Review

Systematic Approaches towards the Development of Host-Directed Antiviral Therapeutics

Andrew Prussia [1,2]**, Pahk Thepchatri** [1,2,]*****, James P. Snyder** [1,2] **and Richard K. Plemper** [3,4,]*****

1   Department of Chemistry, Emory University, Atlanta, GA 30322, USA;
    E-Mails: aprussi@emory.edu (A.P.); jsnyder@emory.edu (J.P.S.)
2   Emory Institute for Drug Discovery (EIDD), Emory University, Atlanta, GA 30322, USA
3   Department of Pediatrics, Emory University, Atlanta, GA 30322, USA
4   Children's Healthcare of Atlanta, Atlanta, GA 30322, USA

*   Authors to whom correspondence should be addressed; E-Mails: pthepch@emory.edu (P.T.);
    rplempe@emory.edu (R.K.P.); Tel.: +1-404-593-0547 (P.T.); +1-404-727-1605 (R.K.P.);
    Fax: +1-404-727-9223 (P.T.); +1-404-727-9223 (R.K.P.).

Abstract: Since the onset of antiviral therapy, viral resistance has compromised the clinical value of small-molecule drugs targeting pathogen components. As intracellular parasites, viruses complete their life cycle by hijacking a multitude of host-factors. Aiming at the latter rather than the pathogen directly, host-directed antiviral therapy has emerged as a concept to counteract evolution of viral resistance and develop broad-spectrum drug classes. This approach is propelled by bioinformatics analysis of genome-wide screens that greatly enhance insights into the complex network of host-pathogen interactions and generate a shortlist of potential gene targets from a multitude of candidates, thus setting the stage for a new era of rational identification of drug targets for host-directed antiviral therapies. With particular emphasis on human immunodeficiency virus and influenza virus, two major human pathogens, we review screens employed to elucidate host-pathogen interactions and discuss the state of database ontology approaches applicable to defining a therapeutic endpoint. The value of this strategy for drug discovery is evaluated, and perspectives for bioinformatics-driven hit identification are outlined.

1. Introduction

With few exceptions, therapeutic approaches to combat infectious diseases have focused in the past decades on targeting unique components or enzymes of viral, bacterial and parasitic origin. It has now become perfectly evident that this traditional, pathogen-directed strategy, while highly successful in numerous cases [1], is inherently compromised by the rapid emergence of resistance or, increasingly, pre-existing pathogen resistance to individual drugs. For example, the efficacy of current neuraminidase inhibitors for the treatment of pandemic swine origin influenza H1N1 isolates is increasingly compromised by the appearance of viral strains with pre-existing resistance in the field [2–4]. Drug-resistant variants have likewise propelled the re-emergence of highly pathogenic bacteria strains such as Mycobacterium tuberculosis decades after they were considered contained [5,6].

Prompted largely by the onset of the global HIV epidemic, the development of combination therapies based on drugs with distinct individual resistance profiles has considerably heightened the barrier against the development of pathogen resistance and frequently boosted the effectiveness of pathogen-directed therapeutics through synergistic effects [7]. Despite these successes, however, combination therapies have not conceptually addressed the problem of pathogen resistance and multidrug resistant variants that emerge frequently in clinical settings and cohorts of highly therapy-experienced patients. Developing new generations of inhibitors that are inherently prohibitive of the rapid development of resistance will rather require a new and complementary paradigm for drug discovery.

1.1. Host-Directed Antivirals, a New Paradigm for Management of Viral Diseases

Of the strategies entertained towards this goal, targeting host factors that are essential for the pathogen life cycle, rather than pathogen components directly, has recently received increasing attention [8–10]. While all pathogenic microbes experience interactions with their host organisms, viruses as obligatory parasites are directly dependent upon their host cell environment for replication, protein expression and assembly of progeny particles. It is anticipated that blocking one or more of these critical host components or cellular pathways will be resilient to the rapid development of viral resistance, since individual point mutations in viral components are unlikely to compensate for the loss of an essential host factor.

Indeed, the currently available data for the experimental use of host cyclin-dependent kinase (CDK) inhibitors to block HSV-1 and HIV-1 replication, for instance, has revealed a remarkably reduced frequency of viral escape from inhibition in tissue culture settings [11,12]. In contrast, single point mutations in viral components are fully sufficient to abrogate high-affinity binding of pathogen-directed antivirals as demonstrated by numerous studies investigating the molecular mechanism of viral drug resistance [1]. Given that replication of related viral pathogens frequently depends on

overlapping host cell pathways, host-directed antiviral strategies have high potential to move beyond the one-bug one-drug paradigm by broadening the pathogen target range of a chemical agent.

1.2. Identification of Suitable Targets for Host-Directed Antivirals

From a therapeutic perspective, the intricate and complex network of virus-host interactions yields a multitude of potential cellular targets for host-directed antivirals. In addition to the aforementioned role of CDK host protein kinases in HSV-1 and HIV-1 replication [11,12], some examples include regulatory kinases of the Abl [13] and Src [14] tyrosine kinase families, inhibition of which blocks poxvirus motility and maturation of West Nile virus particles, respectively. The Raf/MEK/ERK kinases of the mitogen-activated protein kinase (MAPK) cascades [15], when inhibited, induce nuclear retention of the influenza virus ribonucleoprotein complexes [16], preventing their export and ultimately influenza virion assembly. A further example includes inhibition of COX-2, a component of the eicosanoid biosynthesis pathway, which reduces yields of human cytomegalovirus progeny virus [17].

Clearly, the most desirable host target is essential for completion of the pathogen life cycle under investigation but at least temporarily dispensable for host cell survival, thus supporting the prospect that successful inhibition will combine a potent antiviral effect with manageable toxicity. Nevertheless, targeting host factors carries an inherently higher potential for undesirable drug-induced side effects than pathogen-directed antiviral therapies, particularly when the latter is highly selective. While host-directed therapies are being explored for the treatment of some major chronic viral infections such as HSV-1 and HIV-1 [11,12], they appear predestined for the therapy of infections by pathogens predominantly associated with severe acute disease, since anticipated treatment times and concomitant host exposure to the drug remain limited. On the other hand, chronic infections are conceivably treatable by host-protein targets where more than one gene pathway regulates the condition. In some such cases, sequential application of host-gene modifiers could control disease progression without undue side effects associated with chronic application of an anti-viral drug.

Hit candidates for host-directed drug development programs have resulted from a diverse set of experimental approaches. These can be grouped largely into knowledge-driven direct identification of individual targets, automated screening of chemical diversity sets with protocols specifically designed for the discovery of host-directed hits, and systems-wide screens for host factors essential for pathogen replication.

The evolving understanding of critical host-pathogen interactions through molecular virology research of individual viral families has made possible the direct selection of candidate host targets. With an arsenal of approved and experimental therapeutics for inhibition of many cellular pathways at hand, identified candidate targets may in many cases be immediately testable through repurposing of known drugs or commercially available experimental compounds with known bioactivity. The demonstration that Gleevec (Imatinib mesylate), an Abl tyrosine kinase inhibitor licensed for the treatment of several cancer forms [18], is a poxvirus blocker [13] and the repurposing of the MEK kinase inhibitor U0126 to block the Raf/MEK/ERK cascade for influenza virus inhibition [16,19] serve as cases in point. While the former originated from the insight that efficient vaccinia virus spread requires phosphorylation of the viral A36R protein by Abl and Src family tyrosine kinases [20,21], the

latter was triggered by the observation that influenza virus infection induces the activation of MAPK family members [16,21].

The availability of large robotic capacities in both corporate and academic settings combined with the rapid design and production of small-molecule compound libraries in the past two decades has accelerated the pace of discovery of novel drug candidates via high-throughput screening exercises. Applied to the identification of antiviral hits with a host-directed activity profile, for instance, it should be feasible to derive suitable screening protocols based on the hypothesis that host-directed candidates will likely show some cellular interference and return a broadened pathogen target range. While the former will translate into a lower primary screening score represented by the selectivity index (CC_{50}/EC_{50}), the latter should result in efficient inhibition not only of the screening agent but also of pathogens of related viral families when assessed in counter-screening assays. When we explored the general feasibility of this approach conceptually using a ~140,000-entry diversity set, several chemical compound classes were identified that efficiently blocked replication of a panel of distinct members of the myxovirus families [22]. Significantly, a subset of these revealed a host-directed activity profile in secondary assays and counter-screening exercises.
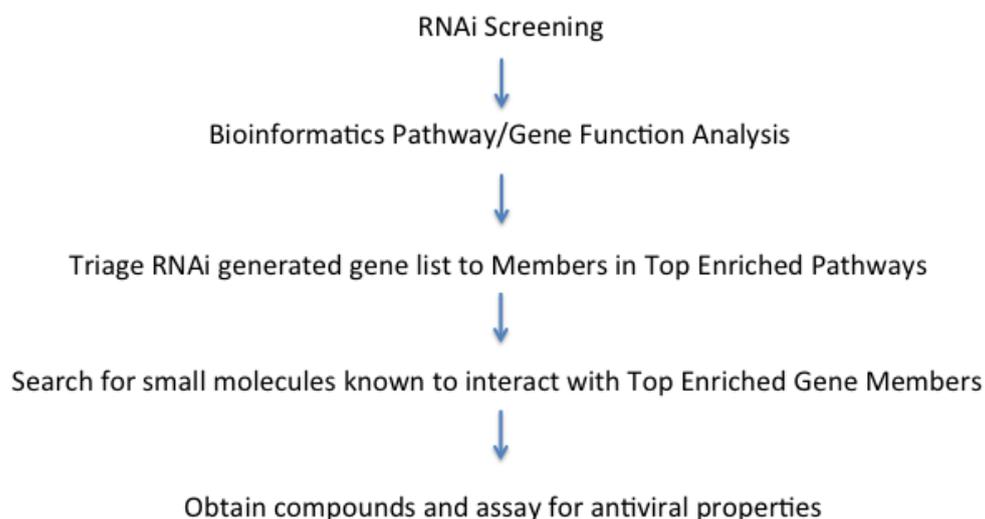
To determine the molecular target of host-directed compounds identified through screening of chemical libraries, a combination of traditional mechanism of action studies, genomics (*i.e.*, gene microarrays), and/or proteomics (*i.e.*, protein profiling) studies is conceivable. Target identification not only sets the stage for possible knowledge-based scaffold optimization through rationale design in conjunction with hit-to-lead chemistry or repurposing of known inhibitors with identical target profile, but also contributes to further elucidating critical pathogen host interactions and, thus, basic insight into pathogen biology.

