT$^{-/-}$ (stock no. 002508), B6;129S6-$Ppp1r14c^{tm1Uhl}$/J (stock no. 013041), B6.129S2-$Tnfrsf1b^{tm1Mwm}$/J (stock no. 002620), B6.129S7-$Tnfrsf1b^{tm1Imx}$/J (stock no. 003243), B6.129S4-$Timp1^{tm1Pds}$/J (stock no. 006243), B6.129P2-$Cxcr3^{tm1Dgen}$/J (stock no. 005796), and B6.129-$Ido1^{tm1Alm}$/J (stock no. 005867). Young female BALB/c (Six-to eight-week-old) mice were purchased from Charles River Laboratories (Wilmington, MA, USA). All animal housing, care and treatment were conducted in accordance with the University of Wisconsin-Madison School of Veterinary Medicine Animal Care and Use Committee, or by the University of North Carolina Institutional Animal Care and Use Committee or under the guidance of the CDC's Institutional Animal Care and Use Committee in an animal facility accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International. For virus infections, mice were anesthetized by isofluorane inhalation (VN1203) or lightly anesthetized with ketamine/xylazine (SARS-CoV and CA04) and intranasally inoculated with either 50 μL of phosphate-buffered saline (PBS) alone (mock infection), or PBS containing a viral agent. Individual or group mouse body weights were collected on a daily basis to monitor the disease course, and mice were humanely euthanized upon reaching the experimental endpoint (i.e. sample collection or severe clinical symptoms). Weight loss significance was determined by Student's t test (Microsoft Excel), and significance in survival data was determined by the Mantel-Cox test (GraphPad). Both viral titers and qPCR assays were used to confirm the infection status of all mice (see below). For more details, see animal strain and cell line descriptions given in Table 6 (available online only).

To prevent RNA degradation, immediately after dissection, lung tissues were directly submerged in RNAlater stabilization solution (Ambion, Catalog number AM7021) following manufacturer's recommendations. Specifically, tissue was cut into small chunks (< 0.5 cm in any single dimension) and placed immediately into 10-20 volumes (w/v) of RNAlater. Samples were thoroughly immersed in RNAlater solution and incubated overnight at 4° C, followed by freezing at −80° C.

Homogenization for later RNA isolation was performed as follows. Lung tissues were thawed, removed from RNAlater, dabbed on a kimwipe to remove excess liquid, transferred to 1 ml TRIzol and homogenized using a TissueLyser or MagnaLyser using several 30 second bursts to prevent overheating of the sample. Samples were transferred to a clean microfuge tube and stored at −80° C until RNA isolation. (CA04M001, IM001, IM002, IM004, IM005, IM006A, IM006B, IM007, IM009, IM010, SM001, SM003, SM004, SM007, SM009, SM012, SM014, SM015, SM019, SM020).

Homogenization of tissue for protein isolation was performed as follows: lung tissue were rinsed in 50 mM ammonium bicarbonate to remove blood and other fluid, then frozen at −80 °C and stored until the next step. The tissue was homogenized in 2 ml of 8 M urea dissolved in 50 mM ammonium bicarbonate buffer [pH 7.8]. The lung homogenates were then centrifuged at low speed (e.g. 5000xg) to remove large pieces of debris. The supernatants were transferred to new 2 mL siliconized microcentrifuge tubes, and incubated at room temperature for 1 hour to inactivate viruses. Samples were then frozen in liquid nitrogen or dry ice/ethanol bath and stored at −80 °C until further processing. (CA04M001, IM001, IM002, IM004, IM005, IM006A, IM006B, IM007, SM001).

Alternatively, mice were anesthetized with 350 μL of Avertin (2, 2, 2-Tribromoethanol) administered intraperitoneally. The virus inoculum was administered intranasally in a total volume of 30 μL (10^5 PFU of virus in 1x PBS). After infection, mice were kept on a heating pad until they regained consciousness. At 12 hours following infection, mice were sacrificed by CO_2 inhalation. A hole was clipped in the trachea and bronchoalveolar fluid (BALF) collected. Lungs were perfused three times with 1 mL cold 1x PBS, for total of 3 mL. Whole lungs were then removed and homogenized in 1 mL TRIzol. (SBRI_AA_E1).

Viral titers and infection confirmation

Influenza viral titers were determined by standard plaque assay[34]. Briefly, MDCK cells were plated in 6-well plates, washed twice and then inoculated with 100 μL of 10-fold diluted virus in minimum essential medium supplemented with 0.3% BSA (MEM-BSA). Cells were incubated at room temperature for 1 hour, after which the inoculation solution was then removed. Cells were washed three times with PBS and overlaid with 1% Seaplaque agarose in MEM-BSA, supplemented with TPCK-treated trypsin (final 0.5 μg/mL). After the agarose solidified, plates were incubated upside down at 37° C and 5% CO_2 for 48 hours. To visualize and count plaques, all wells were fixed with 10% phosphate-buffered formalin overnight, after which the agarose plugs were removed and plates were dried. (CA04M001, ICL004,

ICL006, ICL010, ICL011, ICL012, IM001, IM002, IM004, IM005, IM006A, IM006B, IM007, IM009, IM010, SBRI_AA_E1).

Plaque assays for members of the *Betacoronavirus* genus were performed on supernatants collected at various times post-infection. Briefly, 5x10^5 Vero E6/Vero 81 cells were plated in 6-well plates. 10-fold dilutions of harvested viruses were generated in 1 mL of PBS. Then 200 μL of diluted virus was added to each well. Inoculated cells were incubated at 37° C for 1 hour (rocking gently every 15 minutes). Then 3 mL per well of agarose, consisting of 2x DMEM, 1.6% Seaplaque low melting point agarose, 10% final concentration fetal clone II for SARS-CoV and 4% for MERS-CoV, and 1x Antibiotic/Antimycotic was added to the cells. After solidification, plates were incubated face up at 37° C and 5% CO_2 for ~48 hours. 1x neutral red solution was added to the top of the agar and incubated for at least 2 hours. Neutral red was then removed and the plaques were counted. (ECL001, SCL005, SCL006, SHAE002, SHAE003, SHAE004, SM001, SM003, SM004, SM007, SM009, SM012, SM014, SM015, SM019, SM020).

In addition to viral titers, real-time PCR quantification of viral genomic and viral mRNA species was performed to confirm viral infection and for comparisons between time points both within an infection and across studies. For *Betacoronavirus* genus, first-strand cDNA synthesis was performed using 500 ng of total RNA and Thermo-Script reverse transcriptase (Invitrogen) according to the manufacturer's protocol. For influenza, first-strand cDNA synthesis was performed similar to above except 15 pmol of RNA-specific primers for VN1203, CA04 or NL602 nucleoprotein genomic RNA and mRNA and endogenous control were used to prime the reactions. Primers for NP RNAs were designed with unique 5′ sequences, thereby ensuring differentiation between genomic RNA and mRNA species. The reverse transcriptase reaction primer sequences used were: NP genomic RNA (VN1203, CA04, and NL602) - 5′GGCCGTCATGGTGGCGAATAGCAAAAGCAGGGTAGATAATCACTC3′. NP mRNA (VN1203) - 5′CCAGATCGTTCGAGTCGTTTTTTTTTTTTTTTTTCTTTAATTGTC3′. NP mRNA (NL602 and CA04) - 5′CCAGATCGTTCGAGTCGTTTTTTTTTTTTTTTTCAACTGTCATACTC3′. RPL14 (human endogenous control) - 5′TTCAATCTTCTTGGCCCATC3′. RPL10 (mouse endogenous control) - 5′GAACGATTTGGTAGGGTATAGGAG3′.

The qPCR assay was performed using a SYBR green kit (Applied Biosystems, Carlsbad, CA, USA) with specific primers for the different RNA species, according to the manufacturer's standard protocol. Relative RNA quantities were determined using the comparative threshold cycle (*CT*) method, with human RPL14 or mouse RPL10 as endogenous. The qPCR primer sequences for endogenous controls used were:

RPL14 forward - 5′TTCATCCTCAAGTTTCCGC3′
RPL14 reverse - 5′TTCAATCTTCTTGGCCCATC3′
RPL10 forward - 5′TGAAGACATGGTTGCTGAGAAG3′
RPL10 reverse - 5′GAACGATTTGGTAGGGTATAGGAG3′.

The SARS-CoV qPCR primer sequences used were:
Forward primer for genomic RNA and mRNAs - 5′CCTACCCAGGAAAAGCCAAC3′.
Genomic RNA reverse primer - 5′GGACGAAACCGTAAGCAGTC3′
S mRNA reverse primer - 5′ACCGGTCAAGGTCACTAC3′
E mRNA reverse primer - 5′GCAAGAATACCACGAAAGCA3′.

The MERS-CoV qPCR primer sequences used were:
Forward primer for genomic RNA and mRNAs - 5′GAATAGCTTGGCTATCTCAC3′
Genomic RNA reverse primer - 5′CACAATCCCACCAGACAA3′
ORF2 reverse primer - 5′AGTGTATCATTGTCACGGATAAG3′
ORF5 reverse primer - 5′ACTAGCTGGACGGGTTTA3′

The Influenza qPCR primers sequences used were:
Genomic forward primer (VN1203, NL602 and CA04) - 5′GGCCGTCATGGTGGCGAAT3′
Genomic reverse primer (VN1203) - 5′GTTCCCCACCAGTTTCCATC3′
Genomic reverse primer (NL602 and CA04) - 5′GCTCCCCACCAGTCTCCATT3′
Forward mRNA primer (VN1203, NL602 and CA04) - 5′CCAGATCGTTCGAGTCGT3′
Reverse mRNA primer (VN1203, NL602 and CA04) - 5′CGAACCCGATCGTGCCTTCC3′

Isolation of total RNA and quality control
Isolation of RNA from Trizol lysates was performed using Qiagen RNeasy Mini kit (Qiagen Inc., Valencia, CA). The steps of the protocol were performed as follows: lysates were mixed by vortexing followed by a 5-minute incubation at RT. To each sample, 200 μL of chloroform per mL Trizol was added and the tubes shook vigorously for 15 seconds. In many cases only a portion of the 1 ml Trizol lysate was used and the volume of chloroform was adjusted accordingly. After incubation for 10 minutes at RT, samples were spun at 12,000 g for 15 minutes at 4° C. The aqueous phase was transferred to a clean tube and an equal volume of RLT +BME added. The volume was then doubled by adding 70% ethanol and mixed well by pipetting. This solution was applied to a RNeasy mini column and the RNA was isolated from the column per the manufacturer's instructions. Appropriate amounts of RNA for quality control were removed and the remaining RNA was frozen on dry ice and store at −80° C. Determination of RNA

quality and quantity was done using an Agilent Bioanalyzer 2100 and a Nanodrop spectrophotometer following the manufacturers' recommendations. (CA04M001, ICL004, ICL006, ICL010, ICL011, ICL012, IM001, IM002, IM004, IM005, IM006A, IM006B, IM007, IM009, IM010, ECL001, SCL005, SCL006, SHAE002, SHAE003, SHAE004, SM001, SM003, SM004, SM007, SM009, SM012, SM014, SM015, SM019, SM020).

Alternatively, chloroform (0.2 mL per 1 mL TRIzol) was added to samples, tubes were inverted 20 times, and incubated at room temperature for 3 minutes. Samples were centrifuged at 12,000 g for 15 minutes at 4° C. The aqueous phase (top phase) was transferred to a new tube and an equal volume of isopropyl alcohol was added and samples incubated for 10 minutes. Tubes were centrifuged at 12,000 g for 10 minutes at 4° C. The RNA pellet was washed once with 70% ethanol and re-suspended in 100 µL of H$_2$0, 1 µL of glycogen and 1/10 volume of 3 M NaAc. RNA was re-precipitated with an equal volume of 100% ethanol and washed 2 times with 70% ethanol. The dried pellet was resuspended in 30 µL water. The RNA concentration and purity was determined by absorbance at 260 nm, 270 nm, and 280 nm using a plate reader. RNA integrity was determined using a BioAnalyzer (Agilent, Santa Clara, CA). (SBRI_AA_E1).

