ELSEVIER

Computational biology approaches for selecting host–pathogen drug targets

James R. Brown[1], Michal Magid-Slav[2], Philippe Sanseau[3] and Deepak K. Rajpal[4]

[1] Computational Biology, GlaxoSmithKline, 1250 South Collegeville Road, UP1345, PO Box 5089, Collegeville, PA 19426-0989, USA
[2] 709 Swedeland Road, UW2230, King of Prussia, PA 19406, USA
[3] Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, UK
[4] 5 Moore Drive, 3.2084, Research Triangle Park, NC 27709, USA

The proliferation of genomic platform data, ranging from silencing RNAs through mRNA and microRNA expression to proteomics, is providing new insights into the interplay between human and pathogen genes during infection: the so-called 'host–pathogen interactome'. Exploiting the interactome for novel human drug targets could provide new therapeutic avenues towards the treatment of infectious disease, which could ameliorate the growing clinical challenge of drug-resistant infections. Using the hepatitis C virus interactome as an example, here we suggest a computational biology framework for identifying and prioritizing potential human host targets against infectious diseases.

Introduction

Despite decades of drug research and development, infectious diseases are still the top global healthcare problem, being responsible for millions of morbidities and mortalities each year. Most drugs and vaccines usually target some essential biological function of the pathogen itself. Although initially effective, pathogen gene-targeted treatments exert Darwinian selection pressures on the infectious species, leading to the emergence of drug-resistant strains. Increasing incidents of drug-resistant *Plasmodium falciparum*, the malaria parasite [1] and methicillin-resistant *Staphylococcus aureus* (MRSA) [2] are just two examples of this phenomenon. For other pathogens, existing therapies have either suboptimal outcomes or variable efficacies in different patients.

Pathogen invasion can be broadly considered to be dependent on three components: (i) specific virulence factors encoded by the genome of the pathogen; (ii) the effectiveness of the host immune system response (and the ability of the pathogen to evade it); and (iii) the success of the pathogen to co-opt host factors for its propagation. Our knowledge of these three processes is increasing through the application of genomic technologies. The genomes of most major pathogens have now been sequenced and, for others, such as HIV [3], there are literally dozens, if not hundreds, of examples from different clinical isolates. New DNA sequencing technology will no doubt result in the availability of thousands of pathogen genomes within the next few years. Considerable advances have also been made in understanding the function of innate immunity at the genome-level as the primary trigger for host defense against infection [4]. The host–pathogen interactome is the most recent focus of genomic technologies. Genome-wide studies of host interactions using global RNA interference (RNAi), proteomics, mRNA and microRNA (miRNA) transcriptomics, and other platforms, are beginning to reveal important insights into the human proteins and pathways that are essential for pathogen infection, proliferation and persistence. Recent genome-wide infectivity studies reveal extensive dependency on the co-opting of host systems by viruses, such as HIV [5–7], hepatitis C virus (HCV) [8] and influenza H1N1 [9,10], as well as bacteria, such as *Mycobacterium tuberculosis*, the causative agent of TB [11].

Besides increasing our basic knowledge of infection, these studies defining host–pathogen interactomes could lead to novel infectious disease therapeutic regimes based on the modulation of human proteins. The rationale here is simple and logical. Drug-resistant viruses, bacteria and protozoan parasites (e.g. malaria) spread through clinical settings by a combination of higher mutation rates and selection-driven amplification of resistant strain populations. By contrast, human host-factor proteins are more evolutionary stable, and will not mutate into drug refractory variants. The dependency of the pathogen on a suite of host

Corresponding author:. Rajpal, D.K. (Deepak.K.Rajpal@gsk.com)

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2011.01.008

pathways is essential, provided that new strains do not somehow evolve new routes of infectivity or proliferation. Such an adaptation by the pathogen seems less probable, given the complexity of the host–pathogen interactome. Additionally, host target therapies could improve the outcomes for patients suffering from chronic, persistent infections. However, there are crucial considerations in targeting human host factors, in particular, potential toxicity risks and other side effects. Nonetheless, finding modulators of host–pathogen factors and innate immunity proteins is a rapidly emerging area of infectious disease drug discovery [12].

Computational approaches to host–pathogen interactions

The availability of large genome-wide data sets from multiple platforms naturally leads to integrated systems biology approaches requiring extensive computational biology analyses. For example, Shapira *et al.* [10] used bioinformatic approaches based on yeast two-hybrid analysis and genome-wide expression profiling to identify RNA-binding and WNT signaling pathways as being potentially important in influenza viral infections. Using an alternative approach, that of multiple RNAi screens of different H1N1 and swine-originating influenza A strains, Karlas *et al.* [9], found that inhibition of CDC-like kinase 1 (CLK1) impaired viral RNA splicing. Given the extensive pharmacology available surrounding targeted kinase inhibitors, primarily as anticancer agents, these virus–host studies suggest new applications for these small molecules.

However, there are significant challenges in selecting specific drug targets from genome-wide data sets derived from experiments using RNAi or other genome-wide technologies. Primarily, variation in methodologies and analyses between studies makes it difficult to obtain a consensus set of gene targets. Bushman *et al.* [13] performed a meta-analysis of nine genome-wide studies examining host factors involved with HIV replication in cellular assays. The union set of 'significant' hits across all studies consisted of 2410 protein genes or 9.5% of all human genes. Variables between studies that possibly resulted in gene list discrepancies included the number of small interfering RNAs (siRNAs) used per gene, the type of infected host cell line, filtering thresholds, stage of infection (time points) and experimental noise. RNAi screens can also be biased towards false-positives resulting from off-target effects (OTEs) [14]. Whenever possible, it is desirable to have additional target validation, such as repeatable phenotypes from multiple, redundant siRNAs against the same gene or rescue experiments showing that expression of the intact gene counters the RNAi-induced phenotype. Several known cellular host cofactors for HIV replication, such as the integration cofactor Lens epithelium-derived growth factor (LEDGF/p75) and the cell-surface antigens, human leukocyte antigen (HLA)-B57 and HLA-C, have not shown up in siRNA screens, suggesting limits to the inactivation of highly abundant mRNAs [15].