A second unbiased approach for host-target identification centers on screens for host factors directly interacting with viral components or required for successful completion of the viral life cycle. Of these, avidity-based extracellular interaction screens (AVEXIS) of protein-protein contacts [23] and yeast two-hybrid screens [24,25] appear promising, although they remain inherently limited to specific pathogen factors selected as "baits". In contrast, loss-of-function screens based on aptamers [26,27] or antisense RNA interference [28–31] and gain-of-function approaches utilizing expression libraries [32,33] afford a systems-wide view of host-factors essential for pathogen replication or boosting pathogen success, respectively. In recent years, groundbreaking "loss-of-function" antisense screens were carried out for major viral pathogens including influenza virus [34,35], human immunodeficiency virus (HIV) [36–39] West Nile virus [40] and hepatitis C virus [41]. These have transformed our understanding of the virus-host interplay.

Surprisingly, however, independent large-scale screens directed at influenza virus, for instance, returned very little redundancy for essential host factors identified. This suggests that the screening efforts still lack saturation, and that cross-study bioinformatics efforts for data mining would benefit significantly by commencing on a host cell pathway rather than at the individual protein level. When a set of identified pathways essential for virus replication emerges through bioinformatics selection, individual components can be subjected to secondary screens using known drugs or experimental inhibitors to short-list desirable individual targets. A proposed workflow for a bioinformatics lead-identification process is displayed in Figure 1. In the following, we will discuss in detail the

current state of data mining approaches and how they have been, or could be, applied to screening results reported for HIV-1 and influenza A viruses.

> **Figure 1.** RNAi-Based Lead Identification Workflow. Aberrant expressed genes identified in the RNAi screen are categorized into clusters based on biological function. Members of the largest clusters are literature and database-mined for known small molecule modulators. Candidate inhibitors are subjected to biotesting for hit confirmation. This review focuses on the ability of bioinformatics methods to identify potential medicinal lead compounds.

RNAi Screening

↓

Bioinformatics Pathway/Gene Function Analysis

↓

Triage RNAi generated gene list to Members in Top Enriched Pathways

↓

Search for small molecules known to interact with Top Enriched Gene Members

↓

Obtain compounds and assay for antiviral properties

2. Methods to Analyze and Process Viral Host Factors Identified as Hits

As with the output of chemical genomic screens, genome-wide association studies (GWAS) require bioinformatics support to analyze high density data points and highlight the important gene or gene sets responsible for the phenotype under investigation. False positives need attention beyond the high throughput experiment when choosing targets for further study. These data points often arise when certain genes have been incorrectly identified due to off-target effects associated with siRNA inhibition. In one case, a single siRNA reportedly perturbed the expression of over 300 genes [42]. The potential quantity of false positives generated from a single RNAi screening experiment becomes alarming, when a single RNAi can potentially affect a large subset of gene expressions. It has been hypothesized that clustering methods such as pathway analysis and gene functional analysis are a possible means to discard false positives and highlight true positives, since they readily generate a biological interpretation of a high throughput result. A number of reviews have summarized the state of genetic analysis [43–51]. In this article, we outline the GWAS approach to antiviral identification and then highlight how such methods have been used to probe host-virus interactions.

Currently, several curated databases in the public domain detail known gene product associations (see Table 1 for examples). As will be illustrated in the following, different databases may return a diverse set of answers even when identical RNAi screening results are used as input.

Table 1. Commercial and open source pathway databases.

Database	Description	References
Kyoto Encyclopedia of Genes and Genome (KEGG)	Public Resource links genes to crystal structures and drugs when information is available	[52–54]
Reactome	Public Resource accepts a gene list for the pathway analyzer and returns percentage population per pathway	[55–57]
Protein Analysis Through Evolutionary Relationships (PANTHER)	Free pathway database allows user to identify enrichment in biological pathways, GO terms or protein class	[58,59]
WikiPathways	Community curated pathway database	[60,61]
Ingenuity IPA	Commercial pathway database to identify enrichment in pathways/GO terms; links drugs to specific genes	[62]
Gene Ontology (GO) Consortium	Community database that clusters genes by biological process, molecular function or cellular location across multiple species	[63]
Search Tool for the Retrieval of Interacting Genes/Proteins (STRING)	Freely available functional relationships database displays direct neighborhood relationships between proteins that interact directly or through an intermediary	[64–69]
Search Tool for Interactions of Chemicals (STITCH)	Crosslinks gene products with chemical structures from PubChem	[70,71]
GeneGo Metacore	Commercial manually curated pathway database annotated with 600,000 compounds	[72]
Prolexys HyNet	Commercial database protein-protein interaction identified via in-house yeast two-hybrid screening	[73]
Biomolecular Interaction Network Database (BIND)	Free and Commercial versions describing protein-protein interactions, molecular complexes and pathways	[74,75]
Molecular Interactions Database (MINT)	Public protein-protein interaction database based on peer-reviewed literature. Accessible through web-interface or Simple Object Access Protocol/Representational State Transfer (SOAP/REST) protocols	[76]
Human Protein Reference Database (HPRD)	Public proteonomic database with descriptions for 2750 human proteins taken from the primary literature	[77–79]

Application of GWAS frequently employs graphical networks displaying interconnected genes as exemplified in Figure 2. The ability to interconnect and cluster genes by known function forgoes the need to verify every target that is generated by an RNAi screen through a multitude of single biological experiments. Topological representations of biological function identify enriched regions of perturbed gene expression and their relevant cellular operations. Usefulness of gene enrichment analyses depends

on the quality of the database used [80]. If accurate, gene enrichment separates genes from those that were randomly identified during the RNAi screening and offers the largest source of antiviral target candidates for drug development, based on the assumption that all members of an enriched region are required for efficient viral replication. The fundamentals of network theory and its usage in GWAS analysis are reviewed in [80]. Gene expression studies can utilize enhanced methods of pathway analysis which go beyond treating pathways as simple sets of genes and incorporate the complex gene interactions described by the pathway, such as measurement of total pathway perturbation [81]. Unfortunately such methods require quantitative differential expression data to be applicable, a component lacking in these RNAi screening studies. It is possible, however, that inclusion of RNAi viral inhibition data in a modified version of this enhanced method might lead to improvement in pathway enrichment, but this is yet to be exemplified in the literature.

Figure 2. Network Association Map of RNAi screening results generated by the Brass *et al*. influenza virus infection. The list of perturbed host cell genes were complied in STRING and illustrated as nodes above. Lines between different nodes (edges) represent protein interactions that are either known experimentally (purple) or predicted computationally (yellow). Significant nodes such as those shown around COPA and CRNLK1 suggest these pathways to be critical for the viral life cycle.
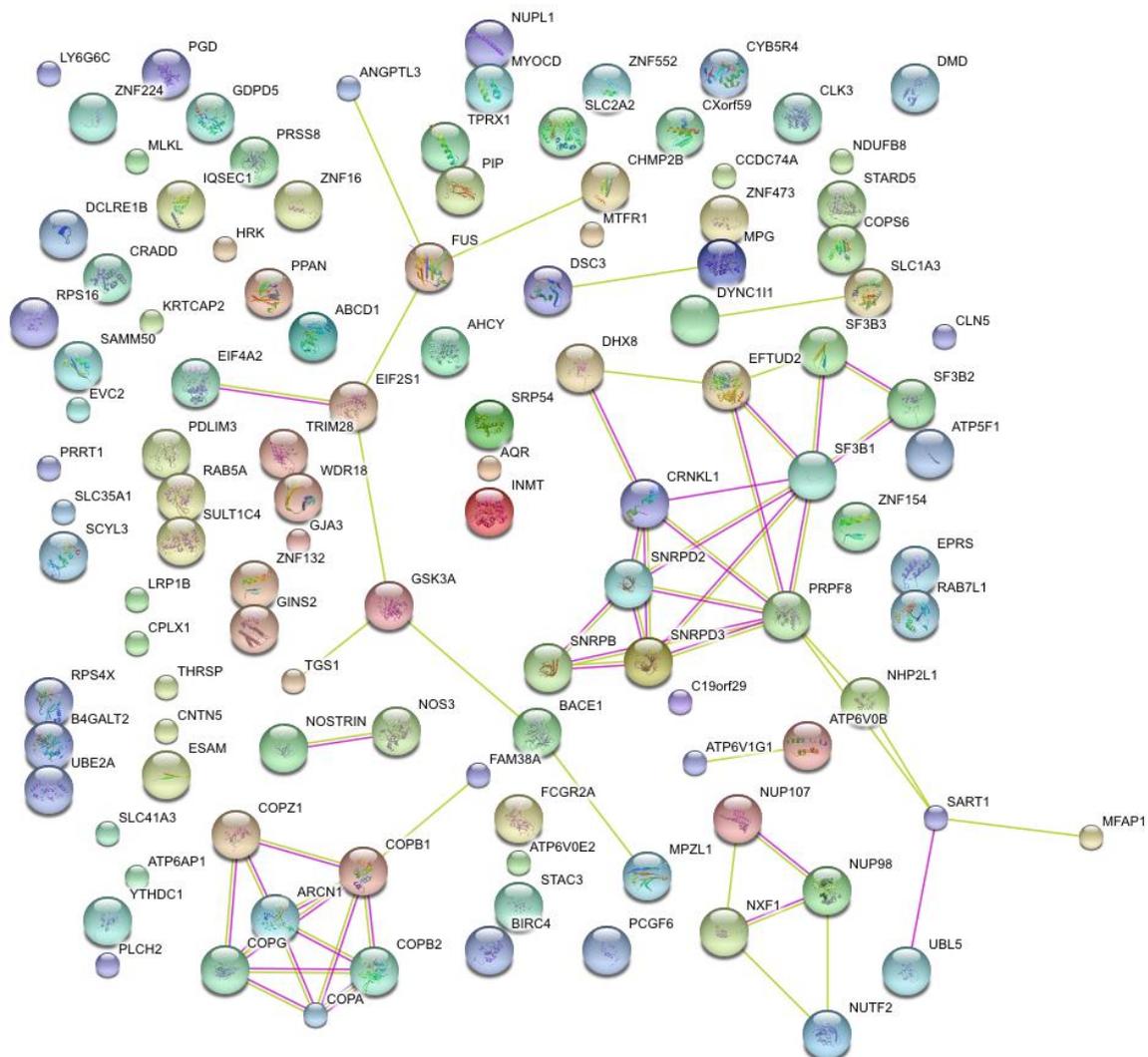
Thus, pathway analysis and gene function annotation offer the drug screening community automated procedures for extracting meaning from a large array of differentially expressed genes. The identified genes can be grouped according to the relationships among protein products, while potential drug targets associated with enriched pathways can be perceived.

An illustration of the identification of therapeutic compounds by means of bioinformatics analysis is the use of a gene list to link Tamoxifen to the treatment of Systemic Lupus Erythematosus (SLE) [82–84]. Investigators discovered that the estrogen receptor pathway is a significantly enriched gene category with respect to a list of SLE's aberrantly expressed genes through the Disease-Drug Correlation Ontology (DDCO). Other databases that similarly infuse chemical knowledge into the pathway databases include KEGG, the Connectivity Map, Ingenuity IPA and the STITCH database (all of which are also included in Table 1). Figure 3 offers an example of a gene interaction map annotated with small molecules.