Labeling and quantification of total RNA for agilent microarray

Fluorescently labeled probes were generated from each total RNA sample using Agilent one-color Low Input Quick Amp Labeling Kit (Agilent Technologies) following the manufacturer's instructions. The cRNA probe synthesis began with adding 100 ng total intact RNA to 3.25 µL water, combining RNA and Master Mix (consisting of 1.0 µL H2O, 0.25 µL Spike-in (dilution 2), 0.8 µL T7 Promoter Primer, 2.05 µL Total Volume per reaction), denaturing for 10 minutes at 65° C, and incubating at 4° C for 5 minutes. Next, double-stranded cDNA was synthesized by combining the above cRNA with 4.7 µL per reaction of First Strand mix (consisting of 2.0 µL 5X first stand Buffer1, 1.0 µL 0.1 M DTT, 0.5 µL 10 mM dNTP mix, and 1.2 µL AffinityScript RNase Block Mix) at room temperature followed by incubations at 40° C for 2 hours, 70° C for 15 minutes, and 4° C for 5 minutes.

Labeling of each sample was done by adding 6.0 µL of transcription master mix (consisting of 0.75 µL H$_2$O, 3.2 µL 5X Transcription Buffer, 0.6 µL 0.1 M DTT, 1.0 µL NTP Mix, 0.21 µL T7 RNA Polymerase Blend, and 0.24 µL of Cy3 per reaction), mixing by pipette, and incubating at 40° C for 2 hours in the dark. Next, probes were purified with the Qiagen Rneasy kit (RNA cleanup protocol). After 84 µL of nuclease free water was added (to 100 µL total volume), samples were transferred to 1.5 mL microfuge tubes. Next, 350 µL of RLT, and 250 µL of 100% ethanol were added and mixed thoroughly by pipetting. 700 µL of this mixture was then applied to a column, centrifuged 30 seconds at 13,000 rpm at 4° C (discard flow through and collection tube). The column was then transferred to a new collection tube, 500 µL of RPE washing buffer was added, and centrifuged for 30 seconds at 13,000 rpm at 4° C (discard flow through, reuse tube). Another 500 µL of RPE was added and centrifuged at 60 seconds at 13,000 rpm at 4° C, then transferred to a clean microfuge tube and spun for 30 seconds on max to dry column.

To elute labelled RNA, the column was transferred to a clean labeled collection microfuge tube and 30 µL of nuclease free H$_2$O was added to the center of the column for 60 seconds followed by a 60 second centrifugation at 13,000 rpm at 4° C. A small aliquot (~4 µL) was removed for quantification (see below). Samples were quick frozen on dry ice and stored at −80° C while protected from light. (CA04M001, ICL004, ICL006, ICL010, ICL011, ICL012, IM001, IM002, IM004, IM005, IM006A, IM006B, IM007, IM009, IM010, ECL001, SCL005, SCL006, SHAE002, SHAE003, SHAE004, SM001, SM003, SM004, SM007, SM009, SM012, SM014, SM015, SM019, SM020).

Quantification of cRNA probes was done by the Nanodrop spectrophotometry. For each sample 1.2 µL of purified probe was measured and cRNA concentration (ng/µL) and dye incorporation (pmol/µL) recorded. Specific activity was calculated by: (dye incorporation/ cRNA concentration)*1000. Results greater than 8.0 were used for hybridization.

Agilent microarray hybridization and scanning

In summary, 250 ng of each RNA sample was hybridized to either an Agilent 4x44K mouse whole-genome oligonucleotide microarray G4122F (design ID 014868) or 4X44K human HG (design ID 014850) array according to the manufacturer's instructions (details below).

Prior to hybridization, cRNA was first fragmented by mixing Cy3-cRNA per reaction volume of 55.0 µL (consisting of Cy3-cRNA 1650 ng, 11 µL of 10X Blocking Reagent, and Nuclease free H$_2$O; Bring volume to 52.8 µL), with fragmentation buffer (consisting of 2.2 µL 25X fragmentation buffer) and then incubating at 60° C for 30 minutes in the dark. 55 µL of 2X Gene Expression Hybridization buffer HI-RPM was then added (Total volume = 110 µL). Samples were mixed well by pipetting, and spun for 1 minute at room temperature at 13,000 rpm.

Subsequent to fragmentation, hybridization of 100 µL of the fragmented cRNA samples was transferred from above onto the center of each of the wells (4x44K format has 4 wells per slide with one sample per well). A slide was carefully lowered, bar code facing up, onto the bead of liquid, followed by the metal chamber cover and clamp. The chamber was rotated vertically to wet the gasket and make sure that all bubbles were mobile. Two chambers were load at a time into the hybridization oven, set at 65° C and level 10-rotation speed (10 rpm) for 17 hours.

Slides were then washed once with slide cover on in Gene Expression Wash Buffer 1, then cover was pried off and the slide washed in Gene Expression Wash Buffer 1 for a second time. Next, slides were washed in Gene Expression Wash Buffer 1 on a stir plate for ~1 minute and placed into pre-warmed Gene Expression Wash buffer 2 on a stir plate for ~1 minute. The excess liquid was blotted onto a Kimwipe and the dry slides placed into a light tight slide box to be scanned at once.

Dry slides were scanned on an Agilent DNA microarray scanner (model G2505B) using the XDR setting. Raw images were analyzed using the Agilent Feature Extraction software (version 9.5.3.1) and the GE1-v5_95_Feb07 extraction protocol. (CA04M001, ICL004, ICL006, ICL010, ICL011, ICL012, IM001, IM002, IM004, IM005, IM006A, IM006B, IM007, IM009, IM010, ECL001, SCL005, SCL006, SHAE002, SHAE003, SHAE004, SM001, SM003, SM004, SM007, SM009, SM012, SM014, SM015, SM019, SM020).

Agilent microarray processing and expression analysis

Normalization and QA/QC: Raw data consisted of the output from the Agilent Feature Extraction (AFE) software. Data was read into the R/Bioconductor Framework and visualization of raw intensity level was performed using boxplots, histograms, and correlation plots. All intensities were background corrected using the norm-exp method and quantile normalized using Agi4x44PreProcess and RMA Bioconductor packages. Distribution of control probes was examined and then those probes were removed. Descriptive statistics were reported for non-control replicate probes and the intensity distributions were examined via histograms. All non-control replicate probes were then summarized by taking the mean intensity measurement. To assess the quality of the reading for each probe, five Agilent QC flags were considered. The first is a test for minimal spot diameter and signal greater than 1.5 times background noise (gIsFound); the second requires positive intensity and background-subtracted intensity greater than 2.6 times background noise (gIsWellAboveBG); the third is true if more than 50% of spot pixels are greater than the saturation threshold (gIsSaturated); the fourth examines variance of spot pixels (gIsFeatNonUnifOL); the fifth checks if replicate probe is greater than 1.42 times the interquartile range (gIsFeatPopnOL).

A given probe must pass all 5 QC flags for all treatment (i.e., infected) replicates of at least one experimental time point. This allowed a conservative criteria regarding probe quality, but was moderate enough to detect transcripts that may be expressed only briefly (i.e., as little as one time point) during an infection time course. It was imperative that this filtering occurred before differential expression analysis so that low intensity (below background) and low quality probes were not considered in the analysis. Final QA/QC involved CV plots, heatmaps, cluster plots, scatter plots and PCA as needed in order to assess if there were sample outliers or additional batch effects.

Differential expression analysis was performed by comparing infected samples to time-matched mock-infected controls, based on a linear model fit for each transcript/probe using the R package Limma[35]. Fold changes were calculated by taking the average $\log_2$ probe intensity of infected replicates and subtracting the average $\log_2$ probe intensity of mock-infected replicates from the same time point. Significant differential expression of a probe was defined using the following criteria: $\log_2$ fold change $>|1.5|$ and an adjusted false discovery rate (FDR) P value of < 0.05 for a given time point.

Software Utilized: Custom R Bioconductor workflow that includes the Agi4x44 Pre-Process and RMA Bioconductor packages[36] plus in house scripts for QA/QC visualization and differential expression analysis. (CA04M001, ICL004, ICL006, ICL010, ICL011, ICL012, IM001, IM002, IM004, IM005, IM006A, IM006B, IM007, IM009, IM010, ECL001, SCL005, SCL006, SHAE002, SHAE003, SHAE004, SM001, SM003, SM004, SM007, SM009, SM012, SM014, SM015, SM019, SM020).

Positive-strand cDNA synthesis, fragmentation, and single-stranded cDNA for affymetrix microarray analysis

cDNA was synthesized from isolated RNA (250 ng from each sample) using the Ambion WT kit (Life Technologies, Grand Island, NY, USA - http://tools.lifetechnologies.com/content/sfs/manuals/cms_064619.pdf) following the manufacturers protocol. Briefly: First strand cDNA was prepared from total RNA using reverse transcriptase and specific primers (5 μL of First-Strand master mix) resulting in single stranded cDNA which contains a T7 promoter sequence. Second strand cDNA was synthesized using Second-Strand master mix and the newly created double stranded DNA was isolated. Antisense cRNA was synthesized from the second strand DNA using T7 RNA polymerase then stored at 4 °C overnight. In the next step, cRNA was purified using cRNA Binding Mix and the yield was determined by absorbance at 260 nm. Purified cRNA was used to synthesize positive sense strand cDNA using reverse transcriptase and random primers. In this reaction, dUTP was added at a fixed ratio relative to dTTP. cRNA was then hydrolyzed using RNase H and cDNA purified using cDNA Binding Mix and cDNA yield determined by absorbance at 260 nm.

Single stranded cDNA was prepared for chip hybridization using the Affymetrix GeneChip WT Terminal Labeling Kit (Affymetrix, Santa Clara, CA, http://www.umich.edu/~caparray/Files/protocols/affymetrix/Ambion%20WT%20Labelling%20and%20Hyb.pdf) following the manufacturer's protocol. Briefly: Single-stranded cDNA was fragmented using Fragmentation master mix from the kit, then labeled by terminal deoxynucleotidyl transferase (TdT) using the Affymetrix proprietary DNA Labeling Reagent. The labeled mix was then hybridized to GeneChip Mouse Exon 1.0 ST Arrays in a hybridization oven for 17 hours. (SBRI_AA_E1).

Affymetrix microarray processing and expression analysis

Probe intensities were measured using the Affymetrix GeneChip Scanner 3000 and processed into CEL files using the Affymetrix GeneChip Operating Software. Microarrays were normalized at the gene level using the BrainArray custom CDF (Entrez Gene, Version 14) for probeset definitions and RMA as implemented in the justRMA function of the Bioconductor package affy for background adjustment, quantile normalization, and summarization. Linearized data was imported into Genedata Analyst software (Genedata AG, Basel, Switzerland). ANOVA was performed on all samples (100 balanced permutations) and genes with a permutation Q-value of 0.01 or less and fold change of at least 2 (between any two groups) were exported. (SBRI_AA_E1).