Computational approaches can help with supporting evidence for potential siRNA screening of candidate host factors by linking positive hits to independently derived interactomes [14]. For example, Evans *et al.* [15] identified short amino acid signatures shared between the HIV and human genomes. These eukaryotic linear motifs (ELMs) are potential sites for protein interactions between viral and host proteins. Conservation of ELMs in the HIV reverse transcriptase is partially predictive of patient response to retroviral therapy, further supporting hypotheses on their functional roles [16]. Statistical tests for the enrichment of siRNA hits in independent interaction protein data sets, such as the infection mapping project (I-MAP) [17], can further refine interaction networks [8]. Transcriptomic and proteomic data sets from clinical studies could also be leveraged for indirect validation of siRNA screen hits. However, multiple patient-related factors, such as age, gender, medications, duration of infection, ethnicity and environmental factors, further introduce complexities during comparative analyses.

These caveats aside, well-structured meta-analysis studies have the potential to provide novel insights into the host–pathogen interactome and generate testable hypotheses regarding possible drug targets. The most valuable computational approaches for illuminating potential drug targets are those based on underlying, measurable criteria for target prioritization. Arguably, priority should be given to those host–pathogen targets that fulfill the following criteria: (i) either essential for pathogen infectivity and proliferation or enhance host response to control and clear pathogens; (ii) low redundancy for the pathogen pathways, yet high pathway redundancy for the host, thus potentially mitigating toxicity effects; (iii) widely and consistently expressed in tissues of infection as well as different patient populations; and (iv) therapeutically tractable using small molecules, siRNA, antibodies or other pharmacological modalities. For the remainder of this review, we provide an illustrative application of this framework for the prioritization of potential host factor targets as a strategy against HCV infection.

Meta-analysis approaches for HCV host factors

Globally, HCV infects more than 170 million people and is a leading cause of serious chronic liver diseases, including liver cirrhosis, liver failure and hepatocellular carcinoma [18]. Current treatment consists of peginterferon α and ribavirin, but only approximately half of patients have a sustained response (the absence of detectable virus RNA 24 weeks post-treatment) [19]. HCV is a single positive-strand RNA virus belonging to the Flaviviridae family; it has a 9.6-kb genome and is entirely dependent on its host cell for its replication and proliferation [20]. Modulation of host cell targets could present a higher barrier to viral propagation and might also offer opportunities to block the HCV lifecycle at multiple stages. Exploiting the HCV–host cell environment could also lead to new therapeutic strategies that might improve patient response with fewer side effects. As a result, there has been substantial interest in the study of HCV–host interactions at the cellular–molecular level [21–23].

In Fig. 1, we propose a systematic integrative approach for evaluating host targets modulated by HCV, leading to the identification of putative therapeutically tractable HCV host targets. This approach takes advantage of the growing number of publicly available '-omics' data sets and the development of innovative computational data mining and analysis methodologies. Here, we review the opportunities with respect to the components shown in Fig. 1, as well as their caveats. Several useful public resources are also listed in Table 1.

RNAi

RNAi methods have been widely used in cell systems that contain replicating HCV to study phenotypes associated with HCV–host interactions [24]. RNAi refers to the sequence-specific