Figure 3. Gene interaction map overlapped with Tamoxifen via the STITCH database. The latter also connects ovals to one another suggesting that these molecules display similar biological behavior towards the same target. Edges refer to interactions as determined by experiment (purple), manual curation (cyan) or computationally predictions (yellow).
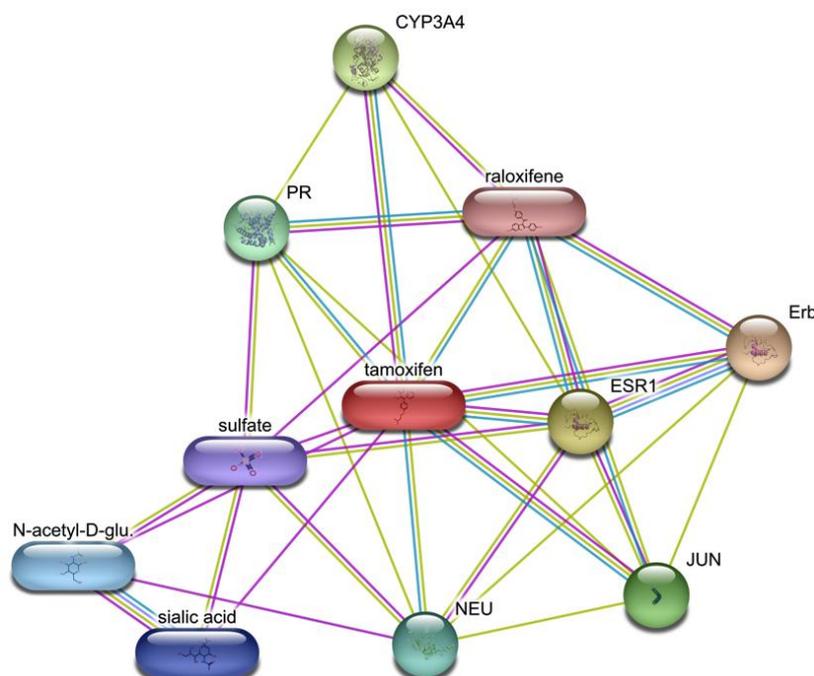
While the previously cited resources offer investigators utmost convenience in immediately accessing lists of available small molecule modulators related to a pathway of interest, other databases connect small molecule modulators with known protein targets. These require a separate pathway analysis to choose a particular set of gene products of interest. Suitable resources include DrugBank [85,86], PDB/sc-PDB [87–89], PubChem [90], Sunset Molecular's WOMBAT-PK 2010 [91,92] and the BRENDA database [93].

Following a bioinformatics selection of target candidates, individual targets must be selected for medicinal chemistry, for instance, based on the previous discovery of small-molecule blockers

or the availability of crystal structures in the Protein Data Bank (PDB). [88,89]. Then, *de novo* drug discovery can be sustained through 3D virtual screening [94] and structure-based design. Application of these methods to the analysis of HIV and flu RNAi screens will be discussed in the next section.

3. Bioinformatics Approaches for Identifying Host-Factors Required for HIV Replication

Each of the systematic studies examined in the following sections employed a unique bioinformatics approach to pathway analysis. Similar to a chemoinformatics clustering analysis of a high-throughput screen to short-list a set of chemical leads for optimization, a goal of the RNAi screening studies is to identify, by means of gene pathway or functional analysis, potential host factor targets that are essential for viral replication. A key question is whether the resulting bioinformatics short list of host factors contains suitable candidates for drug development.

3.1. Bioinformatics Approaches to Identify Host-Factors Required for HIV Virus Replication

For HIV, three independent siRNA studies were published in 2008 by Brass *et al.* [36], Konig *et al.* [95] and Zhou *et al.* [37]. All three siRNA studies utilized the National Center for Biotechnology Information (NCBI) database of HIV-1 and human protein interactions (currently 1443 proteins identified) to evaluate the overlap of hit genes with the curated virus-host interactions available in the NCBI database [61]. Figure 4 illustrates the total number of genes found as well as the pairwise overlap between genes in each study. A meta-analysis of these genome-wide studies was subsequently performed by Bushman *et al*. in 2009 [96].

Figure 4. Illustration of the pairwise overlap between hit genes in the three HIV siRNA studies and the NCBI database. Circle areas are proportional to the number of genes. For clarity, three-way and higher overlaps are not shown.
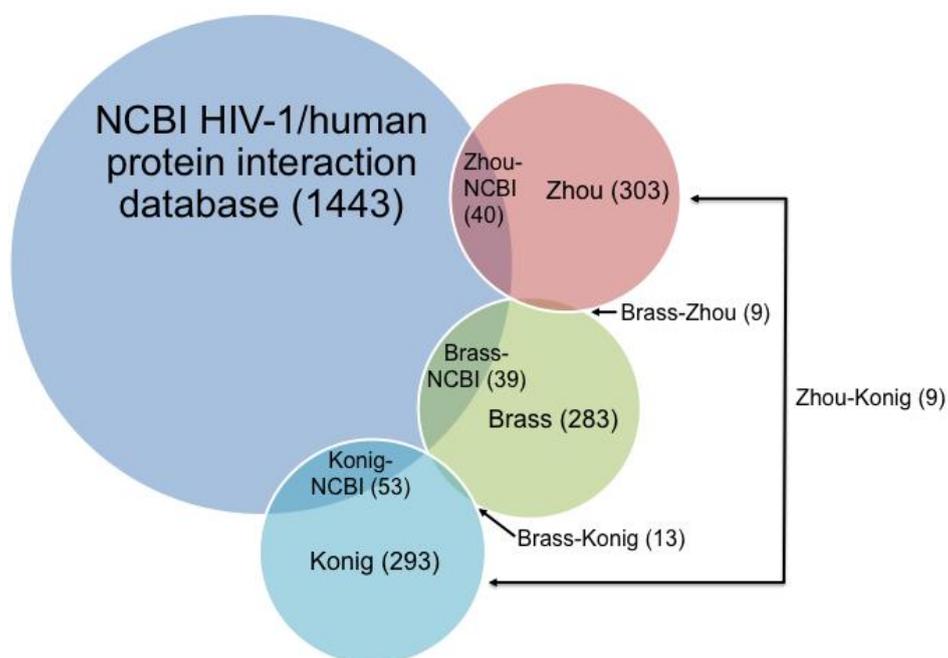
Bushman *et al.* performed an overlap analysis/random distribution comparison based on these data and found associations that were statistically significant (*p*-values < 0.001). While one may safely assume that the hit genes are enriched with respect to independently identified and confirmed host factors required for HIV-1 replication, pairwise overlaps between the studies are low, ranging only from 3 to 6%. While these were still judged statistically significant (*p*-values < 0.024 for all pairs) [96], the overall very low redundancy suggests considerable experimental variability associated with each siRNA screen.

Variation in individual host factors could be accounted for by a number of factors, including: (a) high experimental variance of siRNA transfection efficiencies [42]; (b) harvest of cells at different time points post-infection; (c) the use of different analysis methods and filtering thresholds; (d) an inherent bias of individual assays towards specific stages of the viral life cycle [96]; and (e) overall moderate reproducibility of siRNA-based screens[97]. Some of these variances might be readily controlled by additional replicates examined per screen. For instance, only the study by Konig *et al.* performed the screen in duplicate. As a case in point, the experimental data showed large variances between the replicates: 24% of hit siRNAs (141) exhibit standard deviations greater than 25% of their median values. Furthermore, Bushman *et al.* demonstrated that adjusting the filtering thresholds in this study strongly influences the nature of the identified genes (shown Figure 1D of Bushman *et al.*) [96]. Other parameters, such as non-uniform harvesting time points, are inherent to the design of each individual study and cannot be standardized retroactively. Although capturing different stages of the viral life cycle in separate studies may ultimately be necessary to fully appreciate the scope of the host-pathogen interaction network, different analysis times should be considered as a major contributor to the low level of congruity between the currently available data.

Independent of redundancy between studies, the question remains of whether the gene hits represent *bona fide* host factors required for HIV replication or false positives that may have arisen from experimental variability. Equally important for hit confirmation is the organization of the data sets into groups by gene function and cellular pathways to illuminate distinct parts of the intricate host-pathogen interaction network. Using terms from the Gene Ontology (GO) database Brass *et al.* noted that 103 of their hit genes were assigned with 136 statistically significant (*p*-value < 0.05) biological processes [36]. In brief, the GO database is a consortium established to relate genes to one another in a fixed file format within three categories: biological processes, cellular components and molecular functions [44,63,98,99]. To reduce redundancy, these categories were clustered and manually curated. GO analysis yielded 17 enriched cellular functions in the Brass HIV study. Alternatively, the Zhou study used Ingenuity Pathway Analysis to determine enriched molecular functions and biological pathways. Thirty-two molecular functions were identified, and twelve biological processes were found to be statistically significant (*p*-value < 0.05).

In contrast, Konig and colleagues employed a multi-tiered bioinformatics approach to identify the host factors most important to HIV replication through the use of the Prolexys HyNet database [95]. This resulted in networks of 2468, 4080, and 2850 genes in the HIV, MLV, AAV and toxicity assays, respectively. Using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) [100], the Konig team extracted overrepresented functional clusters for all genes found in the three HIV siRNA studies. Filtering for significance (*p*-value < 0.06 based on a geometric mean for all the terms in a group), redundancy, biological relevance and specificity returned 24 functional groups (A listing

of the overlapping pathways and those identified uniquely are presented in Figure 5), most of which contained genes that were identified in two or more studies. Although each study contains a significant number of genes that may be defined by molecular functions common to each of the three studies, consistent function identification across the siRNA screens is lacking due to the distinctions in each study's bioinformatics methods. These functions may be defined differently between methods due to redundancies present in each database. Slightly different but biologically meaningless distinctions can arise thereby.

