Minireview

Decoding the multifaceted HIV-1 virus-host interactome

Eric Y Chan, Marcus J Korth and Michael G Katze

Address: Department of Microbiology, University of Washington, Seattle, WA 98195-8070, USA.

Correspondence: Michael G Katze. Email: honey@u.washington.edu

Abstract
Recently in *BMC Medical Genomics*, Tozeren and colleagues have uncovered virus-host interactions by searching for conserved peptide motifs in HIV and human proteins. Their computational model provides a novel perspective in the interpretation of high-throughput data on the HIV-host interactome.

The replication of the retrovirus HIV is no simple task. After binding to specific receptors on the cell surface, fusion between the virion and the host membrane releases the viral core into the cytoplasm. The viral RNA genome is then reverse transcribed into DNA by the viral reverse transcriptase and the proviral DNA is integrated into the host genome. The host RNA polymerase II complex reads the integrated DNA and synthesizes viral RNA, which is subsequently translated into proteins. All these viral products then have to be routed properly for assembly so that progeny virions can leave through the plasma membrane.

As its genome encodes just 15 genes, HIV depends on the coordinated actions of cellular factors throughout its replication cycle. However, with more than 25,000 human protein-encoding genes to choose from (according to RefSeq), it is inconceivable that all of the approximately 10^{66} possible virus-host interaction combinations would be involved in virus replication. Computational prediction of protein interactions has typically been based on binding mediated by protein domains, each of which can be up to 500 amino acids in size. However, the limited size of the HIV gene products restricts the number of protein domains available for such analyses; to put this in perspective, the HIV Tat protein is composed of just 88 amino acids. In an alternative approach to the identification of virus-host interactions contributing to HIV replication, Evans *et al.* [1] have developed a strategy that aims to predict interactions of cellular proteins with conserved viral sequence motifs or with short eukaryotic linear motifs (ELMs [2]) common to both virus and host; different ELMs of three to eight amino acids in length were found conserved in all viral proteins, thus providing more candidates for matching with host interaction partners. By the same strategy, Dampier *et al.* [3] have uncovered

correlations between response to HIV drug therapy and the conservation of select ELMs localized to the viral reverse transcriptase (RT), further implicating interactions between viral and cellular factors in disease progression.

Various high-throughput strategies have been used in the search for cellular factors that contribute to HIV replication (Table 1). Recently, cell-line-based RNA interference (RNAi) screens [4-6] have emerged as useful tools for the discovery of such factors. In addition, results from functional proteomic and transcriptomic analyses of primary cell infections have shown that perturbations in the host cellular environment over the course of virus replication vary between cell types and viruses. Arguably, as the virus travels through the host cell, anything that happens in the cell can be considered to have an impact on virus replication. The computational model put forth by Evans *et al.* [1] helps to indicate the mechanism of at least part of such contributions, namely those that occur through conserved binding interactions between cellular factors and viral proteins.

Identification of cellular contributors to HIV replication from conserved sequence motifs

Using conserved ELMs found in viral genes, Evans *et al.* [1] identified complementary 'counter domains' on cellular proteins that serve as binding partners for the viral proteins. Owing to the high mutation rates of RNA viruses, the preservation of these viral sequence motifs would presumably confer viral fitness. From alignments of more than 70 HIV-1 sequences, 56 ELMs were found to be conserved among the individual HIV-1 gene products. At first glance, the identification of 2,348 predicted cellular interaction partners - close to 10% of the human protein-encoding genes - seems unreasonably high. However, the dataset includes both a set of direct viral-host protein interactions (called H1) and a larger set of competitive interactions (H2). The H2 set comprises cellular proteins that share ELMs with viral proteins and that may not, therefore, necessarily bind to viral proteins. The matching of viral proteins to counter domains is likely to increase in the future, because the discovery of H1 and H2 is limited by the number of ELMs in the ELM database, currently only 133 [2].