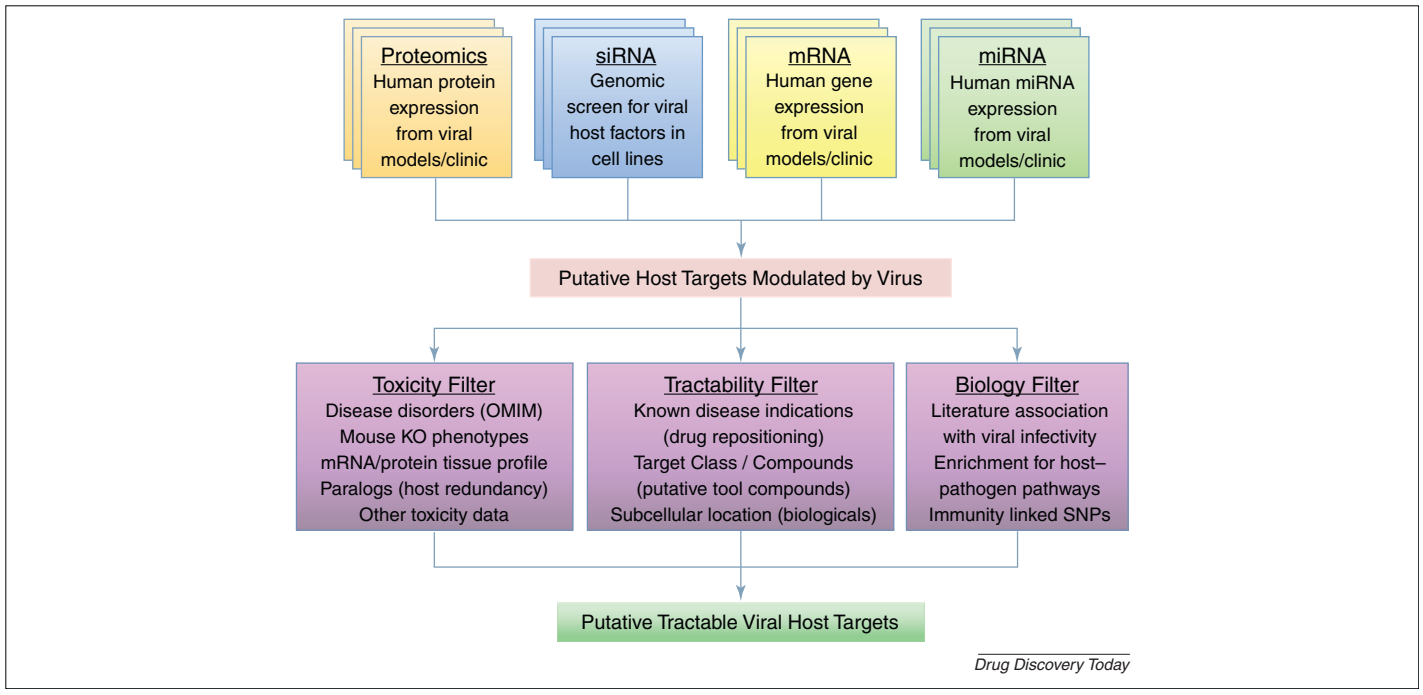

FIGURE 1

A pipeline for selecting putative tractable viral host targets. Conceptually, there are two components. The first is the integrative statistical analysis of multiple platform genomic data sets of the human or animal model host response to a common virus, such as HCV. The second phase involves further refining of putative host targets using quantitative and qualitative filters focused on toxicity, tractability and biological criteria. Specific steps are discussed in the main text.

TABLE 1

Resources

Name	URL	Brief summary
Gene expression		
EBI ArrayExpress	http://www.ebi.ac.uk/microarray-as/ae/	Public collection of microarray gene expression data
GEO (Gene Expression Omnibus)	http://www.ncbi.nlm.nih.gov/geo/	Public gene expression repository
Metabolic pathways, networks and gene signatures		
BioCarta	http://www.biocarta.com/genes/allPathways.asp	Online maps of metabolic and signaling pathways
BioCyc	http://biocyc.org/	A collection of databases integrating genome and metabolic pathway data
GOLD (Genomics Of Lipid-associated Disorders)	https://gold.tugraz.at/	Genes, proteins and pathways implicated in lipid-associated disorders
KEGG (Kyoto Encyclopedia of Genes and Genomes)	http://www.genome.jp/kegg/	Metabolic and regulatory pathways
MetaCyc	http://metacyc.org/	Experimentally studied metabolic pathways from more than 1500 organisms
EndoNet	http://endonet.bioinf.med.uni-goettingen.de/	Information about the components of endocrine networks
Connectivity map	http://www.broadinstitute.org/cmap/	Collection of genome-wide transcriptional expression data from cultured human cells treated with bioactive small molecules
Genetics		
dbSNP	http://www.ncbi.nlm.nih.gov/SNP/	Database of SNPs
MGD (Mouse Genome Database)	http://www.informatics.jax.org/	Mouse genetics, genomics, alleles and phenotypes
Mouse phenome database	http://phenome.jax.org/	Phenotypic and genotypic data from inbred strains of mice
OMIM (Online Mendelian Inheritance in Man)	http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim	Catalog of human genetic and genomic disorders
PhenomicDB	http://www.phenomicdb.de/	A cross-species genotype–phenotype database

TABLE 1 (*Continued*)

Name	URL	Brief summary
Catalog of Published Genome-wide Association Studies	http://genome.gov/gwastudies/	GWAS publications
NIH Genetic Association Database	http://geneticassociationdb.nih.gov/	Archive of human genetic association studies of complex diseases and disorders
HuGE Navigator	http://hugenavigator.net/	Investigates the interactions of genotype and phenotype
NCBI database of genotypes and phenotypes (dbGaP)	http://www.ncbi.nlm.nih.gov/gap	An integrated knowledge base of genetic associations and human genome epidemiology
Metabolome		
HMDB (The Human Metabolome Database)	http://www.hmdb.ca/	Curated human metabolite and metabolism data
Madison-Qingdao Metabolomics Consortium Database	http://mmcd.nmrfam.wisc.edu/	Metabolomics resource based on nuclear magnetic resonance spectroscopy and mass spectrometry
microRNA		
miR2Disease	http://mlg.hit.edu.cn:8080/miR2Disease/	Literature-curated database for validity or potentially pathogenic roles of dysregulated miRNAs in human disease
miRBase	http://www.mirbase.org/	Database of microRNAs (small noncoding RNAs)
Lipid biology		
Lipid MAPS	http://www.lipidmaps.org/	Lipid-associated proteins
Lipase Engineering Database	http://www.led.uni-stuttgart.de/	Integrated information on sequence, structure and function of lipases and esterases
Genes		
Entrez Gene	http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene	Searchable database of genes
UCSC Genome Browser	http://genome.ucsc.edu/	Reference sequence and working draft assemblies for a large collection of genomes
Pharmacogenomic knowledge resources		
PharmGKB	http://www.pharmgkb.org/	Variation in drug response
DrugBank: GenoBrowse	http://www.drugbank.ca/genobrowse	
Proteomics		
Proteomics IDEntifications (PRIDE) Database (EBI)	http://www.ebi.ac.uk/pride/	Molecular/Proteomics
Secretome		
Secreted Protein Database	http://spd.cbi.pku.edu.cn/	Secreted proteins from human, mouse and rat
Clinical Trials		
Clinical Trials.gov	http://www.clinicaltrials.gov/	Registry of federally and privately supported clinical trials conducted in the USA and around the world
Clinical study results	http://www.clinicalstudyresults.org/home/	Repository for clinical study results
Drug interaction		
Drug interaction Database	http://www.druginteractioninfo.org/	
Genomics databases (non-vertebrate)		
European HCV Database	http://euhcvdb.ibcp.fr/euHCVdb/	Towards protein sequence, structure and function analyses and structural biology of HCV
The HVC database	http://hcv.lanl.gov/content/index	Variation in HCV sequences
Viral bioinformatics		