Figure 5. Comparison of HIV-dependent host functions identified by Bushman *et al*. [96], Brass *et al*. [36], Konig *et al*. [95] and Zhou *et al*. [37]. Grey boxes indicate functions unique to an individual study.
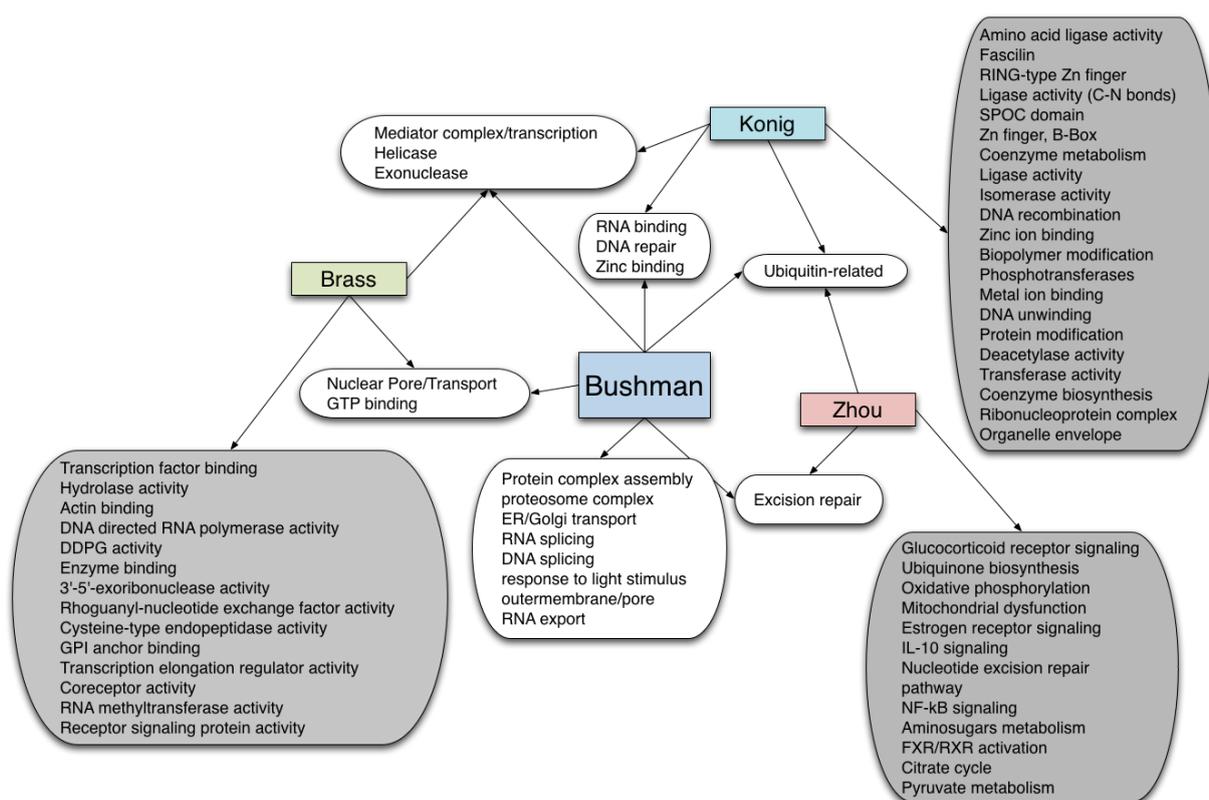
Although some well-documented host factors required for HIV replication (such as CD4, CXCR4, NFκB subunit RELA, activating kinases AKT1 and JAK1, TSG101, and various cofactors of Vpr, Vif, Tat, and Rev) [101–105] were identified in at least one of the three siRNA studies, a variety of other host factors known to engage with HIV (HLA-B57, HLA-C, PSIP1/LEDBF/p75, Sp1, cyclophilin A, ITGB1, ITGB2, and ITGB3) were not discovered [96,106–115]. Of these, the absence of the integration cofactor PSIP1/LEDBF/p75, HIV long terminal repeat transcription factor Sp1, HIV Gag binding protein cyclophilin A, and the three integrin proteins (ITBG1, ITBG2, ITBG3) is most noticeable. [108–112,114–117] These results suggest that even a combined host factor analysis is at risk of missing key host components required for viral replication. Furthermore, no single siRNA study thus far illuminates all relevant, and currently known, cellular factors associated with the pathogen.

The meta-analysis was extended by building an HIV-host factor interaction network of 1657 cellular proteins using an array of protein-protein interaction databases (BIND, HPRD, MINT, and

Reactome) [96]. With MCODE's graph theoretic clustering algorithm, clusters within this interactome map having different functions were identified. Of the 11 clusters found, 10 were associated with distinct cellular functions: the proteasome, transcription/RNA polymerase, the mediator complex, Tat activation/transcriptional elongation, RNA Binding/Splicing, BiP/GRP78/HSPA5 Chaperone, and CCT Chaperone. From a drug discovery perspective, however, small molecule testing or counter-screening with individual siRNAs against target candidates are required to validate individual pathways.

Lack of saturation in host genes identified through current siRNA screens, varying consistency and high overlap of genes in specific areas emerge as future challenges for application to system-wide drug discovery efforts.

3.2. Bioinformatics Approaches to Identify Host-Factors Required for Influenza Virus Replication

In addition to application of system-wide siRNA screens to the HIV system, the technology was applied to the influenza virus. Major siRNA studies were reported by Hao *et al*. [118], Brass *et al*. [119], Shapira *et al*. [120], Konig *et al*. [34] and Karlas *et al*. [35]. Other types of screens were performed by Josset *et al*. [121], which identified a gene list based on gene expression response to influenza; and Coombs *et al*. [122], which performed a quantitative analysis of protein level changes in infected cells. While Hao and colleagues employed a Drosophila cell-based host system for their siRNA screens, both Konig and Karlas relied on human lung cells (A549) and the influenza A/WSN (H1N1) strain or a recombinant variant thereof. Brass and colleagues used a human osteosarcoma cell system (U2OS) and the influenza A/PR/8 (H1N1) strain. The Shapira study is unique in that it combined results for yeast two- hybrid analyses, genome-wide transcriptional gene expression profiling and siRNA screening. Unfortunately, a single publically available resource similar to the NIH/NIAID HIV-1 interaction database does not exist for influenza virus, although many distinct virus-host interactions have been described in the literature (reviewed in [123]).

Watanabe *et al.* summarized five of the six systematic studies reported above and performed bioinformatics analysis on the 1,449 identified genes required for influenza replication [123]. Much like the Bushman *et al*. analysis of HIV host factors [96], 128 genes were found in multiple screens when pairwise comparisons were performed. The highest pairwise overlap (32 genes) is found in the Konig *et al*. and Karlas *et al*. studies, possibly due to the overlaps in host cell type (A549 cells) and influenza virus strain (A/WSN (H1N1)). This lack of overlap is also illustrated in the pairwise analysis given in Table S1. Unlike the meta-analysis for the HIV studies discussed above, this pairwise comparison lacked random distribution simulations, preventing the assessment of statistical significance. Nevertheless, the observed low overlap rate most likely results from factors similar to those discussed above for the HIV siRNA studies, *i.e.*, different harvest times, detection thresholds and host cell lines, coupled with the additional complication of variability introduced through the use of different viral strains.

As described in the HIV siRNA analyses, each study examining influenza virus infections performed individual bioinformatics analyses on siRNA screening results. A summary of these bioinformatics data along with the methodology is reviewed by Min [124]. In unique congruence, three of the influenza virus studies explored the use of known small-molecule inhibitors to obtain

independent proof-of-concept for the importance of cellular targets identified by bioinformatics for virus replication [34,35,121].

Konig reported six compounds with EC_{50} values ranging from 0.5 to 30 μM target FRAP1, HSP90AA1, TUBB, FGFR4, GSK3B, or ANPEP. Of these, FRAP1, TUBB, FGFR4 and GSK3B home to the same GO Term cluster, protein kinase activity, recommending it as a potentially rich source for influenza virus inhibitors. The cytosolic chaperone Hsp90AAP1 was identified in a separate GO Term cluster; interestingly, previous reports have already established a link to influenza virus [125] and HCV [126] replication. The Karlas study reported another efficacious small molecule inhibitor, TG003, which targets the CDC-like kinase 1 (CLK1). CLK1 was retrieved from the Spliceosome GO term cluster, where it ranked seventh in significance among the list of enriched cellular components.

While the above studies identified potential host factor targets through GO term enrichment and then followed up with small molecules available for viral testing, the Josset project searched Connectivity Map with 20 of the most perturbed genes from the 300 initially identified. Of the eight compounds available through commercial vendors, six attenuated influenza virus replication with EC_{50} values ranging from 5.8 to 30 μM. In-house analysis of the 20 genes used to identify these compounds revealed that they are significantly enriched in metabolic processes. Similar to previous studies, the Josset report does not explicitly identify a host pathway essential for viral replication based on the small molecule inhibition studies. To appreciate the full potential of this approach for antiviral drug development, it may be informative to collect all known inhibitors of a particular host pathway and determine the complete extent of virus inhibition.

Watanabe *et al.* performed a meta-analysis of the siRNA results using the set of 128 genes found in two or more studies [123]. The major gene categories were determined through PANTHER, a database that also utilizes GO terms to organize gene lists. Several molecular functions were found significant: nucleic acid-binding proteins, kinases, transcription factors, ribosomal proteins, hydrogen transporters and proteins related to mRNA splicing. Biological processes found to be consequential were protein metabolism and modification, signal transduction, protein phosphorylation, nucleoside, nucleotide and nucleic acid metabolism and intracellular transport. Reactome analysis tagged as significant eukaryotic translation initiation, regulation of gene expression, processing of capped intron-containing pre-mRNAs and Golgi-to-ER retrograde transport. This set of 128 genes was further integrated with the viral protein interaction partners determined by Konig and Shapira, resulting in a network of virus-host interactions. Based on this map, MCODE further identified translation initiation, mRNA processing and proton-transport as crucial. Accordingly, mining of the top MCODE cluster in Figure 6 predicts that compounds such as spectoinomycin, emetine and quercetin will interfere with influenza virus replication.

Successful outcomes for bioinformatics searches predominantly depend on the accuracy of tabulated database interactions. As detailed below, use of different databases may alter the profile of pathways that are enriched from the same gene list. In such cases, users are obligated to formulate a realistic biological interpretation of the relational data to ensure identification of meaningful candidate compounds for an antiviral drug program.

Figure 6. Small molecule (ovals) identification of gene products (spheres) associated with translation initiation. Green edges represent protein-ligand interactions. These compounds have not been reported previously to interfere with influenza infection, although quercetin has been demonstrated to attenuate HCV, however through a different host factor [126].
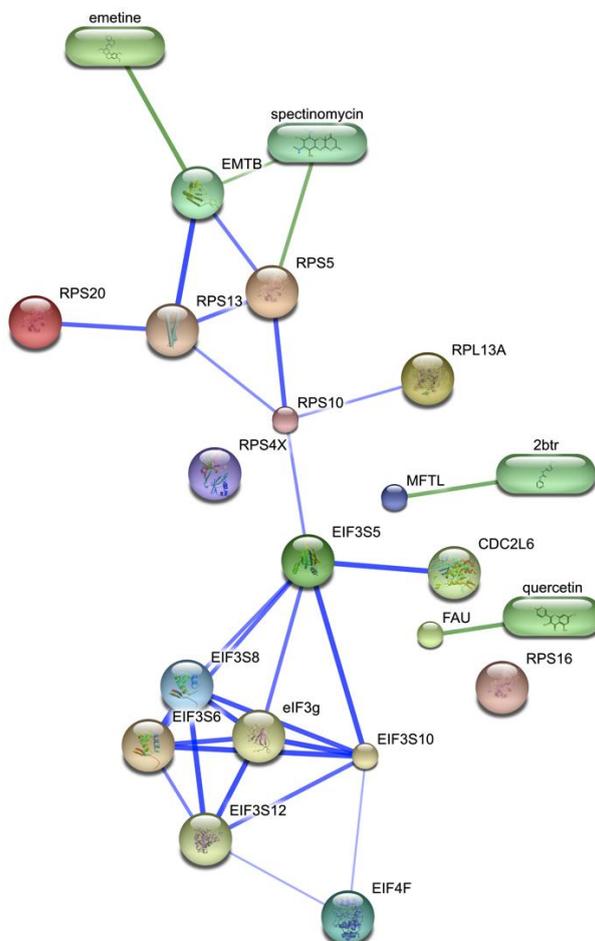
4. Pathway Database Comparisons: Same Source, Different Interpretation

As outlined above, it is a primary function of gene databases to extract biological meaning as well as potential therapeutic host factors from a high throughput RNAi screen by means of descriptive annotations of genes common to a particular biological pathway or gene function. In the realm of antiviral drug discovery, this approach aims at identifying host cell components critical for virus replication.