Table 1

High-throughput analyses of virus-host interactions in HIV infection

Type of screen	Cell type	Virus	References
siRNA screen	HeLa-CD4	HIV-IIIB	[4]
	293T	VSV-G-pseudotyped, HIV pNL4-3	[6]
	HeLa-CD4	HIV HXB2	[5]
Transcriptomic	Primary CD4	HIV NL4-3	[8]
Proteomic	Primary CD4	HIV LAI	[9]
miRNA microarray	H9 T lymphocytes	HIV NL4-3	[10]

Looking at the H1 set alone, the authors [1] reported enrichments in cellular proteins that bind to the viral proteins Tat, Env, and Nef. In accordance with the apoptosis-inducing properties of gp120 (a protein formed from cleavage of Env) and Nef, the authors found (by Gene Ontology and KEGG pathway analysis) an enrichment in cellular partners of Env and Nef involved in apoptosis and protein kinase signaling in the H1 set. In line with the role of Tat as a transactivator, there was an enrichment in cellular partners along the Ras p38/mitogen activating protein kinase signaling cascade, including the small GTPase Ras itself.

Identification of cellular contributory factors using RNAi screens

As partial validation, Evans *et al.* [1] compared their predictions with the findings of several genome-wide RNAi screens, with varying results. In those surveys, almost the entire human genome was screened by individual gene silencing in HIV infection or transfection systems. The RNAi screens rest on the assumption that any reduction in HIV production resulting from the knockdown of an individual gene implies a role for that gene in virus replication.

Using a HeLa-CD4 cell line, Brass *et al.* [4] and Zhou *et al.* [6] both reported more than 200 host-dependency factors. Following transfection with small interfering RNAs (siRNAs), Brass *et al.* [4] implicated cellular factors on the basis of their effects on two indicators: the production of viral Gag p24 antigen during early viral replication; and the infectivity of culture supernatant from siRNA-transfected cells. The screen by Zhou *et al.* [6] was carried out in a similar manner, but with infectivity monitored over a longer period of time. A similar screen by König *et al.* [5] was carried out with a vesicular stomatitis virus glycoprotein (VSV-G)-pseudotyped virus in 293T cells; while substituting the HIV envelope with VSV-G allowed for the incorporation of the HIV core into the 293T cells, the experimental design precluded the identification of host factors that participate in viral entry mediated by fusion of the HIV envelope with the cellular membrane. As a result,

only events from uncoating of the viral core to viral gene expression were detectable by that siRNA screen.

Despite the differences in experimental designs, the three screens [4-6] reported similar findings at the level of cellular functions. In the two screens performed in HeLa-CD4 cells that captured the production of infectious virions, both groups [4,6] identified large numbers of cellular factors involved in Tat-mediated transcription as part of the Mediator complex, and also cellular factors involved in energy metabolism regulated by the Akt kinase. Similarly, Brass *et al.* [4] and König *et al.* [5] both identified constituents of the nuclear pore complex, presumably related to the nuclear entry of the viral pre-integration complex. However, although all three screens [4-6] each reported between 200 and 300 cellular genes, together they reported a total of 842 unique genes as contributing to HIV replication. Only three genes were commonly reported by all three screens.

Perhaps more significantly, as revealed in a recent meta-analysis of the genome-wide screens [7], several known cellular cofactors of HIV replication were not identified by any of the siRNA screens [4-6]. Missing were the cell-surface antigens HLA-B57 and HLA-C, both of which regulate the immune response to HIV and have known effects on viral loads and disease progression; also missing was the integration cofactor LEDGF/p75. Although components of the ubiquitylation machinery were identified, members of the Tsg101/ESCRT pathway, which is specifically co-opted for the virus to leave the cell, were also missing. In addition, proteins such as the Gag-binding protein cyclophilin A and the HIV long terminal repeat binding partner Sp1 were identified only in the 293T-based screen [5]. The false negatives arising from siRNA screens can be attributed to the fact that cellular proteins that are particularly abundant and stable cannot be easily knocked down by siRNA transfection in the time available in these experiments. Furthermore, RNAi specificity requirements and siRNA off-target effects can introduce false positives and negatives into the analysis. Considering the various gaps in identification by the siRNA screens, it should come

as no surprise that Evans *et al.* [1] reported a moderate overlap between findings from the siRNA screens [4-6] and those that they made on the basis of conserved ELMs and counter domains.