Preparation of protein extracts for proteomic analysis

Protein concentrations of cell lysates or lung tissue homogenates were determined by BCA protein assay. The range of final protein amount was between 100-226 μg with an average of 150 μg per sample. Each sample was then diluted to uniform volume in 50 mM ammonium bicarbonate, pH 7.8. Proteins were reduced with 10 mM dithiothreitol, followed by alkylation of free sulfhydryl groups with 40 mM iodoacetamide at 37° C in the dark; each reaction was performed for 1 hour at 37° C with constant shaking at 550 rpm. Denatured and reduced samples were diluted 10-fold with 50 mM ammonium bicarbonate, pH 7.8, and $CaCl_2$ was added to a final concentration of 1 mM prior to enzymatic digestion. Sequencing-grade modified trypsin was activated by adding 20 μL of 50 mM ammonium bicarbonate, pH 7.8, to 20 μg lyophilized trypsin and incubating for 10 min at 37° C. Activated trypsin was added to the samples at 1:50 (w/w) trypsin-to-protein ratio, and samples were digested at 37° C for 3 hours with constant shaking at 800 rpm; reactions were quenched by rapid freezing in liquid nitrogen. Digested samples were desalted using solid phase extraction columns (Discovery C18, Supelco, Bellefonte, PA, USA), which were conditioned with 3 mL of methanol and rinsed with 2 mL of 0.1% trifluoroacetic acid (TFA) in water. Digest-loaded columns were washed with 4 mL of H_2O/acetonitrile (95:5, v/v) containing 0.1% TFA, and peptides were eluted with 1 mL of acetonitrile/H_2O (80:20, v/v) containing 0.1% TFA. Samples were concentrated to 100 μL *in vacuo* (Speed-Vac SC 250 Express, Thermo Savant, Holbrook, NY, USA), and a BCA protein assay was performed to verify final peptide concentrations. Samples were stored at −80° C until either strong cation exchange fractionation followed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) or quantitative LC-MS analyses. (CA04M001, ICL004, ICL006, ICL010, ICL011, ICL012, IM001, IM002, IM004, IM005, IM006A, IM006B, IM007, SCL005, SCL006, SM001).

Strong cation exchange fractionation

Strong cation exchange fractionation was performed on pooled cell or lung protein digests using an Agilent 1100 HPLC System (Agilent, Palo Alto, CA, USA) equipped with a quaternary pump, degasser, diode array detector, peltier-cooled autosampler, and fraction collector (set at 4° C). Peptides (350 μg per injection) were separated with a PolySulfoethyl A (PolyLC Inc., Columbia, MD, USA) column (200 mm × 2.1 mm; 5 μm particles with 300-Å pores) with a 10 mm × 2.1 mm guard column packed with the same material at a flow rate of 0.2 mL/min. The solvents consisted of (a) 10 mM ammonium formate, pH 3.0, and 25% acetonitrile, and (b) 500 mM ammonium formate [pH 6.8] and 25% acetonitrile. The following linear gradient was used: 100% solvent (a) for 10 min; ramp to 50% solvent (b) in 40 min; ramp to 100% solvent (b) in the next 10 min; followed by a 10 min hold in 100% solvent (b). Routinely, 24 fractions were collected from minute 30 to minute 65 of the gradient, and they were subsequently dried *in vacuo* and stored at −80° C until LC-MS/MS analysis. (CA04M001, ICL004, ICL006, ICL010, ICL011, ICL012, IM001, IM002, IM004, IM005, IM006A, IM006B, IM007, SCL005, SCL006, SM001).

Reversed-phase capillary LC-MS-MS and LC-MS analyses

Capillary LC-MS/MS analysis was used to generate separate accurate mass and time (AMT) tag databases for virus-infected cells and lung homogenates (see below). For this, dried peptide fractions from pooled samples were reconstituted in 30 μL of 25 mM ammonium bicarbonate, pH 7.8, and analyzed using a 4-column custom-built capillary LC system coupled online to a linear ion trap mass spectrometer (LTQ; Thermo Scientific, San Jose, CA, USA) by way of an in-house manufactured electrospray ionization interface. Electrospray emitters were custom made using 150 μm outer diameter (o.d.) x 20 μm inner diameter (i.d.) chemically etched fused silica. Reversed-phase capillary columns were prepared by slurry packing 3-μm Jupiter C18 bonded particles (Phenomenex, Torrence, CA, USA) into a 75 μm x 65 cm fused silica capillary (Polymicro Technologies, Phoenix, AZ, USA) using 0.5 cm sol-gel plugs for particle retention. Mobile phases consisted of (a) 0.1% formic acid in water and (b) 0.1% formic acid in acetonitrile, and they were degassed on-line using a Degasys Model DG-2410 vacuum degasser (Dionex, Germany); the HPLC system was equilibrated at 10,000 psi with 100% mobile phase (a) for initial starting conditions. After loading 2.5 μg of peptides onto the column, the mobile phase was held at 100% mobile phase (a) for 50 min. Exponential gradient elution was initiated 50 min after sample loading with a column flow rate of 300 nL/min, and the mobile phase was ramped from 0% to 55% mobile phase (b) over 100 min using a 2.5 mL stainless steel mixing chamber, followed by a rapid increase to ~100% (b) for 10 min to wash the column. To identify the eluting peptides, the LTQ was operated in a data-dependent MS/MS mode (400-2,000 *m/z*), in which a full MS scan was followed by ten MS/MS scans

using a normalized collision energy of 35%. A dynamic exclusion window of 1 min was used to discriminate against previously analyzed ions. The temperature of the heated capillary and the electrospray ionization (ESI) voltage were 200° C and 2.2 kV, respectively.

Following AMT tag database generation, capillary LC-MS analyses were performed on all virus-infected and mock-infected samples to generate quantitative data using the AMT tag approach. For this, individual dried peptide samples (see above) were reconstituted in 30 μL of 25 mM ammonium bicarbonate, pH 7.8, and analyzed in either duplicate or triplicate and in random order using identical chromatographic and electrospray conditions as for capillary LC-MS/MS analyses. The LC system was interfaced to an Exactive Orbitrap mass spectrometer (Thermo Scientific), and the temperature of the heated capillary and the ESI voltage were 250° C and 2.2 kV, respectively. Data were collected over the mass range 400-2,000 m/z. (CA04M001, ICL004, ICL006, ICL010, ICL011, ICL012, IM001, IM002, IM004, IM005, IM006A, IM006B, IM007, SCL005, SCL006, SM001).

Development of the AMT tag database for virus-infected model systems

A novel AMT database was generated for peptides within the specific model system, using mock-infected and virus-infected samples. To generate the AMT tag database, aliquots of the virus infected or mock-infected samples were combined to make the appropriate sample pools. Each pool was subjected to strong cation exchange fractionation as described above, and each fraction was analyzed by capillary LC- MS/MS. The SEQUEST analysis software was used to match the MS/MS fragmentation spectra with sequences from the appropriate UniProt/Swiss-Prot protein database. When searching, SEQUEST used a dynamic mass modification on methionine residues corresponding to oxidation (15.9949 Da) and a static mass modification on cysteinyl residues to account for alkylation by iodoacetamide (57.0215 Da). Peptides passing the following filter criteria were stored as AMT tags in a Microsoft SQL Server database: 1) SEQUEST DelCn2 value ≥ 0.10 (normalized Xcorr difference between the top scoring peptide and the second highest scoring peptide in each MS/MS spectrum) and 2) for the mouse lung tissue database, a SEQUEST correlation score (Xcorr) ≥ 1.6, 2.4, and 3.2 for fully tryptic peptides with 1+, 2+, and 3+ charge states, respectively, and Xcorr ≥ 4.3, and 4.7 for partially tryptic or non-tryptic protein terminal peptides with 2+, and 3+ charge states, respectively. While for the human cell line database, a SEQUEST correlation score (Xcorr) ≥ 2, 2.6, and 3.5 for fully tryptic peptides with 1+, 2+, and 3+ charge states, respectively, and Xcorr ≥ 2.5, 3.6, and 4.1 for partially tryptic or non-tryptic protein terminal peptides with 1+, 2+, and 3+ charge states, respectively. These filter criteria resulted in estimated peptide false discovery rates of < 2% based on target-decoy database searches for both lung and cell samples. Non-tryptic peptides were excluded, and a minimum peptide length of 6 amino acid residues was required. The elution times for these peptides were normalized to a range of 0 to 1 using a predictive peptide LC normalized elution time (NET) model and linear regression. A NET average and standard deviation are assigned to each identified peptide if the same peptide is observed in multiple analyses. Both calculated monoisotopic masses and observed NETs of identified peptides were included in the AMT tag database. From the primary data records corresponding to the data used to generate the AMT tag database (PASS00416 and PASS00417), the user can download the raw data, the assembled database, and parameter files necessary to recreate the AMT-tag approach. In addition to the data files, a detailed tutorial has been made available for download. (CA04M001, ICL004, ICL006, ICL010, ICL011, ICL012, IM001, IM002, IM004, IM005, IM006A, IM006B, IM007, SCL005, SCL006, SM001).

Processing of quantitative LC-MS datasets

Quantitative LC-MS datasets were processed using the PRISM Data Analysis system[37], which is a series of software tools developed in-house (e.g. Decon2LS[38] and VIPER[39]; freely available at http://ncrr.pnl.gov/software/). Individual steps in this data processing approach are reviewed here[40]. The peptide identities for detected features in each dataset (i.e. a single LC-MS analysis) were determined for features matched to AMT tags with high confidence based upon the accurate measured monoisotopic masses and NETs for each of the peptides in the filtered AMT tag databases within initial search tolerances of ± 6 ppm and ± 0.025 NET for monoisotopic mass and elution time, respectively. The peptides identified from this matching process were retained as matrices for subsequent data analysis. (CA04M001, ICL004, ICL006, ICL010, ICL011, ICL012, IM001, IM002, IM004, IM005, IM006A, IM006B, IM007, SCL005, SCL006, SM001).

Proteomic data statistical processing

The integrated LC-MS peak intensity values (i.e. abundances) for the final peptide identifications were processed in a series of steps using MatLab R2010b that included quality control, normalization, protein quantification, and comparative statistical analyses. Quality control processing was performed to identify and remove peptides with an insufficient amount of data across the set of samples[41], and LC-MS runs that showed significant deviation from the standard behavior of all LC-MS analyses[42], using a significance level of 0.0001 (ref. 43). The peptide abundance values were normalized across the technical replicates using a rank invariant subset of peptides (p-value threshold of 0.1) followed by median absolute deviation centering of the data. Normalized $\log_{10}$ abundance values were averaged across the technical replicates within each biological sample. Proteins were quantified using a standard R-Rollup method[44] using the most abundant reference peptide--after filtering the peptides that were redundant, had low data content,

or were outside the dominant significance pattern. Comparative statistical analyses of time-matched mock with viral infected samples were performed using a Dunnett adjusted t-test to assess differences in protein average abundance. (CA04M001, ICL004, ICL006, ICL010, ICL011, ICL012, IM001, IM002, IM004, IM005, IM006A, IM006B, IM007, SCL005, SCL006, SM001).

Data Records

Study designs for all experiments, including relevant repository identifiers, are summarized in Table 1 (Available online only), with individual RNA experiment samples detailed in Table 2 (Available online only) and individual protein experiment samples detailed in Table 3 (Available online only). Sample tracking from animal subjects to experiment samples is given in Table 4 (Available online only). Primary assay results (txt and mzXML files) are archived in public repositories as listed below (Data Record 1 and 2). Derived data resulting from processing and hypothesis testing accompanied by structured metadata describing both the experimental and data analysis methods are available at the Influenza Research Database and the Viral Pathogen Resource (Data Record 3).