post-transcriptional gene silencing mechanism by double-stranded RNA (dsRNA) using either short hairpin RNA (shRNA) or, more frequently, siRNA libraries. Ribonuclease III cleavage of these longer forms of dsRNA leads to the generation of 21-nucleotide and 22-nucleotide siRNAs, which mediate the sequence-specific RNA degradation [25]. Several recent studies have taken the approach of systematically conducting siRNA screens aimed at identifying host proteins that might support HCV replication: for example, whole-genome siRNA library screens [26]; custom siRNA library screens targeting 140 host membrane-trafficking genes (to identify host genes that might have a role in HCV replication and infectious viral production) [27]; siRNA library screens targeting a panel of human protein kinases [28]; and screens of an siRNA library targeting approximately 4000 human genes that might be

potential targets for drug discovery [29]. These approaches have led to the identification of various host targets that might have an important role in HCV replication. One recent area of intense interest revealed by such siRNA studies is the role of PI4K-α and PI4K-β type III phosphatidylinositol 4-kinases in promoting HCV propagation [27]. Given that phosphatidylinositol kinases are also the targets for several cancer drugs presently in clinical trials, there is the attractive potential to repurpose these small molecule inhibitors for antiviral treatments [30]. Despite differences in protocols, thresholds and reproducibility among various studies, bringing together selected siRNA-based data sources offers a statistical basis to identify certain commonalities of host genes and/or pathways associated with HCV replication.

mRNA expression profiling

Understanding transcriptional responses through the measurement of genome-wide changes in mRNA levels is a powerful approach to study molecular events. As such, the volume of public gene expression microarray data has grown exponentially [31]. At this time, the Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) at the National Centre for Biotechnology Information (NCBI) contains more than 470 000 samples and over 18 000 series. A quick search of GEO using the term 'HCV' identified over 40 relevant data sets. These studies concern host transcriptional responses to HCV infections in cell model systems [32], expression of viral proteins in hepatocytes and other cells [32–35], as well as transcription in rodent [36,37] and chimpanzee animal models [38,39], and patient biopsies [40–42]. Gene expression analyses of HCV-infected cells could reveal the roles that specific genes and pathways have in host cell antiviral responses, HCV replication and liver pathologies associated with HCV infections. However, as for RNAi studies, variability across mRNA expression data sets with respect to microarray design, cellular systems, number of replicates, and other factors, must be considered.

miRNA expression profiling

miRNAs are endogenously expressed small noncoding RNAs with regulatory roles in various cellular processes, such as proliferation, transcription, translation, homeostasis, genomic integrity and mRNA stability [43]. The small (21–25 nucleotide) miRNAs are assembled into RNA-induced silencing complexes (RISC), which bind to specific, partially complementary recognition sequences in the 3′ untranslated region (UTR) of the mRNA, resulting in the suppression of translation and/or mRNA degradation. Each miRNA is estimated to interact with up to 100 mRNAs and perhaps almost one third of known human genes are thought to be miRNA regulated [44]. Given that the aberrant expression of the miRNAs has been correlated with various diseased states, miRNAs are of interest as potential biomarkers as well as therapeutic targets.

Recent studies have profiled miRNA expression in cell lines infected with HCV. Steuerwald *et al.* determined a subset of anticorrelated miRNAs–mRNA pairs that displayed altered expression in a hepatoma cell line expressing the full-length HCV genotype 1b [45]. Liu *et al.* measured host miRNA and mRNA expression during different stages of HCV infection in a hepatoma cell *in vitro* system and identified at least four miRNAs that were potentially involved in HCV entry, replication and propagation [46]. An integrated computational approach to identify HCV infection-associated miRNA–mRNA regulatory modules took advantage of miRNA and mRNA profiling data from HCV-infected human liver biopsies combined with systematic target prediction strategy [47]. It has also been suggested that interferon systems modulate the expression of cellular miRNAs and are potentially used by cells for host defense [48].

In addition to being important tools for exploratory disease biology studies, miRNAs have also been implicated as potential therapeutic targets for HCV [49]. Human mir-122 is abundant in the liver and has been reported to be specifically recruited by HCV [50–52]. In HCV-infected human liver biopsies, mir-122 expression was inversely correlated with fibrosis, serum liver transaminase levels and patient age [53]. The therapeutic potential of targeting miRNAs was recently illustrated by Landford *et al.*, who used a locked nucleic acid (LNA) antagomir (SPC3649) against mir-122 to produce long-term suppression of HCV viremia in chimpanzees [54]. This mechanism of action is further bolstered by the observation that interferon (IFN)-β also reduces mir-122 gene expression. Several mir-122 RNA inhibitors are now in HCV clinical trials. However, other miRNAs have also been reported to have a potential regulatory role in HCV replication [55]. Although miRNAs are undoubtedly important regulatory mechanisms, comparative analyses across studies have produced few consensus patterns of miRNA-linked and mRNA-linked expression. In addition, computational predictions of miRNA binding sites are imprecise and tend to predict multiple miRNA binding sites densely packed in the 3′ UTRs of targeted mRNAs [56]. Regardless, the functional importance of miRNAs warrants further investigation into their potential as both biomarkers and therapeutic targets.