Crucial for the success of this strategy is the quality of the pathway database used, which is determined by the curation method of published experimental data of gene associations and the expertise of the curators involved. Soh *et al.* have demonstrated that inconsistencies emerge when gene association data are compared across different pathways databases [127]. This came as a surprise, since most databases share published literature as a data source, suggesting that methodology for curation and criteria for gene associations were not uniform (for this study, the Ingenuity IPA, KEGG, and Wikipathways databases were compared). Assuming curation is performed on available literature data, however, one expects similar genes and gene pairs to be found across the different databases. (Gene pairings are defined as gene product associations confirmed by the database curator).

The Wnt signaling pathway provides a tangible example illustrating the current challenges. This pathway has been implicated in therapeutic interference with cancer and viral entry. Two-way analyses revealed approximately 80% gene similarity when based on the KEGG and Wikipathways databases (Table 4 in [127]). However, only 43% similarity is found when Ingenuity and KEGG are examined (Table 3 in [127]), while comparison of Ingenuity and Wikipathways databases returns only 28% similarity (Table 5 in [127]). Inconsistencies across databases are even more disconcerting when gene pair overlaps are examined: KEGG/Wikipathways (18%), Ingenuity/KEGG (8%), Ingenuity/Wikipathways (0%). It is the quality of the gene pairing data in each database, however, that allows end users to triage the multiple RNAi screening results for pathway congruity.

Looking on a broader scale across 26 cellular pathways described in Soh *et al.*, gene overlap similarity has a mean value of 66.5% when comparing the KEGG database and Wikipathways [127]. By contrast, the mean values of similarity for Ingenuity/KEGG and Ingenuity/Wikipathways were 53.8% (12 pathway categories) and 41% (11 pathway categories), respectively. Despite the higher gene overlap between KEGG and Wikipathways, the pairing overlap is still only approximately 50% for any listed pathway compared across any three of the databases. KEGG is curated by a single lab group, while Wikipathways is curated through a community effort. At the moment, it is not clear to what extent the curation procedures contribute to the highly variable data mismatches. However, there is little doubt that this and other variables would benefit from cross-consolidation between the various databases.

Soh *et al.* also analyzed the comprehensiveness of the databases, which was a measure for the total number of genes from all three databases [127]. This was followed by evaluation of gene members and pairings of each database against the pool, which consisted of 21,314 genes and 60,900 pairings. KEGG was shown to be the most comprehensive of the three databases, but this was influenced by the result of KEGG's inclusion of metabolic pathways specifically not curated by either of the other databases.

Concentrating in particular on viral host factors, we performed an in-house analysis that compares host proteins involved in the influenza virus life cycle across various databases. Databases used in this example included Reactome, Ingenuity IPA and PANTHER. The Reactome database records six host factor genes for influenza in each the categories associated with NS1-mediated effects and virus-induced apoptosis. Databases such as Ingenuity IPA and PANTHER lack pathway categories dedicated to influenza virus. Keyword searches for influenza in the PANTHER database identified no host factor associated with influenza virus infection [58]. Conversely, keyword searching of the Ingenuity database generated a list of five signaling pathways (Lipids/Lipid Rafts, MAPK, PI3K/AKT, Wnt/GSK-3β, hypercytokinemia) involved in the pathogenesis of influenza virus, constituting a list of 38 genes [128]. Import of the latter into all other databases allows the genes to be categorized into signaling pathways such as Wnt, PI3K, and MAPK. However, database annotations suggest that Ingenuity is more likely to alert the user to the genes' roles in influenza virus infection.

Applying this approach to the previously described influenza RNAi screens, we sought to address the question of how does target identification change when different pathway databases are applied to the same dataset? Databases used in this comparison were PANTHER, Reactome and STRING, and the data set analyzed was the commonly identified 128 gene list generated by Watanabe *et al.* [123]. Results are presented in Table 2 with reference to the Wanatabe analysis of the same genes.

Table 2. Comparison of pathway results from the Watanabe pairwise influenza gene set.

GeneGO/MCODE	STRING	PANTHER	Ingenuity IPA	Reactome
Translation Initiation	Translation Initiation	Apoptosis signaling pathway	Chronic Myeloid Leukemia Signaling	Dissolution of Fibrin Clot
Pre-mRNA Processing	Pre-mRNA Processing	T cell activation	B Cell Receptor Signaling	Influenza Life Cycle
Proton-Transporter V-type ATPase	Proton-Transporter V-type ATPase	Angiogenesis	Production of Nitric Oxide and Reactive Oxygen Species in Macrophages	MAP kinase cascade
	COPI coating of Golgi vesicle	Toll receptor signaling	EIF2 Signaling	Metabolism of nitric oxide
	Nuclear Transport	Inflammation mediated by chemokine and cytokine signaling pathways	Rank Signaling in Osteoclast	Eukaryotic Translation Initiation
		Cell cycle	CD40 Signaling	Signaling by FGFR
		PDGF signaling	Molecular Mechanisms of Cancer	Eukaryotic Translation Termination
		FGF signaling	Role of PKR in Interferon Induction and Antiviral Response	Eukaryotic Translation Elongation
		FAS signaling		Regulation of beta-cell development
		Ras Pathway		Signaling by Insulin receptor
		B cell activation		Processing of Capped Intron-Containing Pre-mRNA

It becomes immediately obvious when examining the most enriched pathways that only the STRING database seems to reproduce the results generated by Watanabe *et al.* using GeneGO/MCODE. In all other cases, the different databases returned remarkably different top pathways when the same gene expression set was analyzed. Closer examination reveals that that other top ranking pathways (*i.e.*, translation initiation) rank lower on the Reactome enrichment analysis scale. Pathways associated with B-Cell metabolism are also identified by the Ingenuity IPA and Reactome enrichments, although slightly different naming schemes are used. Since discrete databases identify certain similar pathways at different rankings, a consensus scoring function applicable to available databases appears warranted. This would afford greater confidence in the identification of individual targets for follow-up through small molecule searching.

5. Conclusions

Previous GWAS experiments have attempted to capture the most relevant cellular host pathways utilized by pathogens such as HIV and influenza virus for virus replication [96,123]. As shown by reviewers such as Bushman *et al.* and Watanabe *et al.*, gene lists and enriched pathways vary widely despite the pursuit of similar biological goals. Indeed, the likelihood to successfully identify novel host-directed antivirals would increase significantly if the reproducibility of individual RNAi screens were to be increased [97]. Further challenges emerge from differently curated pathway databases that return unrelated enriched pathways based on analysis of the same gene data set. Preliminary analysis of this situation using the Watanabe 128 pairwise genes suggests that a consensus scoring protocol applicable across different databases would be desirable to clarify this issue. Despite these hurdles associated with experimental false positives and the complexities inherent in interpreting pairwise gene interactions, several tangible examples (*i.e.*, Konig *et al.*, Karlas *et al.*, and Josset *et al.*) demonstrate that RNAi screening coupled with bioinformatics-driven triaging is a viable method to identify small molecule inhibitors of virus replication.

Current databases that infuse chemical knowledge into schemes such as Ingenuity IPA and the Connectivity Map are limited to a small number of compounds, mostly FDA-approved drugs. This narrow focus limits their application to current translational medicine. The STITCH database makes an interesting leap by crosslinking its gene network with multiple chemical-genomic high throughput screening results archived in PubChem. These experimental chemicals along with compounds currently tested *in vitro* for various endpoints offer a rich source for hit candidates with optimization potential. As more databases are used to analyze potential host targets, validation methods employing siRNA are improved and small molecule knowledge is added to the genetic web, more drug discovery initiatives are likely to incorporate this approach in their portfolio of standard operations for the identification of antiviral therapeutics.

Acknowledgments

Supporting Information

Table S1. Pairwise comparison of influenza genome-wide studies.

	Konig [129]	Karlas [130]	Brass [131]	Shapira [132]	Josset [133]	Coombs [134]
Konig		32	9	16	10	1
Karlas	32		12	18	5	3
Brass	9	12		10	6	2
Shapira	16	18	10		20	15
Josset	10	5	6	20		3
Coombs	1	3	2	15	3	

References

1. Coen, D.M.; Richman, D.D. Antiviral Agents. In *Fields Virology*, 5th ed.; Knipe, D.M., Howley, P.M., Eds.; Wolters Kluwer/Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2007; Volume 1, pp. 447–485.

2. Moore, C.; Galiano, M.; Lackenby, A.; Abdelrahman, T.; Barnes, R.; Evans, M.R.; Fegan, C.; Froude, S.; Hastings, M.; Knapper, S.; *et al*. Evidence of person-to-person transmission of oseltamivir-resistant pandemic influenza A(H1N1) 2009 virus in a hematology unit. *J. Infect. Dis.* **2011**, *203*, 18–24.

3. Sheu, T.G.; Fry, A.M.; Garten, R.J.; Deyde, V.M.; Shwe, T.; Bullion, L.; Peebles, P.J.; Li, Y.; Klimov, A.I.; Gubareva, L.V. Dual resistance to adamantanes and oseltamivir among seasonal influenza A(H1N1) viruses: 2008–2010. *J. Infect. Dis.* **2011**, *203*, 13–17.

4. Hayden, F.G.; de Jong, M.D. Emerging influenza antiviral resistance threats. *J. Infect. Dis.* **2011**, *203*, 6–10.

5. Walsh, C. Molecular mechanisms that confer antibacterial drug resistance. *Nature* **2000**, *406*, 775–781.

6. Cegelski, L.; Marshall, G.R.; Eldridge, G.R.; Hultgren, S.J. The biology and future prospects of antivirulence therapies. *Nat. Rev. Microbiol.* **2008**, *6*, 17–27.