Data Record 1

Primary data for Agilent and Affymetrix microarray experiments are available at the NCBI Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) under the accession numbers GSE37569 (Data Citation 1)[8], GSE45042 (Data Citation 2)[5], GSE28166 (Data Citation 3)[4,45], GSE37571 (Data Citation 4), GSE43203 (Data Citation 5), GSE43204 (Data Citation 6), GSE33263 (Data Citation 7)[8,45], GSE36328 (Data Citation 8)[6], GSE37572 (Data Citation 9)[8], GSE43301 (Data Citation 10)[8], GSE43302 (Data Citation 11)[8], GSE44441 (Data Citation 12), GSE44445 (Data Citation 13)[8], GSE40792 (Data Citation 14), GSE37245 (Data Citation 15)[9], GSE33267 (Data Citation 16)[5,12], GSE37827 (Data Citation 17), GSE47960 (Data Citation 18)[11], GSE47961 (Data Citation 19)[11], GSE47962 (Data Citation 20)[11], GSE33266 (Data Citation 21), GSE50000 (Data Citation 22), GSE51386 (Data Citation 23), GSE50878 (Data Citation 24), GSE51387 (Data Citation 25), GSE49262 (Data Citation 26), GSE49263 (Data Citation 27), GSE40824 (Data Citation 28), GSE40827 (Data Citation 29).

Data Record 2

Raw data files and associated parameter files (.raw,.mzXML) for all proteomics experiments including those used for AMT-TAG database generation have been uploaded to PeptideAtlas (http://www.peptideatlas.org). Experiment accession numbers are as follows: PASS00416 (Data Citation 30), PASS00417 (Data Citation 31), PASS00418 (Data Citation 32), PASS00419 (Data Citation 33), PASS00420 (Data Citation 34), PASS00421 (Data Citation 35), PASS00422 (Data Citation 36), PASS00423 (Data Citation 37), PASS00424 (Data Citation 38), PASS00425 (Data Citation 39), PASS00426 (Data Citation 40), PASS00427 (Data Citation 41), PASS00428 (Data Citation 42), PASS00429 (Data Citation 43), PASS00430 (Data Citation 44), PASS00431 (Data Citation 45), PASS00432 (Data Citation 46), PASS00433 (Data Citation 47).

Data Record 3

Structured experiment metadata and derived data (.xls,.txt) including fold change values and p-values are available in the Host Factor Component of the Influenza Research Database (IRD, www.fludb.org) and the Virus Pathogen Resource (ViPR, www.viprbrc.org) under accession numbers: IRD_SV_CA04M001-P (Data Citation 48), IRD_SV_ICL004-P (Data Citation 49), IRD_SV_ICL006-P (Data Citation 50), IRD_SV_ICL010-P (Data Citation 51), IRD_SV_ICL011-P (Data Citation 52), IRD_SV_ICL012-P (Data Citation 53), IRD_SV_IM001-P (Data Citation 54), IRD_SV_IM004-P (Data Citation 55), IRD_SV_IM005-P (Data Citation 56), IRD_SV_IM006A-P (Data Citation 57), IRD_SV_IM006B-P (Data Citation 58), IRD_SV_IM007-P (Data Citation 59), IRD_SV_SCL005-P (Data Citation 60), IRD_SV_SCL006-P (Data Citation 61), IRD_SV_SM001-P (Data Citation 62), IRD_SV_CA04M001-R (Data Citation 63), IRD_SV_ECL001-R (Data Citation 64), IRD_SV_ICL004-R (Data Citation 65), IRD_SV_ICL006-R (Data Citation 66), IRD_SV_ICL010-R (Data Citation 67), IRD_SV_ICL011-R (Data Citation 68), IRD_SV_ICL012-R (Data Citation 69), IRD_SV_IM001-R (Data Citation 70), IRD_SV_IM002-R (Data Citation 71), IRD_SV_IM004-R (Data Citation 72), IRD_SV_IM005-R (Data Citation 73), IRD_SV_IM006A-R (Data Citation 74), IRD_SV_IM006B-R (Data Citation 75), IRD_SV_IM007-R (Data Citation 76), IRD_SV_IM009-R (Data Citation 77), IRD_SV_IM010-R (Data Citation 78), IRD_SI_SBRI_AA_E1 (Data Citation 79), IRD_SV_SCL005-R (Data Citation 80), IRD_SV_SCL006-R (Data Citation 81), IRD_SV_SHAE002-R (Data Citation 82), IRD_SV_SHAE003-R (Data Citation 83), IRD_SV_SHAE004-R (Data Citation 84), IRD_SV_SM001-R (Data Citation 85), IRD_SV_SM003-R (Data Citation 86), IRD_SV_SM004-R (Data Citation 87), IRD_SV_SM007-R (Data Citation 88), IRD_SV_SM009-R (Data Citation 89), IRD_SV_SM012-R (Data Citation 90), IRD_SV_SM014-R (Data Citation 91), IRD_SV_SM015-R (Data Citation 92), IRD_SV_SM019-R (Data Citation 93), IRD_SV_SM020-R (Data Citation 94).

Technical Validation

Two approaches were used to establish the technical validity of the datasets described. First, the biological samples used to generate the transcriptomic and proteomic assay results reported here have also been

assessed using various criteria to ensure technical validity. For example, productive viral infections following inoculation of animals used for *in vivo* studies was confirmed by examining viral titers, viral mRNA levels, and body weight loss to quantify virus replication and host response to infection as detailed below. For *in vitro* studies, viral genomic RNA, viral mRNA, and viral titers were quantified to ensure that the virus was actively replicating and transcribing its genome in cell culture. Second, the processed data produced from *in vivo* and *in vitro* studies were compared, using Boolean logic, to quantify the number of overlapping significant genes identified in multiple studies performed under similar experimental conditions as an assessment of general reproducibility of the derived results between experiments.

In vivo validation

To validate the *in vivo* studies, mean viral mRNA levels, mean % body weight, and mean viral titers were examined as surrogates for viral infection in the animal subjects. The CA04M001 and IM001 experiments, in which mice were infected with influenza A virus strains A/California/04/2009 (pandemic H1N1) and A/Viet Nam/1203/2004 (H5N1) respectively, were used as representative datasets for this validation process. For both experiments, the viral mRNA levels indicate that the virus was actively transcribing mRNA within the host organism with mean viral mRNA levels peaking at either 2 or 4 days post infection (dpi), when either a high or low initial inoculum was used, respectively[8]. Similarly, the mice that were exposed to virus in either experiment showed a dose-dependent decrease in mean % body weight. The viral growth kinetics for A/California/04/2009 in the CA04M001 study showed mean viral titers peaking at 2 dpi regardless of the viral concentration present in the initial inoculation. In contrast, the growth kinetics for A/Viet Nam/1203/2004 used in the IM001 study showed mean viral titers peaking at 2 days post infection when the initial inoculum contained either 10^3 or 10^4 plaque forming units (PFU) of virus; whereas mean viral titers were highest at 4 dpi when the initial inoculum contained 10^2 PFU of virus.

In vitro validation

For *in vitro* studies, measurements from study ICL004, performed in Calu-3 cells infected with the A/Viet Nam/1203/2004 influenza virus strain, were used as representative of all studies. Mean viral genomic RNA levels increased logarithmically between 0 and 12 hours post infection (hpi), eventually peaking at 18 hpi before decreasing at 24 hpi. Similarly, mean viral mRNA levels increased exponentially between 0 and 7 hpi, where they peaked before decreasing through the 24-hour time point. Mean viral titers were observed to increase logarithmically between the 7 and 12-hour time points, and then continued to rise through 24 hpi[8].

Overall, the metrics used to validate the *in vivo* and *in vitro* studies indicate that the biological samples that were taken from these sample sources were productively infected with the intended virus. Consequently, any downstream assays that were performed on these biological samples should give rise to biologically valid results.

Reproducibility of derived results

The primary assay results obtained from the transcriptomic and proteomic experiments were processed to identify specific RNAs or peptides that were differentially expressed under a particular set of conditions (e.g. virus amount, time post infection, virus strain, etc.) in comparison with relevant control samples to derive so called 'Host Factor Biosets'. Another measure of technical validity of the combined laboratory and analytical workflow is that the Host Factor Biosets derived from similar experiment samples treated in a similar way should have similar membership, thereby indicating experimental reproducibility. To examine experimental reproducibility, we determined the overlap that was observed between all experiments that used the same virus strains in the same infection system. In the evaluation reported here, we used datasets reporting results from wild-type A/California/04/2009 (H1N1), A/Viet Nam/1203/04 (H5N1), and SARS-CoV as being representatives of all experiments. In order to avoid complications related to slight differences in infection kinetics, we took the union of all biosets across the entire time course from each experiment (i.e. the non-redundant set of differentially expressed genes) for comparison. This approach can lead to a discrepancy between the total number of genes in each experiment due to differences in number of time points sampled. As an alternative strategy, the data could also be more strictly paired using specific time points and dosages.

Seven situations in which the same treatment conditions (viral strain and viral amount) were reproduced in separate experiments were identified. Since the magnitude of the responses and the number of time points differ between experiments, we calculated the percentage overlap of the number of host factors using the smallest set as the denominator for the sake of comparison. The smallest set was chosen as the denominator to calculate percentage overlap, which may not be an ideal comparison since this would tend to result in higher reproducibility values than if the largest set was used as the denominator. However, since we are primarily interested in measuring the commonality between similar experiments that have slight variations in study designs and are inherently noisy, we chose to focus on determining if the smaller set is largely a subset of the larger sets, which would demonstrate that the these significant hits are reproducible even in the presence of confounding variables such as biological and technical variation.

For transcriptomic experiments the following overlap was observed: 83% for A/California/04/2009 in Calu-3 (Fig. 2a), 91% for A/Viet Nam/1203/2004 in Calu-3 (Fig. 2b, 3-way comparison), 62% for A/California/04/2009 in C57BL/6 (Fig. 2c), 77% for A/Viet Nam/1203/2004 in C57BL/6 (Fig. 2d), 72% for A/California/04/2009 in HAE cells (Fig. 2e), and 40% for SARS-CoV in HAE cells (Fig. 2f). For the one replicated proteomic experiment, in which A/Viet Nam/1203/2004 was used to infect Calu-3, 43% overlap was observed (Fig. 2g). The reproducibility of host responses to both pandemic and highly pathogenic avian influenza viruses was high in both the Calu-3 cell line and mouse infection systems. Reproducibility using HAE was somewhat lower both in terms of the absolute number of host factor responding (Fig. 2e) and the percentage overlap (only 39% in Fig. 2f), perhaps due to the variability in establishing similar cell culture conditions using primary epithelial cells. Reproducibility of proteomic results was also somewhat lower with an overlap of 43% following infection of Calu-3 with A/Viet Nam/1203/2004 (Fig. 2g).

In general, the reproducibility of this collection of experiments using well-controlled standard operating procedures was quite high. In other microarray studies it has been shown that technical variability can result in 10% difference between replicate RNA samples[46,47], while biological variability can cause anywhere from 10% to 30% differences depending on the biological system sampled[48]. In addition, variability in stimulation conditions has been found to have an even more profound impact than either technical or biological variability[47]. Thus, the studies compared here demonstrated considerable reproducibility in results in spite of the number of sources of variability inherent in their designs. Furthermore, reduction in the number of false positive results in any individual experiment (high specificity) could be achieved by focusing on the host factors that were reproducibly identified in similar experiments. Reduction in the number of false negative results (high sensitivity) could be achieved by using the union of host factors identified among similar experiments.

Overall, the results from the independent *in vitro* and *in vivo* validation experiments and the high degree of reproducibility of the derived host factor lists show that the datasets described here are of high quality and can be useful for in-depth investigation and hypothesis generation in follow-up analyses.