Proteomics

The proteome refers to all the proteins expressed in a given cell, tissue or organ. Proteomic approaches involve experimental methodologies such as two-dimensional (2D) gel electrophoresis, liquid chromatography, mass spectrometry, amino acid sequencing, and protein microarrays and imaging. These data, combined with complex bioinformatics analyses, are used to identify and map the component proteins in particular biological systems. Although proteomic studies are potentially powerful, there are several complexities that affect their analysis and interpretation, including the extent of post-translational protein modifications, the dynamic range of protein expression and problems in detecting low abundance or membrane-associated proteins [57]. Several studies have produced proteomics data sets that contribute towards understanding HCV infections by focusing on host interactions with specific viral proteins. For example, Kang *et al.* identified cytokeratins and vimentin as host proteins that interact with the HCV core protein that packages the viral RNA genome into a nucleocapsid [58]. Other studies concentrated on overall HCV interactions with specific portions of the host proteome, such as the mitochondrial proteome [59] or subcellular locations known to be crucial for HCV replication, such as membrane lipids [60].

Meta-analysis of host–HCV data sets

Systematically combining data from different studies and '-omic' platforms can provide a better understanding of the various processes the host cells undergo following HCV infection. Given that the various platforms measure different cellular processes (i.e. transcription, translation, etc.), such integration might not

necessarily reveal the same changes in gene regulation [61]. Further inconsistencies might be introduced by different screening methodologies, the degree of false-positives and lack of access to raw data, especially for siRNA studies. However, in a properly designed and controlled meta-analysis, each platform might complimentarily reveal those pathways that are significantly enriched across diverse data sets.

One approach to overcome issues of comparing multiple gene lists is to focus the analysis on hierarchies above the gene level, such as cellular pathway or protein functional categories, using enrichment analyses for ontology terms, including gene ontology (GO) molecular functions, biological processes and canonical pathways. As an example, there has recently been a significant increase in the number of publicly available data sets investigating the effects of HCV infection on host cells. Woodhouse *et al.* [62] studied the effects of infecting HuH 7.5 cells with HCV genotype 2 by measuring changes using proteomics, mRNA microarray and RNA-Seq technologies. Both Li *et al.* [8] and Randall *et al.* [63] used siRNA against human host factors using the same HCV genotype and model human cell system. When comparing these different data sets using only gene identifiers, the overlap appears to be only eight genes across all platforms. However, performing pathway enrichment using GeneGO™ MetaCore™ software reveals a greater overlap at the pathway level (Fig. 2, Table 2). Pathways