The contribution of global transcriptomic and proteomic perturbations to HIV replication

Advances in high-throughput transcriptomic and proteomic approaches have allowed the comprehensive characterization of the 'natural' environment of HIV replication, namely the cellular factors around the virus in the cell types it infects. Following a transcriptomic approach, Imbeault *et al.* [8] have recently reported a microarray study of HIV-infected primary CD4 T cells. This study revealed the induction of type-I interferons and the subsequent upregulation of p53 transcription following infection with the NL4-3 strain of HIV-1. Likewise, in a recent proteomic analysis of primary CD4 cells infected with the LAI strain of HIV-1, our group [9] reported signs of apoptosis following infection on the basis of accumulation of poly-(ADP)-ribose polymerase (PARP) and indications of compromised mitochondrial integrity. Furthermore, treatment with a reverse transcriptase inhibitor reversed many of the protein abundance changes associated with robust virus replication, thus allowing us to identify perturbations in karyopherin-mediated nuclear trafficking and mRNA-splicing factors associated with DNA repair as unique host response signatures that accompanied HIV replication [9].

More recently, Nathans *et al.* [10] performed a microRNA microarray analysis, revealing a specific cellular microRNA (miR-29) that targeted the HIV-1 3' untranslated region, thereby repressing virus production. Unlike the examples discussed thus far, this represents a possible mechanism by which the virus modulates immune evasion rather than facilitating viral replication directly.

Complementary approaches for decoding the virus-host interactome

The identification of cellular factors that contribute to HIV replication will pave the way for the development of antivirals that are more resilient to the evolution of viral drug resistance; Maraviroc, an analog of the chemokine receptor CCR5, is a step in this direction. To this end, one could argue we can simply focus on identifying cellular factors that promote HIV infectivity (Figure 1a). However, given the technical limitations with the coverage gaps of siRNA screens and the difficulty in adapting primary cells for screening purposes, complementary high-throughput approaches need to be adopted for the full elucidation of virus-host interactions and their contribution to HIV replication. As reported by Dampier *et al.* [3], the conservation of different ELMs in the HIV reverse transcriptase can predict a patient's response to therapy, pointing to the potential utility of virus-host interactions as a prognostic
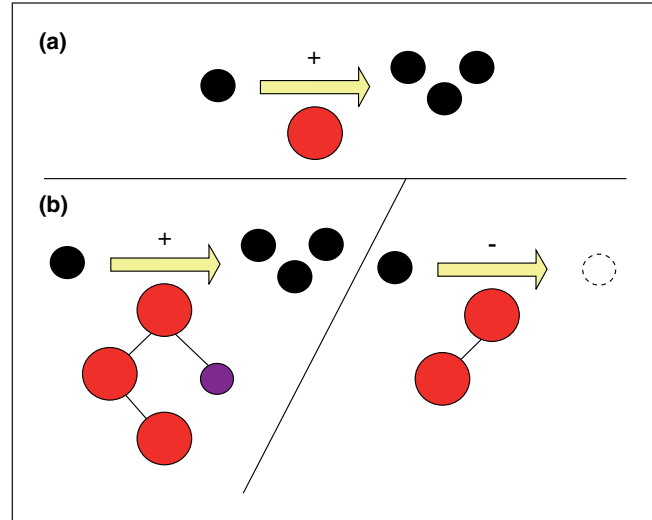