Usage Notes

The systems biology studies described herein were designed to characterize how the host responds to viral infection. Separate experiments were performed to measure the response to multiple independent variables that include sampling at multiple time points after infection, at different viral dosages, in various model systems, and in systems with various genetic backgrounds (both viral and host). A number of publications have reported the use of these datasets to address questions concerning pathogenesis. For instance, comparisons between the similarities and dissimilarities of the host response at early and late stages of infection with genetically modified viral strains have been performed to identify possible correlates of disease severity[9]. Likewise, the host response was measured in different model systems to identify host factors that have a role in disease progression[45]. Lastly, all of these variables were combined to conduct a detailed evaluation of HPAI H5N1 infection based on time, dosage, and genetically-modified viruses[8].

Various analytical and statistical tools can be used to interpret the results from systems biology experiments. The commonly-used methods fall into three categories: enrichment analysis, co-expression network analysis, and transcription factor predictions. Enrichment analysis tools, such as DAVID or GSEA[49–51], use diverse statistical methods to identify annotations that are over-represented in a list of significant genes or proteins. Co-expression network analysis methods, such as Weighted Correlation Network Analysis (WGCNA)[52], find correlated patterns across multiple microarray samples. Lastly, transcription factor prediction tools, such as PSCAN and PRISM[53,54], examine the promoter regions of co-expressed genes to identify over-represented transcription factor binding sites. All of these tools allow researchers to generate hypotheses concerning how different components of the biological system being studied are modulated.

The host factor component within IRD and ViPR contains the processed data derived from these studies focusing primarily on the host response to virus infection. Here we present several scientific use cases that highlight some of the exploration and data mining uses that are currently possible using these data in the IRD/ViPR systems.

Host responses to related viruses

Comparisons of host responses to infection by viruses that are phylogenetically related but associated with subtle or dramatic differences in disease symptoms or severity can be used to help identify the genetic determinants of virulence.

The IM002 experiment was designed to interrogate the different host responses that occurred when mice were infected with more distantly related H1N1 influenza A viruses: A/Brisbane/59/2007, A/New Jersey/8/1976, A/Mexico/4482/2009, and a reconstructed 'Spanish flu' strain r1918 at 1, 3, and 5 days post infection. Although all four viruses are of the H1N1 subtype, they differ in their apparent virulence characteristics in humans, with the original isolates of the A/New Jersey/8/1976, A/Mexico/4482/2009, and the r1918 strain all described as causing severe respiratory illness, whereas the A/Brisbane/59/2007 strain was isolated from a relatively mild case, and was included in the trivalent vaccine from 2008-2010. Previous analyses of the pathways and functions that were over-represented in the list of differentially expressed genes revealed that inflammatory response, cell-to-cell signaling, immune cell trafficking, lipid
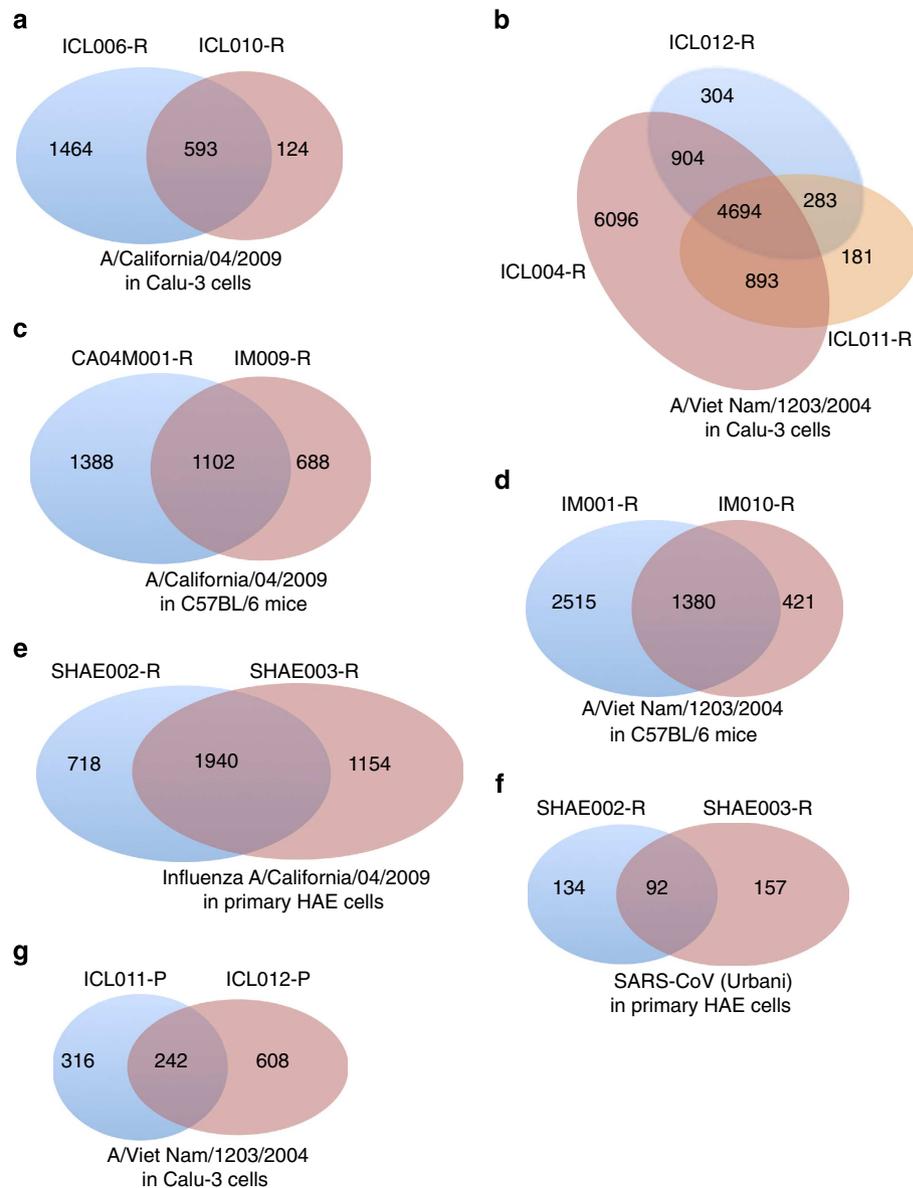
Figure 2. Computational validation of results. (**a**) Shows the comparison performed between all significant host factors identified by transcriptomics in the ICL006-R and ICL010-R experiments, both of which involved infection of Calu-3 cells with A/California/04/2009 at 3 multiplicity of infection (MOI). (**b**) Compares all significant host factors identified by transcriptomics in the ICL004-R, ICL011-R and ICL012-R experiments, which involved infection of Calu-3 cells with the A/Viet Nam/1203/2004 strain at 1 MOI. (**c**) A comparison between all significant host factors identified by transcriptomics in the CA04M001-R and IM009-R experiments, both of which involved infection of C57BL/6 mice with 10^6 plaque forming units (PFU) of A/California/04/2009. (**d**) Compares all significant host factors identified by transcriptomics in the IM001-R and IM010-R experiments, both of which involved infection of C57BL/6 mice with 10^3 PFU of A/Viet Nam/1203/2004. (**e**) Compares all significant host factors identified by transcriptomics in the SHAE002-R and SHAE003-R experiments, which involved infection of HAE cells with A/California/04/2009 at 2 MOI. (**f**) Compares all significant host factors identified by transcriptomics in the SHAE002-R and SHAE003-R experiments, which involved infection of HAE cells with the SARS-CoV Urbani strain at 2 MOI. (**g**) Compares all significant host factors identified by proteomics in the ICL011-P and ICL012-P experiments, which involved infection of Calu-3 cells with A/Viet Nam/1203/2004 at 1 MOI.

metabolism, amino acid metabolism, cardiovascular disease, and cancer gene responses were associated with virulence differences[6].

To illustrate one potential way to compare datasets in ViPR/IRD, we quantified the overlap between differentially expressed genes identified from samples infected with this collection of 'pre-pandemic' H1N1 influenza A virus. As before, we compared the non-redundant union of significant genes derived from the transcriptomic results reported in the IM002 experiment to determine differences detected in mice infected with these viruses. Although the pairwise comparisons of the host response to virus infection were somewhat variable, the differences appeared to correlate with the virulence of the infecting virus. No overlap was observed between all three virulent viruses and the A/Brisbane/59/2007 strain. In contrast, 262 host factors (66%) were shared in mice infected with the three virulent viruses (Fig. 3a). Performing a functional annotation analysis for this list of 262 genes identified the same list of GO terms
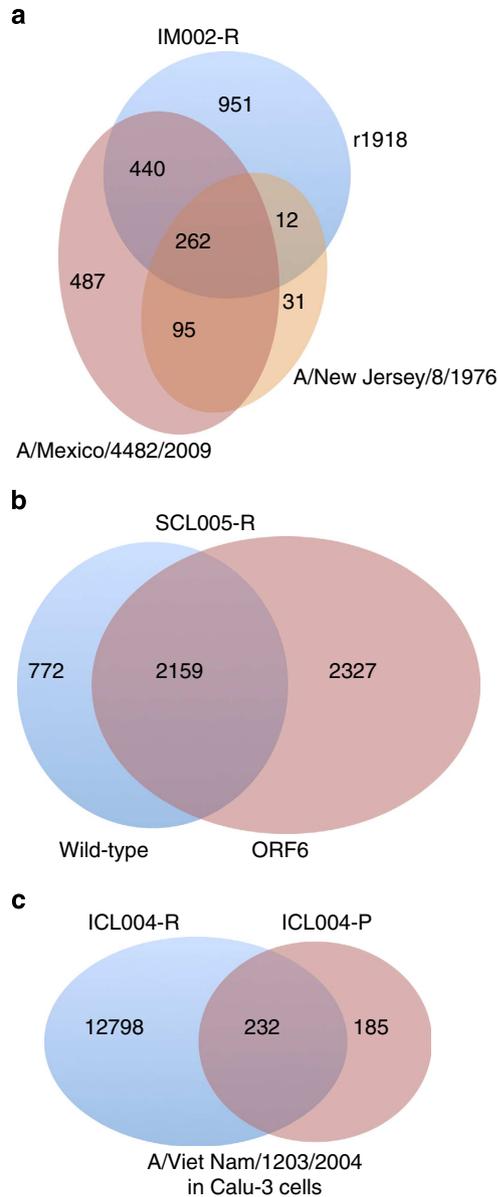
Figure 3. Comparative analyses usage examples. (**a**) Compares the significant host factors identified in the IM002-R experiment by transcriptomics at all time points by infecting C57BL/6 mice with 10^6 PFU of H1N1 influenza strains A/Mexico/4482/2009, A/New Jersey/8/1976, or r1918. (**b**) Shows the comparison of significant factors identified by transcriptomics in the SCL005-R experiment that infected Calu-3 cells at 5 MOI with either wild-type SARS-CoV or a mutant containing a functional ORF6 knockout. (**c**) Results from the ICL004-R and ICL004-P experiments show the overlapping and unique host factors identified by transcriptomics and proteomics respectively for time points after (and including) 12 hpi in Calu-3 cells infected at 1 MOI with influenza strain A/Viet Nam/1203/2004.

as reported in the original study[6] except for the exclusion of cardiovascular disease and cancer.

Similar analyses designed to compare the host response to other related viruses could be performed using the datasets that are currently available in IRD/ViPR. For example, the ICL010-R experiment examines the host transcriptional response to 2 different 2009 influenza A (H1N1) pandemic strains in Calu-3 cells at 12-, 24, and 48 hpi. The ECL001-R and SCL005-R experiments could be analysed to better understand the similarities and differences in the host response during infection with either MERS or SARS-CoV in Calu-3 cells across multiple time points.