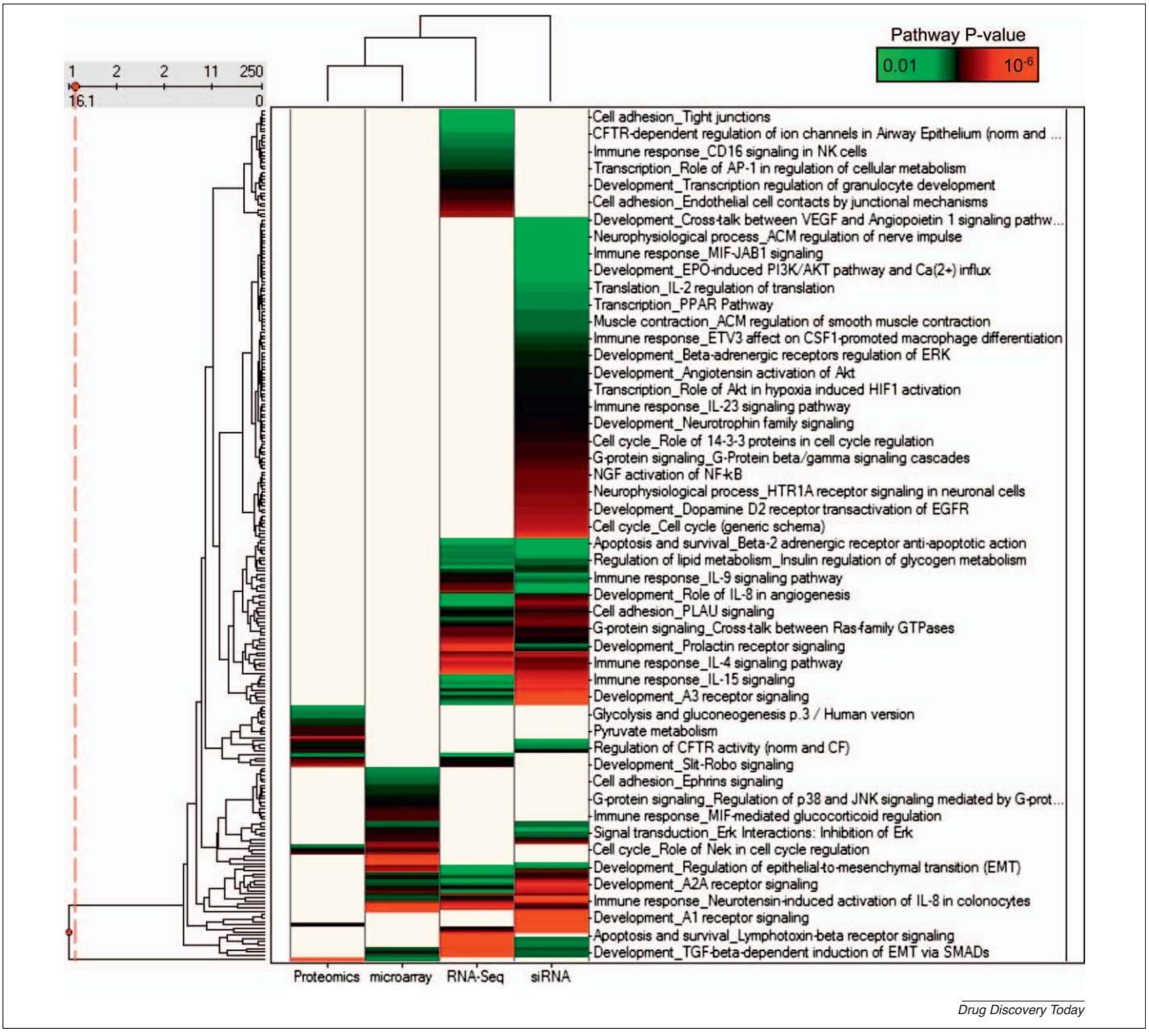

FIGURE 2

An example of integrative analysis of multiple human infectivity data sets for HCV. Hierarchical clustering of $-\log_{10}$(pathway enrichment *P* value) across multiple published data types (proteomics, microarray, RNA-Seq and siRNA) [8,62,63]. The figure shows only pathways that have significant representation ($P \leqslant 0.01$) in at least one platform. Pathways where $P \geqslant 0.01$ are in white. The colors represent the pathway significance level as shown in the legend. The colors represent an increase in significance from green to black to red.

TABLE 2

The 16 pathways that are significantly increased across three out of four data types (RNA-Seq, microarray and siRNA) measuring HCV infectivity

Cell adhesion_Chemokines and adhesion
Cell cycle_ESR1 regulation of G1/S transition
Development_A2A receptor signaling
Development_EGFR signaling pathway
Development_PDGF signaling via STATs and NF-kB
Development_Regulation of epithelial-to-mesenchymal transition (EMT)
Development_TGF-beta receptor signaling
Development_TGF-beta-dependent induction of EMT via MAPK
Development_TGF-beta-dependent induction of EMT via SMADs
DNA damage_ATM/ATR regulation of G1/S checkpoint
Immune response_CD40 signaling
Immune response_Fc epsilon RI pathway
Immune response_Gastrin in inflammatory response
Immune response_Histamine H1 receptor signaling in immune response
Immune response_IFN gamma signaling pathway
Immune response_Neurotensin-induced activation of IL-8 in colonocytes

Proteomics data was enriched for other pathways. These pathways were enriched only in these three platforms.

that are commonly enriched across platforms (mRNA, RNA-Seq and siRNA) and studies include those that are known to be modulated by HCV infection, such as IFN-γ, transforming growth factor (TGF)-β and epithelial to mesenchymal transition (EMT) pathways. Further insights into the adaptive processes of the host and co-opted host factors that are crucial for HCV propagation can be gained by performing pathway enrichment analysis on the combined list of genes essential for HCV replication (siRNA) with host cell responses to HCV infection (mRNA, proteomics and RNA-Seq). This approach takes into account the fact that not all levels of regulation are captured by one platform and can result in the increased significance of some pathways, such as interferon β, nuclear factor (NF)-κB and interleukin (IL)-6.

Further refining the target list

Further refinement of candidate host targets emerging from meta-analyses involves the creative use of various computational tools and databases to address key drug development issues, such as potential toxicity liabilities, druggability and biological rationale (Fig. 1). This subsequent stage is less tractable for the pipelining of computational analyses than the earlier '-omic' platform data phase. However, multiple data sources do exist that can be used to evaluate target risk and infer opportunity. Clues about the potential toxicity liabilities regarding the inactivation of a host protein can be gained from multiple sources of loss-of-function phenotypes (Table 1). Online Mendelian Inheritance in Man (OMIM; http://www.ncbi.nlm.nih.gov/omim) is a publicly available and well curated source of information on known human genotype–phenotypes associations as well as other disease interactions. A query of 'Hepatitis C virus' retrieved over 100 records on human genes associated with HCV pathogenesis. Similarly, novel human targets can be searched for putative genotype–phenotype interactions. Analysis of pathways, as well as gene paralogy, can be used to infer gene redundancy, which might buffer the host from loss-of-function side effects. Many large pharmaceutical companies also have proprietary databases on toxicity from past drug discovery efforts.

The 'druggability' of human proteins by small molecules is an area of significant biomedical and commercial interest. Approved drugs bind to less than 300 human targets [64], although some estimates of the universe of human druggable proteome ranges from 2000 to 3000 proteins [65]. Analysis of target lists for known drug classes, as well as opportunities for repositioning existing drugs, can be used to further refine the list of putative host defense targets. Drugs are also now more broadly defined to include non-small molecules, such as biologicals or antibodies, which can attack extracellular or exposed membrane targets, and siRNAs that, in theory, could be applied to nonconventional targets, including miRNAs.

Finally, additional biological information from the literature and other expert sources can be applied to the validation or invalidation of the proposed target. For genes that are derived from HCV-infected host data sets, literature surveys might reveal additional information on their roles in viral infection, replication and propagation. Further evidence of disease associations might come from pathway analyses as well as from genetics analyses in the form of genome-wide association studies (GWAS).

Pulling together

In this review, we highlighted the opportunities for the field of computational biology to contribute to the identification of potential host defense targets against infectious diseases. There are multiple, rich data sets available for tackling this problem and no doubt such resources will continue to expand rapidly with new technologies, such as next generation sequencing. The specific computational methodologies and data sources used will be highly dependent upon the specific pathogen as well as the therapeutic strategy. Ultimately, the value of any such endeavor will not be solely determined by some grand computational meta-analysis. Rather it will be computational biologists, molecular biologists, disease specialists and clinicians pulling together as interdisciplinary teams to develop new medicines against some of the most devastating human diseases.