7. Schinazi, R.F.; Peters, J.; Williams, C.C.; Chance, D.; Nahmias, A.J. Effect of combinations of acyclovir with vidarabine or its 5'-monophosphate on herpes simplex viruses in cell culture and in mice. *Antimicrob. Agents Chemother.* **1982**, *22*, 499–507.

8. Kellam, P. Attacking pathogens through their hosts. *Genome Biol.* **2006**, *7*, 201.

9. Schwegmann, A.; Brombacher, F. Host-directed drug targeting of factors hijacked by pathogens. *Sci. Signal.* **2008**, *1*, re8.

10. Tan, S.L.; Ganji, G.; Paeper, B.; Proll, S.; Katze, M.G. Systems biology and the host response to viral infection. *Nat. Biotechnol.* **2007**, *25*, 1383–1389.

11. Koon, H.B.; Bubley, G.J.; Pantanowitz, L.; Masiello, D.; Smith, B.; Crosby, K.; Proper, J.; Weeden, W.; Miller, T.E.; Chatis, P.; *et al*. Imatinib-induced regression of AIDS-related Kaposi's sarcoma. *J. Clin. Oncol.* **2005**, *23*, 982–989.

12. Salerno, D.; Hasham, M.G.; Marshall, R.; Garriga, J.; Tsygankov, A.Y.; Grana, X. Direct inhibition of CDK9 blocks HIV-1 replication without preventing T-cell activation in primary human peripheral blood lymphocytes. *Gene* **2007**, *405*, 65–78.

13. Reeves, P.M.; Bommarius, B.; Lebeis, S.; McNulty, S.; Christensen, J.; Swimm, A.; Chahroudi, A.; Chavan, R.; Feinberg, M.B.; Veach, D.; *et al*. Disabling poxvirus pathogenesis by inhibition of Abl-family tyrosine kinases. *Nat. Med.* **2005**, *11*, 731–739.

14. Hirsch, A.J.; Medigeshi, G.R.; Meyers, H.L.; DeFilippis, V.; Fruh, K.; Briese, T.; Lipkin, W.I.; Nelson, J.A. The Src family kinase c-Yes is required for maturation of West Nile virus particles. *J. Virol.* **2005**, *79*, 11943–11951.

15. Pearson, G.; Robinson, F.; Beers Gibson, T.; Xu, B.E.; Karandikar, M.; Berman, K.; Cobb, M.H. Mitogen-activated protein (MAP) kinase pathways: Regulation and physiological functions. *Endocr. Rev.* **2001**, *22*, 153–183.

16. Pleschka, S.; Wolff, T.; Ehrhardt, C.; Hobom, G.; Planz, O.; Rapp, U.R.; Ludwig, S. Influenza virus propagation is impaired by inhibition of the Raf/MEK/ERK signalling cascade. *Nat. Cell Biol.* **2001**, *3*, 301–305.

17. Zhu, H.; Cong, J.P.; Yu, D.; Bresnahan, W.A.; Shenk, T.E. Inhibition of cyclooxygenase 2 blocks human cytomegalovirus replication. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 3932–3937.

18. Goldman, J.M.; Druker, B.J. Chronic myeloid leukemia: current treatment options. *Blood* **2001**, *98*, 2039–2042.

19. Ludwig, S. Targeting cell signalling pathways to fight the flu: Towards a paradigm change in anti-influenza therapy. *J. Antimicrob. Chemother.* **2009**, *64*, 1–4.

20. Newsome, T.P.; Scaplehorn, N.; Way, M. SRC mediates a switch from microtubule- to actin-based motility of vaccinia virus. *Science* **2004**, *306*, 124–129.

21. Kujime, K.; Hashimoto, S.; Gon, Y.; Shimizu, K.; Horie, T. p38 mitogen-activated protein kinase and c-jun-NH2-terminal kinase regulate RANTES production by influenza virus-infected human bronchial epithelial cells. *J. Immunol.* **2000**, *164*, 3222–3228.

22. Yoon, J.J.; Chawla, D.; Paal, T.; Ndungu, M.; Du, Y.; Kurtkaya, S.; Sun, A.; Snyder, J.P.; Plemper, R.K. High-throughput screening-based identification of paramyxovirus inhibitors. *J. Biomol. Screen.* **2008**, *13*, 591–608.

23. Bushell, K.M.; Sollner, C.; Schuster-Boeckler, B.; Bateman, A.; Wright, G.J. Large-scale screening for novel low-affinity extracellular protein interactions. *Genome Res.* **2008**, *18*, 622–630.

24. Fields, S.; Sternglanz, R. The two-hybrid system: an assay for protein-protein interactions. *Trends Genet.* **1994**, *10*, 286–292.

25. Chien, C.T.; Bartel, P.L.; Sternglanz, R.; Fields, S. The two-hybrid system: A method to identify and clone genes for proteins that interact with a protein of interest. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 9578–9582.

26. Kaur, G.; Roy, I. Therapeutic applications of aptamers. *Expert Opin. Invest. Drugs* **2008**, *17*, 43–60.

27. Borghouts, C.; Kunz, C.; Groner, B. Peptide aptamer libraries. *Comb. Chem. High Throughput Screen.* **2008**, *11*, 135–145.

28. Grimm, D.; Kay, M.A. Therapeutic application of RNAi: Is mRNA targeting finally ready for prime time? *J. Clin. Invest.* **2007**, *117*, 3633–3641.

29. Fewell, G.D.; Schmitt, K. Vector-based RNAi approaches for stable, inducible and genome-wide screens. *Drug Discov. Today* **2006**, *11*, 975–982.

30. Filipowicz, W.; Bhattacharyya, S.N.; Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: Are the answers in sight? *Nat. Rev. Genet.* **2008**, *9*, 102–114.

31. Stenvang, J.; Kauppinen, S. MicroRNAs as targets for antisense-based therapeutics. *Expert Opin. Biol. Ther.* **2008**, *8*, 59–81.

32. Nguyen, D.G.; Yin, H.; Zhou, Y.; Wolff, K.C.; Kuhen, K.L.; Caldwell, J.S. Identification of novel therapeutic targets for HIV infection through functional genomic cDNA screening. *Virology* **2007**, *362*, 16–25.

33. Basha, S.; Rai, P.; Poon, V.; Saraph, A.; Gujraty, K.; Go, M.Y.; Sadacharan, S.; Frost, M.; Mogridge, J.; Kane, R.S. Polyvalent inhibitors of anthrax toxin that target host receptors. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 13509–13513.

34. Konig, R.; Stertz, S.; Zhou, Y.; Inoue, A.; Hoffmann, H.H.; Bhattacharyya, S.; Alamares, J.G.; Tscherne, D.M.; Ortigoza, M.B.; Liang, Y.; *et al*. Human host factors required for influenza virus replication. *Nature* **2010**, *463*, 813–817.

35. Karlas, A.; Machuy, N.; Shin, Y.; Pleissner, K.P.; Artarini, A.; Heuer, D.; Becker, D.; Khalil, H.; Ogilvie, L.A.; Hess, S.; *et al*. Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. *Nature* **2010**, *463*, 818–822.

36. Brass, A.L.; Dykxhoorn, D.M.; Benita, Y.; Yan, N.; Engelman, A.; Xavier, R.J.; Lieberman, J.; Elledge, S.J. Identification of host proteins required for HIV infection through a functional genomic screen. *Science* **2008**, *319*, 921–926.

37. Zhou, H.; Xu, M.; Huang, Q.; Gates, A.T.; Zhang, X.D.; Castle, J.C.; Stec, E.; Ferrer, M.; Strulovici, B.; Hazuda, D.J.; Espeseth, A.S. Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe* **2008**, *4*, 495–504.

38. Borner, K.; Hermle, J.; Sommer, C.; Brown, N.P.; Knapp, B.; Glass, B.; Kunkel, J.; Torralba, G.; Reymann, J.; Beil, N.; *et al*. From experimental setup to bioinformatics: An RNAi screening platform to identify host factors involved in HIV-1 replication. *Biotechnol. J.* **2010**, *5*, 39–49.

39. Pache, L.; Konig, R.; Chanda, S.K. Identifying HIV-1 host cell factors by genome-scale RNAi screening. *Methods* **2011**, *53*, 3–12.

40. Krishnan, M.N.; Ng, A.; Sukumaran, B.; Gilfoy, F.D.; Uchil, P.D.; Sultana, H.; Brass, A.L.; Adametz, R.; Tsui, M.; Qian, F.; *et al*. RNA interference screen for human genes associated with West Nile virus infection. *Nature* **2008**, *455*, 242–245.

41. Li, Q.; Brass, A.L.; Ng, A.; Hu, Z.; Xavier, R.J.; Liang, T.J.; Elledge, S.J. A genome-wide genetic screen for host factors required for hepatitis C virus propagation. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 16410–16415.

42. Sigoillot, F.D.; King, R.W. Vigilance and Validation: Keys to Success in RNAi Screening. *ACS Chem. Biol.* **2011**, *6*, 47–60.

43. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **2009**, *37*, 1–13.

44. Bard, J.B.; Rhee, S.Y. Ontologies in biology: design, applications and future challenges. *Nat. Rev. Genet.* **2004**, *5*, 213–222.

45. Rhee, S.Y.; Wood, V.; Dolinski, K.; Draghici, S. Use and misuse of the gene ontology annotations. *Nat. Rev. Genet.* **2008**, *9*, 509–515.

46. Chuang, H.Y.; Hofree, M.; Ideker, T. A decade of systems biology. *Annu. Rev. Cell Dev. Biol.* **2010**, *26*, 721–744.

47. Jensen, L.J.; Bork, P. Ontologies in quantitative biology: a basis for comparison, integration, and discovery. *PLoS Biol.* **2010**, *8*, e1000374.

48. Khatri, P.; Draghici, S. Ontological analysis of gene expression data: Current tools, limitations, and open problems. *Bioinformatics* **2005**, *21*, 3587–3595.

49. Ooi, H.S.; Schneider, G.; Chan, Y.L.; Lim, T.T.; Eisenhaber, B.; Eisenhaber, F. Databases of protein-protein interactions and complexes. *Methods Mol. Biol.* **2010**, *609*, 145–159.

50. Ooi, H.S.; Schneider, G.; Lim, T.T.; Chan, Y.L.; Eisenhaber, B.; Eisenhaber, F. Biomolecular pathway databases. *Methods Mol. Biol.* **2010**, *609*, 129–144.

51. Quackenbush, J. Computational analysis of microarray data. *Nat. Rev. Genet.* **2001**, *2*, 418–427.

52. Kanehisa, M.; Goto, S.; Furumichi, M.; Tanabe, M.; Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **2010**, *38*, D355–D360.

53. Kanehisa, M.; Goto, S.; Hattori, M.; Aoki-Kinoshita, K.F.; Itoh, M.; Kawashima, S.; Katayama, T.; Araki, M.; Hirakawa, M. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* **2006**, *34*, D354–D357.

54. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30.

55. Matthews, L.; Gopinath, G.; Gillespie, M.; Caudy, M.; Croft, D.; de Bono, B.; Garapati, P.; Hemish, J.; Hermjakob, H.; Jassal, B.; *et al*. Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* **2009**, *37*, D619–D622.

56. Vastrik, I.; D'Eustachio, P.; Schmidt, E.; Gopinath, G.; Croft, D.; de Bono, B.; Gillespie, M.; Jassal, B.; Lewis, S.; Matthews, L.; *et al*. Reactome: A knowledge base of biologic pathways and processes. *Genome Biol.* **2007**, *8*, R39.

57. Joshi-Tope, G.; Vastrik, I.; Gopinath, G.R.; Matthews, L.; Schmidt, E.; Gillespie, M.; D'Eustachio, P.; Jassal, B.; Lewis, S.; Wu, G.; *et al*. The Genome Knowledgebase: A resource for biologists and bioinformaticists. *Cold Spring Harb. Symp. Quant. Biol.* **2003**, *68*, 237–243.

58. Thomas, P.D.; Campbell, M.J.; Kejariwal, A.; Mi, H.; Karlak, B.; Daverman, R.; Diemer, K.; Muruganujan, A.; Narechania, A. PANTHER: A library of protein families and subfamilies indexed by function. *Genome Res.* **2003**, *13*, 2129–2141.

59. Thomas, P.D.; Kejariwal, A.; Campbell, M.J.; Mi, H.; Diemer, K.; Guo, N.; Ladunga, I.; Ulitsky-Lazareva, B.; Muruganujan, A.; Rabkin, S.; *et al*. PANTHER: A browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res.* **2003**, *31*, 334–341.

60. Pico, A.R.; Kelder, T.; van Iersel, M.P.; Hanspers, K.; Conklin, B.R.; Evelo, C. WikiPathways: Pathway editing for the people. *PLoS Biol.* **2008**, *6*, e184.

61. Karp, P.D.; Ouzounis, C.A.; Moore-Kochlacs, C.; Goldovsky, L.; Kaipa, P.; Ahren, D.; Tsoka, S.; Darzentas, N.; Kunin, V.; Lopez-Bigas, N., Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res.* **2005**, *33*, 6083–6089.

62. Lee, S.J.; Ways, J.A.; Barbato, J.C.; Essig, D.; Pettee, K.; DeRaedt, S.J.; Yang, S.; Weaver, D.A.; Koch, L.G.; Cicila, G.T. Gene expression profiling of the left ventricles in a rat model of intrinsic aerobic running capacity. *Physiol. Genomics* **2005**, *23*, 62–71.

63. Gene Ontology (GO). Available online: http://www.geneontology.org/ (accessed on 9 June 2011).

64. Szklarczyk, D.; Franceschini, A.; Kuhn, M.; Simonovic, M.; Roth, A.; Minguez, P.; Doerks, T.; Stark, M.; Muller, J.; Bork, P.; *et al*. The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **2011**, *39*, D561–D568.

65. Jensen, L.J.; Kuhn, M.; Stark, M.; Chaffron, S.; Creevey, C.; Muller, J.; Doerks, T.; Julien, P.; Roth, A.; Simonovic, M.; *et al*. STRING 8—A global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* **2009**, *37*, D412–D416.

66. von Mering, C.; Jensen, L.J.; Kuhn, M.; Chaffron, S.; Doerks, T.; Kruger, B.; Snel, B.; Bork, P. STRING 7—recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res.* **2007**, *35*, D358–D362.

67. von Mering, C.; Jensen, L.J.; Snel, B.; Hooper, S.D.; Krupp, M.; Foglierini, M.; Jouffre, N.; Huynen, M.A.; Bork, P. STRING: Known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* **2005**, *33*, D433–D437.

68. von Mering, C.; Huynen, M.; Jaeggi, D.; Schmidt, S.; Bork, P.; Snel, B. STRING: A database of predicted functional associations between proteins. *Nucleic Acids Res.* **2003**, *31*, 258–261.

69. Snel, B.; Lehmann, G.; Bork, P.; Huynen, M.A. STRING: A web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.* **2000**, *28*, 3442–3444.

70. Kuhn, M.; Szklarczyk, D.; Franceschini, A.; Campillos, M.; von Mering, C.; Jensen, L.J.; Beyer, A.; Bork, P. STITCH 2: An interaction network database for small molecules and proteins. *Nucleic Acids Res.* **2010**, *38*, D552–D556.

71. Kuhn, M.; von Mering, C.; Campillos, M.; Jensen, L.J.; Bork, P. STITCH: Interaction networks of chemicals and proteins. *Nucleic Acids Res.* **2008**, *36*, D684–688.

72. Dezso, Z.; Nikolsky, Y.; Nikolskaya, T.; Miller, J.; Cherba, D.; Webb, C.; Bugrim, A. Identifying disease-specific genes based on their topological significance in protein networks. *BMC Syst. Biol.* **2009**, *3*,DOI: 10.1186/1752-0509-3-36.

73. LaCount, D.J.; Vignali, M.; Chettier, R.; Phansalkar, A.; Bell, R.; Hesselberth, J.R.; Schoenfeld, L.W.; Ota, I.; Sahasrabudhe, S.; Kurschner, C.; *et al*. A protein interaction network of the malaria parasite Plasmodium falciparum. *Nature* **2005**, *438*, 103–107.

74. Bader, G.D.; Hogue, C.W. BIND—a data specification for storing and describing biomolecular interactions, molecular complexes and pathways. *Bioinformatics* **2000**, *16*, 465–477.

75. Bader, G.D.; Donaldson, I.; Wolting, C.; Ouellette, B.F.; Pawson, T.; Hogue, C.W. BIND—The Biomolecular Interaction Network Database. *Nucleic Acids Res.* **2001**, *29*, 242–245.

76. Ceol, A.; Chatr Aryamontri, A.; Licata, L.; Peluso, D.; Briganti, L.; Perfetto, L.; Castagnoli, L.; Cesareni, G. MINT, the molecular interaction database: 2009 update. *Nucleic Acids Res.* **2010**, *38*, D532–D539.

77. Prasad, T.S.; Kandasamy, K.; Pandey, A. Human protein reference database and human proteinpedia as discovery tools for systems biology. *Methods Mol. Biol.* **2009**, *577*, 67–79.

78. Mishra, G.R.; Suresh, M.; Kumaran, K.; Kannabiran, N.; Suresh, S.; Bala, P.; Shivakumar, K.; Anuradha, N.; Reddy, R.; Raghavan, T.M.; *et al*. Human protein reference database—2006 update. *Nucleic Acids Res* **2006,** *34*, D411–D414.

79. Peri, S.; Navarro, J.D.; Amanchy, R.; Kristiansen, T.Z.; Jonnalagadda, C.K.; Surendranath, V.; Niranjan, V.; Muthusamy, B.; Gandhi, T.K.; Gronborg, M.; *et al*. Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res.* **2003**, *13*, 2363–2371.

80. Diez, D.; Wheelock, A.M.; Goto, S.; Haeggstrom, J.Z.; Paulsson-Berne, G.; Hansson, G.K.; Hedin, U.; Gabrielsen, A.; Wheelock, C.E. The use of network analyses for elucidating mechanisms in cardiovascular disease. *Mol. Biosyst.* **2010**, *6*, 289–304.

81. Tarca, A.L.; Draghici, S.; Khatri, P.; Hassan, S.S.; Mittal, P.; Kim, J.S.; Kim, C.J.; Kusanovic, J.P.; Romero, R. A novel signaling pathway impact analysis. *Bioinformatics* **2009**, *25*, 75–82.

82. Qu, X.A.; Gudivada, R.C.; Jegga, A.G.; Neumann, E.K.; Aronow, B.J. Inferring novel disease indications for known drugs by semantically linking drug action and disease mechanism relationships. *BMC Bioinformatics* **2009**, *10*, DOI: 10.1186/1471-2105-10-S5-S4.

83. Sthoeger, Z.M.; Zinger, H.; Mozes, E. Beneficial effects of the anti-oestrogen tamoxifen on systemic lupus erythematosus of (NZBxNZW)F1 female mice are associated with specific reduction of IgG3 autoantibodies. *Ann. Rheum. Dis.* **2003**, *62*, 341–346.

84. Cohen-Solal, J.F.; Jeganathan, V.; Grimaldi, C.M.; Peeva, E.; Diamond, B. Sex hormones and SLE: influencing the fate of autoreactive B cells. *Curr. Top. Microbiol. Immunol.* **2006**, *305*, 67–88.

85. Wishart, D.S.; Knox, C.; Guo, A.C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. DrugBank: A knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **2008**, *36*, D901–D906.

86. Wishart, D.S.; Knox, C.; Guo, A.C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **2006**, *34*, D668–D672.

87. Kellenberger, E.; Muller, P.; Schalon, C.; Bret, G.; Foata, N.; Rognan, D. sc-PDB: An annotated database of druggable binding sites from the Protein Data Bank. *J. Chem. Inf. Model.* **2006**, *46*, 717–727.

88. Berman, H.M.; Bhat, T.N.; Bourne, P.E.; Feng, Z.; Gilliland, G.; Weissig, H.; Westbrook, J. The Protein Data Bank and the challenge of structural genomics. *Nat. Struct. Biol.* **2000**, *7*, 957–959.

89. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

90. Li, Q.; Cheng, T.; Wang, Y.; Bryant, S.H. PubChem as a public resource for drug discovery. *Drug Discov. Today* **2010**, *15*, 1052–1057.

91. Olah, M.; Rad, R.; Ostopovici, L.; Bora, A.; Hadaruga, N.; Hadaruga, R.; Moldovan, R.; Fulias, A.; Mracec, M.; Opera, T.I. WOMBAT and WOMBAT-PK: Bioactivity Databases for Lead and Drug Discovery. In *Chemical Biology: From Small Molecules to Systems Biology and Drug Design*; Schreiber, S.L., Kapoor, T.M., Wess, G., Eds.; Wiley-VCH: New York, NY, USA, 2007; pp. 760–786.

92. Opera, T.I.; Benedetti, P.; Berellini, G.; Olah, M.; Fejgin, K.; Boyer, S. Rapid ADME Filters for Lead Discovery In *Molecular Interactions Field*; Cruciani, G., Ed.; Wiley-VCH: New York, NY, USA, 2006; pp. 249–272.

93. Scheer, M.; Grote, A.; Chang, A.; Schomburg, I.; Munaretto, C.; Rother, M.; Sohngen, C.; Stelzer, M.; Thiele, J.; Schomburg, D. BRENDA, the enzyme information system in 2011. *Nucleic Acids Res.* **2011**, *39*, D670–D676.