Host responses to mutant viruses

In order to investigate how specific viral proteins might function to regulate host intracellular processes, experiments were performed to identify shared and unique host genes that are differentially expressed in hosts infected with either wild-type or mutant viruses[5,11,13]. For example, analysis of experimental results involving variants of SARS-CoV with a deletion in the ORF6 coding region performed in human cell lines (SCL005) or mice (SM012) determined that antagonism of karyopherin-based transport was an important function of the ORF6 gene product in SARS-CoV. As nuclear signaling disruption is a strategy used by many viral pathogens, identifying trends in these data presents a possible area for general therapeutic intervention.

Using Host Factor Bioset data in ViPR/IRD comparing human cell lines infected with either wild-type or ORF6 mutant viruses, 74% of significant host factors were shared between the two infection conditions (Fig. 3b). When an enrichment analysis was performed on the non-overlapping host factors (i.e. unique to infection with the ORF6 mutant strain), we found enrichment for terms relating to cell proliferation and immune response ($P < 0.05$), which matches what had been reported previously[12].

Similar analyses can be performed with data from other studies in IRD/ViPR to compare infection with wild-type versus genetically-engineered mutant strains. For instance, there are multiple studies that characterize the host responses to SARS-CoV variant strains in a human cell line (SCL006) and a mouse model (SM003, and SM012), and host responses to four influenza mutant variants of VN1203 in mice (IM004, IM005, IM006A, IM006B, and IM007) and a human cell line (ICL011 and ICL012).

Differences between RNA and protein responses

While proteins carry the functional workload in cells, it is still challenging to achieve the levels of dynamic range necessary to measure their abundances and activities at a comprehensive genome-wide level to support systems-level analysis, given practical constraints upon measurement throughput and sensitivity. In contrast, while gene expression arrays allow for the comprehensive measurement of gene expression levels across the entire genome, these levels do not directly reflect functional physiology mediated by the encoded proteins. Therefore, it is interesting to compare quantitative proteomic and transcriptomic experiments using similar samples to estimate how often transcriptional profiles are reflected in protein profiles.

Prior reports have compared transcriptomic and proteomic data as a way of validating experimental results[12]. In these cases, proteomic data was integrated with the corresponding transcriptomic data to improve the biological networks produced and to support the associated statistical analysis of SARS-CoV host responses in general, and the karyopherin transport hubs in particular.

We performed a similar comparison of RNA and protein Host Factor Biosets in ViPR/IRD from the same tissue/cell samples using either microarray or mass spectrometry technologies (ICL004). In this case, transcriptomic data from the 12 h time point of human Calu-3 cells infected with A/Viet Nam/ 1203/2004 influenza virus was chosen because this time point contained the highest number of significant probes. These transcriptomic data were compared with a pooled list from the time-matched and all subsequent time points from the related proteomics data in order to control for the 'lag' that occurs between when the changes in mRNA expression are translated into corresponding changes in the translated protein product. We found approximately 56% overlap between the statistically significant RNA at 12 hpi and the statistically significant protein at the combined 12 and 18/24 hpi time points (Fig. 3c), similar to what has been reported previously[55,56].

This type of analysis can be conducted on a number of other studies hosted by ViPR/IRD that evaluated both the transcript and protein host responses. These experiments included influenza infection in human cell lines (ICL006, ICL010, ICL011, ICL012), mice (CA04M001, IM001, IM004, IM005, IM006A, IM006B, IM007), and SARS-CoV infections in human cell lines (SCL005, SCL006) and mice (SM001).

Meta-analysis

While many of the usage examples provided involve comparative analysis of results obtained in the same experiment, in some cases similar comparisons can be performed between experiments. Indeed, the Validation section focused on the reproducibility of derived results and involved the comparison of similar experiment conditions performed in separate experiments. Combining and contrasting results from different studies is often referred to as meta-analysis. One of the goals behind making these datasets freely available through public database resources like GEO, ViPR and IRD is to support their incorporation into meta-analysis approaches that include new experiments being performed by the research community. To facilitate this kind of meta-analysis, considerable effort has been dedicated to providing complete descriptions of all experimental and data analysis methodologies, combined with structured, standardized metadata describing selected key aspects of the experiment design, especially regarding the experiment

factors that serve as the independent (manipulated) variables in the study in ViPR and IRD. Clear representations of the experiment metadata will make it relatively simple to appropriately combine these primary and derived results with data from future related experiments for meta-analysis.

Re-analysis

Although we have emphasized the use of the derived Host Factor Biosets, in every case a close linkage is maintained between the processed data and the primary assay results from which they were derived-even though they may be stored in distinct resources. Thus, the final usage example is related to the re-processing and re-analysis of primary assay results.

Different processing algorithm(s) chosen to perform a specific processing or analysis step in a multistep workflow can yield different results. Many bioinformatics data processing workflows currently implement a 'one size fits all' approach with a single algorithm being applied to all datasets without regard for the specific nuances of each individual dataset. The problem with this approach is that the processing algorithms and parameters that are ideal for one dataset may not be optimal for other datasets—especially when dealing with different experimental systems. Hence, processing of the data using a single universally-applied algorithm may not sufficiently address the individual requirements of each dataset. To truly solve this underlying problem, an analytical workflow that explores the relevant parameter space for a diverse set of algorithms would need to be implemented. This approach would provide an objective and automated process to intelligently 'choose' the optimal combination of parameters and algorithms that produce the best set of results for each individual dataset. Such an approach could be implemented in analytical workflow suites like Galaxy or GenePattern[57,58] to objectively and confidently reprocess the raw data from each '-omics' dataset in a data-driven fashion to further improve the quality of the derived data. Until such a system is developed, we advocate for using the processed data that was provided by the original submitters and subjected to the peer-review process when performing meta-analyses given its overall high quality.

References

1. Peng, X. *et al.* Virus-host interactions: from systems biology to translational research. *Curr. Opin. Microbiol.* **12,** 432–438 (2009).
2. Zak, D. E. & Aderem, A. Systems biology of innate immunity. *Immunol. Rev.* **227,** 264–282 (2009).
3. Aderem, A. *et al.* A systems biology approach to infectious disease research: innovating the pathogen-host research paradigm. *MBio* **2,** e00325–10 (2011).
4. Li, C. *et al.* Host regulatory network response to infection with highly pathogenic H5N1 avian influenza virus. *J. Virol.* **85,** 10955–10967 (2011).
5. Josset, L. *et al.* Cell host response to infection with novel human coronavirus EMC predicts potential antivirals and important differences with SARS coronavirus. *MBio.* **4,** e00165–13 (2013).
6. Josset, L. *et al.* Implication of inflammatory macrophages, nuclear receptors, and interferon regulatory factors in increased virulence of pandemic 2009 H1N1 influenza A virus after host adaptation. *J. Virol.* **86,** 7192–7206 (2012).
7. Gralinski, L. E. *et al.* Mechanisms of severe acute respiratory syndrome coronavirus-induced acute lung injury. *MBio.* **4** (2013).
8. Tchitchek, N. *et al.* Specific mutations in H5N1 mainly impact the magnitude and velocity of the host response in mice. *BMC Syst. Biol.* **7,** 69 (2013).
9. Askovich, P. S. *et al.* Differential host response, rather than early viral replication efficiency, correlates with pathogenicity caused by influenza viruses. *PLoS ONE* **8,** e74863 (2013).
10. Tam, V. C. *et al.* Lipidomic profiling of influenza infection identifies mediators that induce and resolve inflammation. *Cell* **154,** 213–227 (2013).
11. Mitchell, H. D. *et al.* A network integration approach to predict conserved regulators related to pathogenicity of influenza and SARS-CoV respiratory viruses. *PLoS ONE* **8,** e69374 (2013).
12. Sims, A. C. *et al.* Release of severe acute respiratory syndrome coronavirus nuclear import block enhances host transcription in human lung cells. *J. Virol.* **87,** 3885–3902 (2013).
13. Menachery, V. D. *et al.* Attenuation and restoration of SARS-CoV mutant lacking 2' O methyltransferase activity. *J. Virol.* **88,** 4251–4264 (2014).
14. Squires, R. B. *et al.* Influenza research database: an integrated bioinformatics resource for influenza research and surveillance. *Influenza Other Respir. Viruses* **6,** 404–416 (2012).
15. Pickett, B. E. *et al.* ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res.* **40,** D593–D598 (2012).
16. Peiris, J. S., Yuen, K. Y., Osterhaus, A. D. & Stohr, K. The severe acute respiratory syndrome. *N. Engl. J. Med.* **349,** 2431–2441 (2003).
17. Neumann, G. *et al.* Generation of influenza A viruses entirely from cloned cDNAs. *Proc. Natl Acad. Sci. USA* **96,** 9345–9350 (1999).
18. Watanabe, T., Watanabe, S., Kim, J. H., Hatta, M. & Kawaoka, Y. Novel approach to the development of effective H5N1 influenza A virus vaccines: use of M2 cytoplasmic tail mutants. *J. Virol.* **82,** 2486–2492 (2008).
19. Ozawa, M. *et al.* Impact of amino acid mutations in PB2, PB1-F2, and NS1 on the replication and pathogenicity of pandemic (H1N1) 2009 influenza viruses. *J. Virol.* **85,** 4596–4601 (2011).
20. Imai, M. *et al.* Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature* **486,** 420–428 (2012).
21. Hatta, M., Gao, P., Halfmann, P. & Kawaoka, Y. Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. *Science* **293,** 1840–1842 (2001).
22. Manicassamy, B. *et al.* Protection of mice against lethal challenge with 2009 H1N1 influenza A virus by 1918-like and classical swine H1N1 based vaccines. *PLoS Pathog.* **6,** e1000745 (2010).
23. Kash, J. C. *et al.* Genomic analysis of increased host immune and cell death responses induced by 1918 influenza virus. *Nature* **443,** 578–581 (2006).
24. Neumann, G. & Kawaoka, Y. Reverse genetics systems for the generation of segmented negative-sense RNA viruses entirely from cloned cDNA. *Curr. Top. Microbiol. Immunol.* **283,** 43–60 (2004).
25. Sanders, C. J. *et al.* Compromised respiratory function in lethal influenza infection is characterized by the depletion of type I alveolar epithelial cells beyond threshold levels. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **304,** L481–L488 (2013).