Figure 1

Is there a straight path to identifying cellular contributory factors to HIV infection? **(a)** In an ideal world, the identification of cellular factors (red circles) that contribute (+) to propagation of virions (black circles) would be the most relevant piece of information. However, for reasons reviewed in this article, there is no straight path that would provide 100% coverage. **(b)** By taking advantage of the computational approach of Evans *et al.* [1] and complementary high-throughput approaches [4-6,8-10], we will maximize our ability to identify such contributory cellular factors. In the process, we will also uncover interaction networks among cellular and viral proteins (purple circle); we might also uncover contributory (left, +) and inhibitory (right, -) factors of infection. The multitude of information will provide additional matrices for data evaluation across platforms, freeing us from the cell-type-specific and viral-strain-specific variables that became apparent in the high-throughput screens reviewed here [4-6,8-10].

tool for response to drug regimens. In this regard, defining the structural basis of interaction - such as by predicting virus-host binding interactions between ELMs and counter domains as described by Evans *et al.* [1] - helps provide the proper context for interpreting the high-density, high-throughput data (Figure 1b).

Acknowledgements

References

1. Evans P, Dampier W, Ungar L, Tozeren A: **Prediction of HIV-1 virus-host protein interactions using virus and host sequence motifs.** *BMC Med Genomics* 2009, **2**:27.
2. **The Eukaryotic Linear Motif Resource for Functional Sites in Proteins** [http://elm.eu.org/]
3. Dampier W, Evans P, Ungar L, Tozeren A: **Host sequence motifs shared by HIV predict response to antiretroviral therapy.** *BMC Medical Genomics* 2009, **2**:47.
4. Brass AL, Dykxhoorn DM, Benita Y, Yan N, Engelman A, Xavier RJ, Lieberman J, Elledge SJ: **Identification of host proteins**

required for HIV infection through a functional genomic screen. *Science* 2008, **319:**921-926.

5. König R, Zhou Y, Elleder D, Diamond TL, Bonamy GM, Irelan JT, Chiang CY, Tu BP, De Jesus PD, Lilley CE, Seidel S, Opaluch AM, Caldwell JS, Weitzman MD, Kuhen KL, Bandyo-padhyay S, Ideker T, Orth AP, Miraglia LJ, Bushman FD, Young JA, Chanda SK: **Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication.** *Cell* 2008, **135:**49-60.

6. Zhou H, Xu M, Huang Q, Gates AT, Zhang XD, Castle JC, Stec E, Ferrer M, Strulovici B, Hazuda DJ, Espeseth AS: **Genome-scale RNAi screen for host factors required for HIV replication.** *Cell Host Microbe* 2008, **4:**495-504.

7. Bushman FD, Malani N, Fernandes J, D'Orso I, Cagney G, Diamond TL, Zhou H, Hazuda DJ, Espeseth AS, König R, Bandyopadhyay S, Ideker T, Goff SP, Krogan NJ, Frankel AD, Young JAT, Chanda SK: **Host cell factors in HIV replication: meta-analysis of genome-wide studies.** *PLoS Pathog* 2009, **5:**e1000437.

8. Imbeault M, Ouellet M, Tremblay M: **Microarray study reveals that HIV-1 induces rapid type-I interferon-dependent p53 mRNA up-regulation in human primary CD4+ T cells.** *Retrovirology* 2009, **6:**5.

9. Chan EY, Sutton JN, Jacobs JM, Bondarenko A, Smith RD, Katze MG: **Dynamic host energetics and cytoskeletal proteomes in human immunodeficiency virus type 1-infected human primary CD4 cells: analysis by multiplexed label-free mass spectrometry.** *J Virol* 2009, **83:**9283-9295.

10. Nathans R, Chu CY, Serquina AK, Lu CC, Cao H, Rana TM: **Cellular microRNA and P bodies modulate host-HIV-1 interactions.** *Mol Cell* 2009, **34:**696-709.

© 2009 BioMed Central Ltd