Acknowledgments

We wish to thank Heather Madsen and Chris Moore for helpful discussions.

References

1 WHO, (2009) *World Malaria Report*. WHO

2 Rossolini, G.M. *et al.* (2010) Epidemiology and clinical relevance of microbial resistance determinants versus anti-Gram-positive agents. *Curr. Opin. Microbiol.* 13, 582–588

3 Los Alamos National Laboratory, (2010) *HIV Databases*. National Institute of Allergy and Infectious Diseases

4 Medzhitov, R. and Janeway, C., Jr (2000) Innate immunity. *N. Engl. J. Med.* 343, 338–344

5 Gautier, V.W. *et al.* (2009) *In vitro* nuclear interactome of the HIV-1 Tat protein. *Retrovirology* 6, 47
6 Konig, R. *et al.* (2008) Global analysis of host–pathogen interactions that regulate early-stage HIV-1 replication. *Cell* 135, 49–60
7 Rotger, M. *et al.* (2010) Genome-wide mRNA expression correlates of viral control in CD4+ T-cells from HIV-1-infected individuals. *PLoS Pathog.* 6, e1000781
8 Li, Q. *et al.* (2009) A genome-wide genetic screen for host factors required for hepatitis C virus propagation. *Proc. Natl. Acad. Sci. U. S. A.* 106, 16410–16415
9 Karlas, A. *et al.* (2010) Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. *Nature* 463, 818–822
10 Shapira, S.D. *et al.* (2009) A physical and regulatory map of host–influenza interactions reveals pathways in H1N1 infection. *Cell* 139, 1255–1267
11 Kumar, D. *et al.* (2010) Genome-wide analysis of the host intracellular network that regulates survival of *Mycobacterium tuberculosis*. *Cell* 140, 731–743
12 Georgel, P. *et al.* (2010) Virus–host interactions in hepatitis C virus infection: implications for molecular pathogenesis and antiviral strategies. *Trends Mol. Med.* 16, 277–286
13 Bushman, F.D. *et al.* (2009) Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS Pathog.* 5, e1000437
14 Chan, E.Y. *et al.* (2009) Decoding the multifaceted HIV-1 virus–host interactome. *J. Biol.* 8, 84
15 Evans, P. *et al.* (2009) Prediction of HIV-1 virus–host protein interactions using virus and host sequence motifs. *BMC Med. Genomics* 2, 27
16 Dampier, W. *et al.* (2009) Host sequence motifs shared by HIV predict response to antiretroviral therapy. *BMC Med. Genomics* 2, 47
17 de Chassey, B. *et al.* (2008) Hepatitis C virus infection protein network. *Mol. Syst. Biol.* 4, 230
18 Shepard, C.W. *et al.* (2005) Global epidemiology of hepatitis C virus infection. *Lancet Infect. Dis.* 5, 558–567
19 Cornberg, M. *et al.* (2006) Present and future therapy for hepatitis C virus. *Expert. Rev. Anti. Infect. Ther.* 4, 781–793
20 Lindenbach, B.D. and Rice, C.M. (2005) Unravelling hepatitis C virus replication from genome to function. *Nature* 436, 933–938
21 Lemon, S.M. *et al.* (2010) Development of novel therapies for hepatitis C. *Antiviral Res.* 86, 79–92
22 Pan, Q.W. *et al.* (2007) New therapeutic opportunities for hepatitis C based on small RNA. *World J. Gastroenterol.* 13, 4431–4436
23 Pezacki, J.P. *et al.* (2010) Host–virus interactions during hepatitis C virus infection: a complex and dynamic molecular biosystem. *Mol. Biosyst.* 6, 1131–1142
24 Randall, G. *et al.* (2003) Clearance of replicating hepatitis C virus replicon RNAs in cell culture by small interfering RNAs. *Proc. Natl. Acad. Sci. U. S. A.* 100, 235–240
25 Elbashir, S.M. *et al.* (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411, 494–498
26 Tai, A.W. *et al.* (2009) A functional genomic screen identifies cellular cofactors of hepatitis C virus replication. *Cell Host Microbe* 5, 298–307
27 Berger, K.L. *et al.* (2009) Roles for endocytic trafficking and phosphatidylinositol 4-kinase III alpha in hepatitis C virus replication. *Proc. Natl. Acad. Sci. U. S. A.* 106, 7577–7582
28 Supekova, L. *et al.* (2008) Identification of human kinases involved in hepatitis C virus replication by small interference RNA library screening. *J. Biol. Chem.* 283, 29–36
29 Ng, T.I. *et al.* (2007) Identification of host genes involved in hepatitis C virus replication by small interfering RNA technology. *Hepatology* 45, 1413–1421
30 Workman, P. *et al.* (2010) Drugging the PI3 kinome: from chemical tools to drugs in the clinic. *Cancer Res.* 70, 2146–2157
31 Barrett, T. *et al.* (2009) NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res.* 37, D885–D890
32 Pezacki, J. *et al.* (2009) Transcriptional profiling of the effects of 25-hydroxycholesterol on human hepatocyte metabolism and the antiviral state it conveys against the hepatitis C virus. *BMC Chem. Biol.* 9, 2
33 Basu, A. *et al.* (2006) Microarray analyses and molecular profiling of Stat3 signaling pathway induced by hepatitis C virus core protein in human hepatocytes. *Virology* 349, 347–358
34 Budhu, A. *et al.* (2007) Induction of a unique gene expression profile in primary human hepatocytes by hepatitis C virus core, NS3 and NS5A proteins. *Carcinogenesis* 28, 1552–1560