94. Shoichet, B.K. Virtual screening of chemical libraries. *Nature* **2004**, *432*, 862–865.

95. Konig, R.; Zhou, Y.Y.; Elleder, D.; Diamond, T.L.; Bonamy, G.M.C.; Irelan, J.T.; Chiang, C.Y.; Tu, B.P.; De Jesus, P.D.; Lilley, C.E.; *et al*. Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell* **2008**, *135*, 49–60.

96. Bushman, F.D.; Malani, N.; Fernandes, J.; D'Orso, I.; Cagney, G.; Diamond, T.L.; Zhou, H.L.; Hazuda, D.J.; Espeseth, A.S.; Konig, R.; *et al*. Host Cell Factors in HIV Replication: Meta-Analysis of Genome-Wide Studies. *PLoS Pathog.* **2009**, *5*, e1000437.

97. Barrows, N.J.; Le Sommer, C.; Garcia-Blanco, M.A.; Pearson, J.L. Factors affecting reproducibility between genome-scale siRNA-based screens. *J. Biomol. Screen.* **2010**, *15*, 735–747.

98. Gene Ontology Consortium.The Gene Ontology in 2010: Extensions and refinements. *Nucleic Acids Res.* **2010**, *38*, D331–D335.

99. The Reference Genome Group of the Gene Ontology Consortium. The Gene Ontology's Reference Genome Project: A unified framework for functional annotation across species. *PLoS Comput. Biol.* **2009**, *5*, e1000431.

100. Dennis, G., Jr.; Sherman, B.T.; Hosack, D.A.; Yang, J.; Gao, W.; Lane, H.C.; Lempicki, R.A. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **2003**, *4*, R60:1–R60:11.

101. Fu, W.; Sanders-Beer, B.E.; Katz, K.S.; Maglott, D.R.; Pruitt, K.D.; Ptak, R.G. Human immunodeficiency virus type 1, human protein interaction database at NCBI. *Nucleic Acids Res.* **2009**, *37*, D417–D422.

102. Garrus, J.E.; von Schwedler, U.K.; Pornillos, O.W.; Morham, S.G.; Zavitz, K.H.; Wang, H.E.; Wettstein, D.A.; Stray, K.M.; Cote, M.; Rich, R.L.; *et al*. Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* **2001**, *107*, 55–65.

103. Kilzer, J.M.; Stracker, T.; Beitzel, B.; Meek, K.; Weitzman, M.; Bushman, F.D. Roles of host cell factors in circularization of retroviral dna. *Virology* **2003**, *314*, 460–467.

104. Wei, P.; Garber, M.E.; Fang, S.M.; Fischer, W.H.; Jones, K.A. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* **1998**, *92*, 451–462.

105. Yedavalli, V.S.; Neuveut, C.; Chi, Y.H.; Kleiman, L.; Jeang, K.T. Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function. *Cell* **2004**, *119*, 381–392.

106. Fellay, J.; Shianna, K.V.; Ge, D.; Colombo, S.; Ledergerber, B.; Weale, M.; Zhang, K.; Gumbs, C.; Castagna, A.; Cossarizza, A.; *et al*. A whole-genome association study of major determinants for host control of HIV-1. *Science* **2007**, *317*, 944–947.

107. Matthews, P.C.; Prendergast, A.; Leslie, A.; Crawford, H.; Payne, R.; Rousseau, C.; Rolland, M.; Honeyborne, I.; Carlson, J.; Kadie, C.; *et al*. Central role of reverting mutations in HLA associations with human immunodeficiency virus set point. *J. Virol.* **2008**, *82*, 8548–8559.

108. Cherepanov, P. LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro. *Nucleic Acids Res.* **2007**, *35*, 113–124.

109. Cherepanov, P.; Maertens, G.; Proost, P.; Devreese, B.; Van Beeumen, J.; Engelborghs, Y.; De Clercq, E.; Debyser, Z. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J. Biol. Chem.* **2003**, *278*, 372–381.

110. Ciuffi, A.; Bushman, F.D. Retroviral DNA integration: HIV and the role of LEDGF/p75. *Trends Genet.* **2006**, *22*, 388–395.

111. Gamble, T.R.; Vajdos, F.F.; Yoo, S.; Worthylake, D.K.; Houseweart, M.; Sundquist, W.I.; Hill, C.P. Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell* **1996**, *87*, 1285–1294.

112. Llano, M.; Saenz, D.T.; Meehan, A.; Wongthida, P.; Peretz, M.; Walker, W.H.; Teo, W.; Poeschla, E.M. An essential role for LEDGF/p75 in HIV integration. *Science* **2006**, *314*, 461–464.

113. Maertens, G.; Cherepanov, P.; Pluymers, W.; Busschots, K.; De Clercq, E.; Debyser, Z.; Engelborghs, Y. LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *J. Biol. Chem.* **2003**, *278*, 33528–33539.

114. Marshall, H.M.; Ronen, K.; Berry, C.; Llano, M.; Sutherland, H.; Saenz, D.; Bickmore, W.; Poeschla, E.; Bushman, F.D. Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. *PLoS One* **2007**, *2*, e1340.

115. Turlure, F.; Devroe, E.; Silver, P.A.; Engelman, A. Human cell proteins and human immunodeficiency virus DNA integration. *Front. Biosci.* **2004**, *9*, 3187–3208.

116. Ciuffi, A.; Llano, M.; Poeschla, E.; Hoffmann, C.; Leipzig, J.; Shinn, P.; Ecker, J.R.; Bushman, F. A role for LEDGF/p75 in targeting HIV DNA integration. *Nat. Med.* **2005**, *11*, 1287–1289.

117. Llano, M.; Vanegas, M.; Fregoso, O.; Saenz, D.; Chung, S.; Peretz, M.; Poeschla, E.M. LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. *J. Virol.* **2004**, *78*, 9524–9537.

118. Hao, L.; Sakurai, A.; Watanabe, T.; Sorensen, E.; Nidom, C.A.; Newton, M.A.; Ahlquist, P.; Kawaoka, Y. Drosophila RNAi screen identifies host genes important for influenza virus replication. *Nature* **2008**, *454*, 890–893.

119. Brass, A.L.; Huang, I.C.; Benita, Y.; John, S.P.; Krishnan, M.N.; Feeley, E.M.; Ryan, B.J.; Weyer, J.L.; van der Weyden, L.; Fikrig, E.; *et al*. The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. *Cell* **2009**, *139*, 1243–1254.

120. Shapira, S.D.; Gat-Viks, I.; Shum, B.O.; Dricot, A.; de Grace, M.M.; Wu, L.; Gupta, P.B.; Hao, T.; Silver, S.J.; Root, D.E.; *et al*. A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. *Cell* **2009**, *139*, 1255–1267.

121. Josset, L.; Textoris, J.; Loriod, B.; Ferraris, O.; Moules, V.; Lina, B.; N'Guyen, C.; Diaz, J.J.; Rosa-Calatrava, M. Gene expression signature-based screening identifies new broadly effective influenza a antivirals. *PLoS One* **2010**, *5*, e13169.

122. Coombs, K.M.; Berard, A.; Xu, W.; Krokhin, O.; Meng, X.; Cortens, J.P.; Kobasa, D.; Wilkins, J.; Brown, E.G. Quantitative proteomic analyses of influenza virus-infected cultured human lung cells. *J. Virol.* **2010**, *84*, 10888–10906.

123. Watanabe, T.; Watanabe, S.; Kawaoka, Y. Cellular networks involved in the influenza virus life cycle. *Cell Host Microbe* **2010**, *7*, 427–439.

124. Min, J.Y.; Subbarao, K. Cellular targets for influenza drugs. *Nat. Biotechnol.* **2010**, *28*, 239–240.

125. Chase, G.; Deng, T.; Fodor, E.; Leung, B.W.; Mayer, D.; Schwemmle, M.; Brownlee, G. Hsp90 inhibitors reduce influenza virus replication in cell culture. *Virology* **2008**, *377*, 431–439.

126. Gonzalez, O.; Fontanes, V.; Raychaudhuri, S.; Loo, R.; Loo, J.; Arumugaswami, V.; Sun, R.; Dasgupta, A.; French, S.W. The heat shock protein inhibitor Quercetin attenuates hepatitis C virus production. *Hepatology* **2009**, *50*, 1756–1764.

127. Soh, D.; Dong, D.; Guo, Y.; Wong, L. Consistency, comprehensiveness, and compatibility of pathway databases. *BMC Bioinformatics* **2010**, *11*, 449.F

128. Ingenuity Systems. Available online: http://www.ingenuity.com/index.html (accessed on 6 January 2011).

129. Konig, R.; Stertz, S.; Zhou, Y.; Inoue, A.; Hoffmann, H.H.; Bhattacharyya, S.; Alamares, J. G.; Tscherne, D. M.; Ortigoza, M. B.; Liang, Y.; Gao, Q.; Andrews, S.E.; Bandyopadhyay, S.; De Jesus, P.; Tu, B.P.; Pache, L.; Shih, C.; Orth, A.; Bonamy, G.; Miraglia, L.; Ideker, T.; Garcia-Sastre, A.; Young, J.A.; Palese, P.; Shaw, M.L.; Chanda, S.K. Human host factors required for influenza virus replication. *Nature* **2010**, *463*, 813–817.

130. Karlas, A.; Machuy, N.; Shin, Y.; Pleissner, K. P.; Artarini, A.; Heuer, D.; Becker, D.; Khalil, H.; Ogilvie, L.A.; Hess, S.; Maurer, A.P.; Muller, E.; Wolff, T.; Rudel, T.; Meyer, T.F. Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. *Nature* **2010**, *463*, 818–822.

131. Brass, A.L.; Huang, I.C.; Benita, Y.; John, S.P.; Krishnan, M.N.; Feeley, E.M.; Ryan, B.J.; Weyer, J.L.; van der Weyden, L.; Fikrig, E.; Adams, D.J.; Xavier, R.J.; Farzan, M.; Elledge, S.J. The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. *Cell* **2009**, *139*, 1243–1254.

132. Shapira, S.D.; Gat-Viks, I.; Shum, B.O.; Dricot, A.; de Grace, M.M.; Wu, L.; Gupta, P.B.; Hao, T.; Silver, S. J.; Root, D. E.; Hill, D. E.; Regev, A.; Hacohen, N., A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. *Cell* **2009**, *139*, 1255–1267.

133. Josset, L.; Textoris, J.; Loriod, B.; Ferraris, O.; Moules, V.; Lina, B.; N'Guyen, C.; Diaz, J.J.; Rosa-Calatrava, M., Gene expression signature-based screening identifies new broadly effective influenza a antivirals. *PLoS One* **2010**, *5*.

134. Chen, Y.L.; Chen, W.W.; Wang, Y.F.; Li, R.L.; Guo, W.F.; Lao, S.X.; Wang, J.H.; Huang, S.P. Bioinformatics research on chronic superficial gastritis of Pi-deficiency syndrome by gene arrays. *Chin. J. Integr. Med.* **2009**, *15*, 341–346.