26. Rutigliano, J. A. *et al.* Protective memory responses are modulated by priming events prior to challenge. *J. Virol.* **84,** 1047–1056 (2010).
27. Ilyushina, N. A. *et al.* Adaptation of pandemic H1N1 influenza viruses in mice. *J. Virol.* **84,** 8607–8616 (2010).
28. Becker, M. M. *et al.* Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. *Proc. Natl Acad. Sci. USA* **105,** 19944–19949 (2008).
29. Yount, B. *et al.* Severe acute respiratory syndrome coronavirus group-specific open reading frames encode nonessential functions for replication in cell cultures and mice. *J. Virol.* **79,** 14909–14922 (2005).
30. Yount, B. *et al.* Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus. *Proc. Natl Acad. Sci. USA* **100,** 12995–13000 (2003).
31. Roberts, A. *et al.* A mouse-adapted SARS-coronavirus causes disease and mortality in BALB/c mice. *PLoS Pathog.* **3,** e5 (2007).
32. Yoshikawa, T. *et al.* Dynamic innate immune responses of human bronchial epithelial cells to severe acute respiratory syndrome-associated coronavirus infection. *PLoS ONE* **5,** e8729 (2010).
33. Fulcher, M. L., Gabriel, S., Burns, K. A., Yankaskas, J. R. & Randell, S. H. Well-differentiated human airway epithelial cell cultures. *Methods Mol. Med.* **107,** 183–206 (2005).
34. Zeng, H. *et al.* Highly pathogenic avian influenza H5N1 viruses elicit an attenuated type i interferon response in polarized human bronchial epithelial cells. *J. Virol.* **81,** 12439–12449 (2007).
35. Smyth, G. K. in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* 397–420 (Springer: New York, NY, 2005).
36. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5,** R80 (2004).
37. Kiebel, G. R. *et al.* PRISM: a data management system for high-throughput proteomics. *Proteomics* **6,** 1783–1790 (2006).
38. Jaitly, N. *et al.* Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. *BMC Bioinformatics* **10,** 87 (2009).
39. Monroe, M. E. *et al.* VIPER: an advanced software package to support high-throughput LC-MS peptide identification. *Bioinformatics* **23,** 2021–2023 (2007).
40. Zimmer, J. S., Monroe, M. E., Qian, W. J. & Smith, R. D. Advances in proteomics data analysis and display using an accurate mass and time tag approach. *Mass Spectrom. Rev.* **25,** 450–482 (2006).
41. Webb-Robertson, B. J. *et al.* Combined statistical analyses of peptide intensities and peptide occurrences improves identification of significant peptides from MS-based proteomics data. *J. Proteome Res.* **9,** 5748–5756 (2010).
42. Matzke, M. M. *et al.* Improved quality control processing of peptide-centric LC-MS proteomics data. *Bioinformatics* **27,** 2866–2872 (2011).
43. Webb-Robertson, B. J., Matzke, M. M., Jacobs, J. M., Pounds, J. G. & Waters, K. M. A statistical selection strategy for normalization procedures in LC-MS proteomics experiments through dataset-dependent ranking of normalization scaling factors. *Proteomics* **11,** 4736–4741 (2011).
44. Polpitiya, A. D. *et al.* DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* **24,** 1556–1558 (2008).
45. McDermott, J. E. *et al.* Conserved host response to highly pathogenic avian influenza virus infection in human cell culture, mouse and macaque model systems. *BMC Syst. Biol.* **5,** 190 (2011).
46. Shi, L. *et al.* The balance of reproducibility, sensitivity, and specificity of lists of differentially expressed genes in microarray studies. *BMC Bioinformatics* **9**(Suppl 9): S10 (2008).
47. Bryant, P. A., Smyth, G. K., Robins-Browne, R. & Curtis, N. Technical variability is greater than biological variability in a microarray experiment but both are outweighed by changes induced by stimulation. *PLoS ONE* **6,** e19556 (2011).
48. Xu, J. *et al.* Variability of DNA microarray gene expression profiles in cultured rat primary hepatocytes. *Gene Regul. Syst. Bio.* **1,** 235–249 (2007).
49. Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37,** 1–13 (2009).
50. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4,** 44–57 (2009).
51. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102,** 15545–15550 (2005).
52. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9,** 559 (2008).
53. Wenger, A. M. *et al.* PRISM offers a comprehensive genomic approach to transcription factor function prediction. *Genome Res.* **23,** 889–904 (2013).
54. Zambelli, F., Pesole, G. & Pavesi, G. Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. *Nucleic Acids Res.* **37,** W247–W252 (2009).
55. Waters, K. M. *et al.* Network analysis of epidermal growth factor signaling using integrated genomic, proteomic and phosphorylation data. *PLoS ONE* **7,** e34515 (2012).
56. Waters, K. M., Pounds, J. G. & Thrall, B. D. Data merging for integrated microarray and proteomic analysis. *Brief Funct. Genomic Proteomic* **5,** 261–272 (2006).
57. Reich, M. *et al.* GenePattern 2.0. *Nat. Genet.* **38,** 500–501 (2006).
58. Giardine, B. *et al.* Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* **15,** 1451–1455 (2005).

Data Citations

1. Baric, R., Katze, M., Chang, J., Gralinski, L. & Law, L. *Gene Expression Omnibus* GSE37569 (2012).
2. Josset, L. *et al. Gene Expression Omnibus* GSE45042 (2013).
3. Li, C. *et al. Gene Expression Omnibus* GSE28166 (2011).
4. Li, C. *et al. Gene Expression Omnibus* GSE37571 (2012).
5. Katze, M. G. *et al. Gene Expression Omnibus* GSE43203 (2013).
6. Katze, M. G. *et al. Gene Expression Omnibus* GSE43204 (2013).
7. Li, C. *et al. Gene Expression Omnibus* GSE33263 (2011).
8. Josset, L. *et al. Gene Expression Omnibus* GSE36328 (2011).
9. Kawaoka, Y. *et al. Gene Expression Omnibus* GSE37572 (2012).
10. Kawaoka, Y. *et al. Gene Expression Omnibus* GSE43301 (2013).
11. Kawaoka, Y. *et al. Gene Expression Omnibus* GSE43302 (2013).
12. Kawaoka, Y. *et al. Gene Expression Omnibus* GSE44441 (2013).
13. Kawaoka, Y. *et al. Gene Expression Omnibus* GSE44445 (2013).
14. Kawaoka, Y. *et al. Gene Expression Omnibus* GSE40792 (2012).
15. Aderem, A. *Gene Expression Omnibus* GSE37245 (2011).
16. Bankhead, A. *et al. Gene Expression Omnibus* GSE33267 (2011).

17. Bankhead, A. III *et al. Gene Expression Omnibus* GSE37827 (2012).
18. Baric, R., Katze, M., Sims, A., Law, G. L. & Chang, J. *Gene Expression Omnibus* GSE47960 (2013).
19. Baric, R., Katze, M., Sims, A., Law, G. L. & Chang, J. *Gene Expression Omnibus* GSE47961 (2013).
20. Baric, R., Katze, M., Sims, A., Law, G. L. & Chang, J. *Gene Expression Omnibus* GSE47962 (2013).
21. Gralinski, L. *et al. Gene Expression Omnibus* GSE33266 (2011).
22. Katze, M., Baric, R., Law, L., Gralinski, L. & Chang, J. *Gene Expression Omnibus* GSE50000 (2014).
23. Katze, M., Baric, R., Gralinski, L., Law, L. & Chang, J. *Gene Expression Omnibus* GSE51386 (2014).
24. Baric, R., Katze, M., Gralinski, L., Chang, J. & Law, L. *Gene Expression Omnibus* GSE50878 (2014).
25. Katze, M., Baric, R., Law, L., Gralinski, L. & Chang, J. *Gene Expression Omnibus* GSE51387 (2014).
26. Katze, M., Baric, R., Law, L., Gralinski, L. & Chang, J. *Gene Expression Omnibus* GSE49262 (2014).
27. Katze, M., Baric, R., Law, L., Gralinski, L. & Chang, J. *Gene Expression Omnibus* GSE49263 (2014).
28. Baric, R., Katze, M., Gralinski, L., Law, L. & Chang, J. *Gene Expression Omnibus* GSE40824 (2012).
29. Baric, R., Katze, M., Gralinski, L., Law, L. & Chang, J. *Gene Expression Omnibus* GSE40827 (2012).
30. Metz, T. *PeptideAtlas* PASS00416 (2014).
31. Metz, T. *PeptideAtlas* PASS00417 (2014).
32. Metz, T. *PeptideAtlas* PASS00418 (2014).
33. Metz, T. *PeptideAtlas* PASS00419 (2014).
34. Metz, T. *PeptideAtlas* PASS00420 (2014).
35. Metz, T. *PeptideAtlas* PASS00421 (2014).
36. Metz, T. *PeptideAtlas* PASS00422 (2014).
37. Metz, T. *PeptideAtlas* PASS00423 (2014).
38. Metz, T. *PeptideAtlas* PASS00424 (2014).
39. Metz, T. *PeptideAtlas* PASS00425 (2014).
40. Metz, T. *PeptideAtlas* PASS00426 (2014).
41. Metz, T. *PeptideAtlas* PASS00427 (2014).
42. Metz, T. *PeptideAtlas* PASS00428 (2014).
43. Metz, T. *PeptideAtlas* PASS00429 (2014).
44. Metz, T. *PeptideAtlas* PASS00430 (2014).
45. Metz, T. *PeptideAtlas* PASS00431 (2014).
46. Metz, T. *PeptideAtlas* PASS00432 (2014).
47. Metz, T. *PeptideAtlas* PASS00433 (2014).
48. Metz, T., Smith, R., Kawaoka, Y. & Katze, M. *Influenza Research Database* IRD_SV_CA04M001-P (2013).
49. Metz, T., Smith, R., Kawaoka, Y. & Katze, M. *Influenza Research Database* IRD_SV_ICL004-P (2013).
50. Metz, T., Smith, R., Kawaoka, Y. & Katze, M. *Influenza Research Database* IRD_SV_ICL006-P (2013).
51. Metz, T., Smith, R., Kawaoka, Y. & Katze, M. *Influenza Research Database* IRD_SV_ICL010-P (2013).
52. Metz, T., Smith, R., Kawaoka, Y. & Katze, M. *Influenza Research Database* IRD_SV_ICL011-P (2013).
53. Metz, T., Smith, R., Kawaoka, Y. & Katze, M. *Influenza Research Database* IRD_SV_ICL012-P (2013).
54. Metz, T., Smith, R., Kawaoka, Y. & Katze, M. *Influenza Research* Database IRD_SV_IM001-P (2013).
55. Metz, T., Smith, R., Kawaoka, Y. & Katze, M. *Influenza Research Database* IRD_SV_IM004-P (2013).
56. Metz, T., Smith, R., Kawaoka, Y. & Katze, M. *Influenza Research Database* IRD_SV_IM005-P (2013).
57. Metz, T., Smith, R., Kawaoka, Y. & Katze, M. *Influenza Research Database* IRD_SV_IM006A-P (2013).
58. Metz, T., Smith, R., Kawaoka, Y. & Katze, M. *Influenza Research Database* IRD_SV_IM006B-P (2013).
59. Metz, T., Smith, R., Kawaoka, Y. & Katze, M. *Influenza Research Database* IRD_SV_IM007-P (2013).
60. Metz, T., Smith, R., Baric, R. & Katze, M. *Influenza Research Database* IRD_SV_SCL005-P (2013).
61. Metz, T., Smith, R., Baric, R. & Katze, M. *Influenza Research Database* IRD_SV_SCL006-P (2013).
62. Metz, T., Smith, R., Kawaoka, Y. & Katze, M. *Influenza Research Database* IRD_SV_SM001-P (2013).
63. Gralinski, L., Baric, R. & Katze, M. *Influenza Research Database* IRD_SV_CA04M001-R (2013).
64. Josset, L. *et al. Influenza Research Database* IRD_SV_ECL001-R (2013).
65. Kawaoka, Y., Smith, D., Katze, M., McWeeney, S. & Waters, K. *Influenza Research Database* IRD_SV_ICL004-R (2012).
66. Eisfeld, A., Katze, M. & Kawaoka, Y. *Influenza Research Database* IRD_SV_ICL006-R (2012).
67. Ellis, A., Eisfeld, A., Chang, J., Katze, M. & Kawaoka, Y. *Influenza Research Database* IRD_SV_ICL010-R (2013).
68. Fan, S., Eisfeld, A., Chang, J., Katze, M. & Kawaoka, Y. *Influenza Research Database* IRD_SV_ICL011-R (2013).
69. Fan, S., Eisfeld, A., Chang, J., Katze, M. & Kawaoka, Y. *Influenza Research Database* IRD_SV_ICL012-R (2013).
70. Kawaoka, Y., Smith, D., Katze, M., McWeeney, S. & Waters, K. *Influenza Research Database* IRD_SV_IM001-R (2012).
71. Carter, V., Katze, M. & Tumpey, M. T. *Influenza Research Database* IRD_SV_IM002-R (2013).
72. Li, C., Eisfeld, A., Katze, M. & Kawaoka, Y. *Influenza Research Database* IRD_SV_IM004-R (2012).
73. Li, C., Eisfeld, A., Chang, J., Katze, M. & Kawaoka, Y. *Influenza Research Database* IRD_SV_IM005-R (2013).
74. Li, C., Eisfeld, A., Chang, J., Katze, M. & Kawaoka, Y. *Influenza Research Database* IRD_SV_IM006A-R (2013).
75. Li, C., Eisfeld, A., Chang, J., Katze, M. & Kawaoka, Y. *Influenza Research Database* IRD_SV_IM006B-R (2013).
76. Li, C., Eisfeld, A., Chang, J., Katze, M. & Kawaoka, Y. *Influenza Research Database* IRD_SV_IM007-R (2013).
77. Chang, J., Katze, M. & Tumpey, T. *Influenza Research Database* IRD_SV_IM009-R (2013).
78. Li, C., Eisfeld, A., Chang, J., Katze, M. & Kawaoka, Y. *Influenza Research Database* IRD_SV_IM010-R (2013).
79. Aderem, A. & Thomas, P. *Influenza Research Database* IRD_SI_SBRI_AA_E1 (2013).
80. Baric, R., Smith, D., Katze, M., McWeeney, S. & Waters, K. *Influenza Research Database* IRD_SV_SCL005-R (2012).
81. Baric, R., Smith, D., Katze, M., McWeeney, S. & Waters, K. *Influenza Research Database* IRD_SV_SCL006-R (2012).
82. Sims, A., Burkett, S., Chang, J., Katze, M. & Baric, R. *Influenza Research Database* IRD_SV_SHAE002-R (2013).
83. Sims, A., Burkett, S., Chang, J., Katze, M. & Baric, R. *Influenza Research Database* IRD_SV_SHAE003-R (2013).
84. Sims, A., Burkett, S., Chang, J., Katze, M. & Baric, R. *Influenza Research Database* IRD_SV_SHAE004-R (2013).
85. Baric, R., Smith, D., Katze, M., McWeeney, S. & Waters, K. *Influenza Research Database* IRD_SV_SM001-R (2012).
86. Gralinski, L., Chang, J., Katze, M. & Baric, R. *Influenza Research Database* IRD_SV_SM003-R (2013).
87. Gralinski, L., Chang, J., Katze, M. & Baric, R. *Influenza Research Database* IRD_SV_SM004-R (2013).
88. Gralinski, L., Chang, J., Katze, M. & Yoshihiro, K. *Influenza Research Database* IRD_SV_SM007-R (2013).
89. Gralinski, L., Chang, J., Katze, M. & Baric, R. *Influenza Research Database* IRD_SV_SM009-R (2013).
90. Gralinski, L., Chang, J., Katze, M. & Baric, R. *Influenza Research Database* IRD_SV_SM012-R (2013).
91. Menachery, V., Chang, J., Katze, M. & Baric, R. *Influenza Research Database* IRD_SV_SM014-R (2013).
92. Gralinski, L., Chang, J., Katze, M. & Yoshihiro, K. *Influenza Research Database* IRD_SV_SM015-R (2013).
93. Gralinski, L., Chang, J., Katze, M. & Yoshihiro, K. *Influenza Research Database* IRD_SV_SM019-R (2013).
94. Gralinski, L., Chang, J., Katze, M. & Yoshihiro, K. *Influenza Research Database* IRD_SV_SM020-R (2013).