35 Blackham, S. *et al.* (2010) Gene expression profiling indicates the roles of host oxidative stress, apoptosis, lipid metabolism, and intracellular transport genes in the replication of hepatitis C virus. *J. Virol.* 84, 5404–5414
36 Joyce, M.A. *et al.* (2009) HCV induces oxidative and ER stress, and sensitizes infected cells to apoptosis in SCID/Alb-uPA mice. *PLoS Pathog.* 5, e1000291
37 Walters, K.A. *et al.* (2006) Host-specific response to HCV infection in the chimeric SCID-beige/Alb-uPA Mouse model: role of the innate antiviral immune response. *PLoS Pathog.* 2, e59
38 Bigger, C.B. *et al.* (2001) DNA microarray analysis of chimpanzee liver during acute resolving hepatitis C virus infection. *J. Virol.* 75, 7059–7066
39 Su, A.I. *et al.* (2002) Genomic analysis of the host response to hepatitis C virus infection. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15669–15674
40 Lederer, S. *et al.* (2006) Distinct cellular responses differentiating alcohol- and hepatitis C virus-induced liver cirrhosis. *Virol. J.* 3, 98
41 Smith, M.W. *et al.* (2006) Gene expression patterns that correlate with hepatitis C and early progression to fibrosis in liver transplant recipients. *Gastroenterology* 130, 179–187
42 Bièche, I. *et al.* (2005) Molecular profiling of early stage liver fibrosis in patients with chronic hepatitis C virus infection. *Virology* 332, 130–144
43 Carthew, R.W. and Sontheimer, E.J. (2009) Origins and mechanisms of miRNAs and siRNAs. *Cell* 136, 642–655
44 Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233
45 Steuerwald, N.M. *et al.* (2010) Parallel microRNA and mRNA expression profiling of (genotype 1b) human hepatoma cells expressing hepatitis C virus. *Liver Int.* 30, 1490–1504
46 Liu, X. *et al.* (2010) Systematic identification of microRNA and messenger RNA profiles in hepatitis C virus-infected human hepatoma cells. *Virology* 398, 57–67
47 Peng, X. *et al.* (2009) Computational identification of hepatitis C virus associated microRNA-mRNA regulatory modules in human livers. *BMC Genomics* 10, 373
48 Pedersen, I.M. *et al.* (2007) Interferon modulation of cellular microRNAs as an antiviral mechanism. *Nature* 449, 919–922
49 Jackson, A. and Linsley, P.S. (2010) The therapeutic potential of microRNA modulation. *Discov. Med.* 9, 311–318
50 Jopling, C.L. *et al.* (2006) Positive and negative modulation of viral and cellular mRNAs by liver-specific microRNA miR-122. *Cold Spring Harb. Symp. Quant. Biol.* 71, 369–376
51 Henke, J.I. *et al.* (2008) microRNA-122 stimulates translation of hepatitis C virus RNA. *EMBO J.* 27, 3300–3310
52 Jopling, C.L. (2008) Regulation of hepatitis C virus by microRNA-122. *Biochem. Soc. Trans.* 36, 1220–1223
53 Marquez, R.T. *et al.* (2010) Correlation between microRNA expression levels and clinical parameters associated with chronic hepatitis C viral infection in humans. *Lab. Invest.* 90, 1727–1736
54 Lanford, R.E. *et al.* (2010) Therapeutic silencing of microRNA-122 in primates with chronic hepatitis C virus infection. *Science* 327, 198–201
55 Hou, W. *et al.* (2010) MicroRNA-196 represses Bach1 protein and hepatitis C virus gene expression in human hepatoma cells expressing hepatitis C viral proteins. *Hepatology* 51, 1494–1504
56 Alexiou, P. *et al.* (2009) Lost in translation: an assessment and perspective for computational microRNA target identification. *Bioinformatics* 25, 3049–3055
57 Chakravarti, B. *et al.* (2010) Proteomics and systems biology: application in drug discovery and development. *Methods Mol. Biol.* 662, 3–28
58 Kang, S.M. *et al.* (2005) Proteomic profiling of cellular proteins interacting with the hepatitis C virus core protein. *Proteomics* 5, 2227–2237
59 Tsutsumi, T. *et al.* (2009) Proteomics analysis of mitochondrial proteins reveals overexpression of a mitochondrial protein chaperon, prohibitin, in cells expressing hepatitis C virus core protein. *Hepatology* 50, 378–386
60 Mannova, P. *et al.* (2006) Modification of host lipid raft proteome upon hepatitis C virus replication. *Mol. Cell. Proteomics* 5, 2319–2325
61 Shen, K. and Tseng, G.C. (2010) Meta-analysis for pathway enrichment analysis when combining multiple genomic studies. *Bioinformatics* 26, 1316–1323
62 Woodhouse, S.D. *et al.* (2010) Transcriptome sequencing, microarray, and proteomic analyses reveal cellular and metabolic impact of hepatitis C virus infection *in vitro*. *Hepatology* 52, 443–453
63 Randall, G. *et al.* (2007) Cellular cofactors affecting hepatitis C virus infection and replication. *Proc. Natl. Acad. Sci. U. S. A.* 104, 12884–12889
64 Landry, Y. and Gies, J.P. (2008) Drugs and their molecular targets: an updated overview. *Fundam. Clin. Pharmacol.* 22, 1–18
65 Russ, A.P. and Lampel, S. (2005) The druggable genome: an update. *Drug Discov. Today* 10, 1607–1610