Acknowledgements

Author Contributions

Brian D. Aevermann prepared manuscript materials and curated data for loading into IRD/ViPR. Brett E. Pickett prepared manuscript materials and performed data analysis. Sanjeev Kumar responsible for databasing data in IRD/ViPR. Edward B. Klem provided input on manuscript. Sudhakar Agnihothram performed SARS-CoV mouse and cell experiments. Peter S. Askovich contributed to experiment design, data analysis, and methods development. Armand Bankhead III contributed to conception and experimental design, data analysis, and computational modelling. Meagen Bolles performed SARS-CoV mouse and cell experiments. Victoria Carter processed samples for microarray analysis and performed microarray assays. Jean Chang processed samples for microarray analysis and performed microarray assays. Therese R.W. Clauss operated LC-MS instrumentation for quantitative phosphoproteomics analyses. Pradyot Dash contributed to experiment design, mouse infections, tissue collection. Alan H. Diercks contributed to experiment design, data analysis, and methods development. Amie J. Eisfel contributed to experimental design, data analysis, and sample processing; edited manuscript. Amy Ellis performed A/California/04/2009 pandemic influenza virus cell experiments. Shufang Fan performed A/Viet Nam/1203/2004 avian influenza virus cell experiments. Martin T. Ferris titered A/California/04/2009 samples. Lisa E. Gralinski performed mouse infections using A/California/04/2009 virus, contributed to experimental design. Richard R. Green data dissemination and data submissions to ViPR/IRD. Marina A. Gritsenko prepared cell samples for quantitative phosphoproteomics analyses. Masato Hatta generated mutant influenza viruses and characterized viruses in cell culture and animal models. Robert A Heegel operated LC-MS instrumentation for quantitative proteomics analyses. Jon M. Jacobs participated in design of proteomics experiments. Sophia Jeng contributed to data analysis. Laurence Josset contributed to data QC and analysis. Shari M. Kaiser contributed to experiment design, sample preparation. Sara Kelly processed samples for microarray analysis and performed microarray assays. G. Lynn Law designed experiments, monitored data quality, data analysis, data dissemination and data submissions to ViPR/IRD. Chengjun Li performed A/Viet Nam/1203/2004 avian influenza virus cell and mouse experiments. Jianing Li performed data analysis. Casey Long performed viral infections, performed plaque assays, and prepared samples. Maria L. Luna prepared cell and lung samples for quantitative proteomics analyses. Melissa Matzke performed initial proteomic data analysis. Jason McDermott contributed to conception and experimental design and performed computational analysis. Vineet Menachery performed SARS-CoV mouse and cell experiments. Thomas O. Metz managed the Center for Systems Virology's Proteomics and Metabolomics Core, designed proteomics experiments, monitored data quality, and processed quantitative proteomics data. Hugh Mitchell performed the computational analysis. Matthew E. Monroe constructed cell and lung accurate mass and time tag databases, monitored automated processing of quantitative proteomics data, and transferred proteomics raw data files to public repositories. Garnet Navarro performed sample preparation, methods development. Gabriele Neumann contributed to conception and experimental design. Rebecca L. Podyminogin performed sample preparation. Samuel O. Purvine maintained protein reference data files. Carrie M Rosenberger participated in experiment design. Catherine. J Sanders participated in experiment design, mouse infections. Athena A. Schepmoes prepared cell and lung samples for quantitative proteomics analyses. Anil K. Shukla operated LC-MS instrumentation for quantitative proteomics analyses. Amy Sims contributed to conception and experimental design, performed SARS-CoV cell and HAE experiments. Pavel Sova performed qPCR assays. Vincent C. Tam participated experiment design, sample preparation, methods development. Nicolas Tchitchek performed computational analysis, Paul G. Thomas participated experiment design, tissue collection. Susan C. Tilton performed initial proteomic data analysis. Allison Totura performed mouse infections. Jing Wang performed initial proteomic data analysis. Bobbie-Jo Webb-Robertson performed initial proteomic data analysis. Ji Wen performed initial proteomic data analysis. Jeffrey M. Weiss performed qPCR assays, data dissemination and data submissions to ViPR/IRD. Feng Yang participated in design of phosphoproteomics and monitored data quality. Boyd Yount maintained viral stocks; safety test materials for

inactivation of the SARS-CoV. Qibin Zhang participated in design of proteomics experiments. Shannon McWeeney contributed to conception, experimental design, manuscript editing and approval. Richard D. Smith contributed to conception, experimental design, manuscript editing and approval. Katrina M. Waters contributed to conception, experimental design and proteomic measurement capabilities, manuscript editing and approval. Yoshihiro Kawaoka contributed to conception, experimental design, manuscript editing and approval. Ralph Baric contributed to conception, experimental design, manuscript editing and approval. Alan Aderem contributed to conception, participated in experiment design. Michael G. Katze contributed to conception, experimental design, manuscript editing and approval. Richard H. Scheuermann prepared manuscript materials, designed the validation and usage examples, interpreted results.

Additional information

Tables 1–6 are only available in the online version of this paper.

Competing financial interests: The authors declare no competing financial interests.

Brian D. Aevermann[1], Brett E. Pickett[1], Sanjeev Kumar[2], Edward B. Klem[2], Sudhakar Agnihothram[3], Peter S. Askovich[4], Armand Bankhead III[5,6], Meagen Bolles[7], Victoria Carter[8], Jean Chang[8], Therese R.W. Clauss[9], Pradyot Dash[10], Alan H. Diercks[4], Amie J. Eisfeld[11], Amy Ellis[11], Shufang Fan[11], Martin T. Ferris[12], Lisa E. Gralinski[7], Richard R. Green[8], Marina A. Gritsenko[9], Masato Hatta[11], Robert A. Heegel[9], Jon M. Jacobs[9], Sophia Jeng[5], Laurence Josset[8], Shari M. Kaiser[4,17], Sara Kelly[8], G. Lynn Law[8], Chengjun Li[13], Jiangning Li[4], Casey Long[3], Maria L. Luna[9], Melissa Matzke[9], Jason McDermott[9], Vineet Menachery[3], Thomas O. Metz[9], Hugh Mitchell[9], Matthew E. Monroe[9], Garnet Navarro[4], Gabriele Neumann[11], Rebecca L. Podyminogin[4], Samuel O. Purvine[14], Carrie M. Rosenberger[4,17], Catherine J. Sanders[10], Athena A. Schepmoes[9], Anil K. Shukla[9], Amy Sims[3], Pavel Sova[8], Vincent C. Tam[4], Nicolas Tchitchek[8], Paul G. Thomas[10], Susan C. Tilton[9], Allison Totura[7], Jing Wang[9], Bobbie-Jo Webb-Robertson[9], Ji Wen[9], Jeffrey M. Weiss[8], Feng Yang[9], Boyd Yount[3], Qibin Zhang[9], Shannon McWeeney[5,6], Richard D. Smith[9], Katrina M. Waters[9], Yoshihiro Kawaoka[11], Ralph Baric[3,7], Alan Aderem[4], Michael G. Katze[8,15] & Richard H. Scheuermann[1,16]

[1]J. Craig Venter Institute, La Jolla, CA 92037, USA. [2]Northrop Grumman Information Systems, Health IT, Rockville, MD 20850, USA. [3]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7400, USA. [4]Seattle Biomedical Research Institute, Seattle, WA 98109, USA. [5]Oregon Clinical & Translational Research Institute, Portland, Oregon 97239-3098, USA. [6]Division of Bioinformatics and Computational Biology, Department of Medical Informatics and Clinical Epidemiology, Oregon Health Sciences University, Portland, Oregon 97239-3098, USA. [7]Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7290, USA. [8]Department of Microbiology, University of Washington, Seattle, WA 98195, USA. [9]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA. [10]Department of Immunology, St. Jude Children's Research Hospital, Memphis, TN 38105-3678, USA. [11]School of Veterinary Medicine, Department of Pathobiological Sciences, Influenza Research Institute, University of Wisconsin-Madison, Madison, WI 53706, USA. [12]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7264, USA. [13]Division of Animal influenza, Harbin Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Harbin, Heilongjiang Province 150001, China. [14]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99354, USA. [15]Washington National Primate Research Center, University of Washington, Seattle, WA 98195, USA. [16]Department of Pathology, University of California, San Diego, CA 92093, USA. [17]Present addresses: Kineta, Inc. 219 Terry Ave N Suite 300, Seattle, WA, USA (S.M.K.) and Diagnostic Discovery, Genentech, Inc., South San Francisco, CA, USA (C.M.R.).