



Predicting response to hepatitis C therapy

Thomas S. Oh and Charles M. Rice

Center for the Study of Hepatitis C, The Rockefeller University, New York, New York, USA.

Current treatment for chronic hepatitis C is expensive, is often accompanied by burdensome side effects, and, sadly, fails in almost half of cases. The ability to predict such failures prior to treatment could save a great deal of pain and expense for the patient with HCV. In this issue of the *JCI*, Aurora and colleagues describe the development of genetic markers predictive of treatment response based on a study of viral sequence variation (see the related article beginning on page 225). Genome-wide covariation analyses of pretreatment virus sequences from 94 patients showed distinct patterns of mutations strongly associated with the ultimate success or failure of treatment. Such analyses suggest markers predictive of response to therapy and may lead to new insights into the underlying biology of hepatitis C.

An estimated 130 million people worldwide (1) and nearly 4 million in the United States are chronically infected with HCV, leading to liver damage and increased risk of hepatocellular carcinoma. In the United States, 10,000 deaths each year are attributed to chronic HCV infection (2). The current treatment regime, pegylated IFN- α and ribavirin, is long and difficult, requiring months of weekly injections, with serious side effects ranging from flu-like symptoms to depression and autoimmune disorders. Success of treatment is far from guaranteed: in HCV genotype 1 infections, which account for the majority of cases in the US, only about half of patients display the long-term suppression of virus indicative of cure.

Numerous studies in recent years have proposed markers for predicting HCV patient response to therapy. Markers may be based on viral factors, such as viral sequence variation (3); host factors, such as gene expression profiles (4) or polymorphisms in specific host genes (5); or combinations thereof (6, 7). Interestingly, very different types of biomarkers can give similar results, indicative of the intimate interactions between the manifold host and viral players in virus replication and disease progression.

In this issue of the *JCI*, Aurora et al. define a set of biomarkers predictive of the response to HCV therapy (8). These markers are purely viral factors, com-

posed of sets of varying residues in the HCV amino acid sequence identified by covariation analysis.

Covariation analysis reveals functional relationships

A statistical measure, covariance quantifies the degree of linkage between 2 variables; variables that are completely independent have a low covariance, whereas variables that vary synchronously have a high covariance.

Covariance between residues in a protein or set of proteins can be estimated from the variation observed in a population. An alignment of multiple HCV sequences shows both conserved and varying residues. The varying positions are compared in pairwise fashion; for each pair of positions, the linkage between the 2 residues will affect the pattern of variation observed. For a pair of positions with a 10% mutation frequency at each site, both mutations would be shared by 1% of sequences if they are perfectly independent and 10% if they are perfectly covariant.

Because covariation implies a relation between 2 residues in a sequence, it has been used to infer information about direct interactions in the 3-dimensional structure of a protein (9) and to identify protein-protein interactions (10). However, covariance arises from all functional interactions between residues, both direct and indirect, as well as from phylogenetic relationships (Figure 1). Distinguishing between the many sources of covariance is a continuing challenge for anyone wishing to use this technique (11, 12).

Covariation patterns are highly correlated with treatment outcome

The Viral Resistance to Antiviral Therapy of Chronic Hepatitis C (Virahep-C) clinical

study (13) evaluated the efficacy of treatment in HCV genotype 1a and 1b patients. The complete HCV coding sequence was determined for pretreatment isolates from each of 94 patients, who were followed during and after treatment to determine the final outcome of therapy.

In the present study, Aurora et al. analyzed the 94 HCV sequences obtained during the Virahep-C study for amino acid covariance in each of the genotype 1 subtypes as well as stratified within each subtype by treatment response (8). From this analysis they made an important, and perhaps surprising, observation: the sets of covariant pairs were markedly different between the responsive and nonresponsive patient groups. In the HCV genotype 1a sequences, about 2,000 covariant residue pairs were identified; three-quarters of the covariant pairs found in the responsive genomes did not appear in the nonresponsive sequence set, and vice versa. The results of the HCV genotype 1b sequence analysis was even more striking: 90% of the residue pairs identified as being covariant in one response group were independent in the other group.

The strong correlation between covariance sets and therapeutic outcome immediately suggests the possibility of finding a reliable predictor for response to therapy in the pretreatment HCV sequence. However, there is still an additional step that must be made; a patient coming in for treatment generally harbors a range of closely related viral sequences. Covariance, on the other hand, is an aggregate property determined from a sequence alignment of an entire group of responders or nonresponders. The covariance sets reported by Aurora et al. showed a clear difference between groups of sequences depending on response to therapy (8), but a biomarker must be able to place a single sequence of unknown response into the correct group. In order to bridge this gap, the authors looked to the interconnected nature of the covariance sets they had generated.

Covariance networks

Each covariation analysis performed by Aurora et al. identified on the order of 2,000 pairs of correlated residues (8). How-

Conflict of interest: The authors have declared that no conflict of interest exists.

Nonstandard abbreviations used: Virahep-C, Viral Resistance to Antiviral Therapy of Chronic Hepatitis C [study].

Citation for this article: *J. Clin. Invest.* 119:5–7 (2009). doi:10.1172/JCI38069.

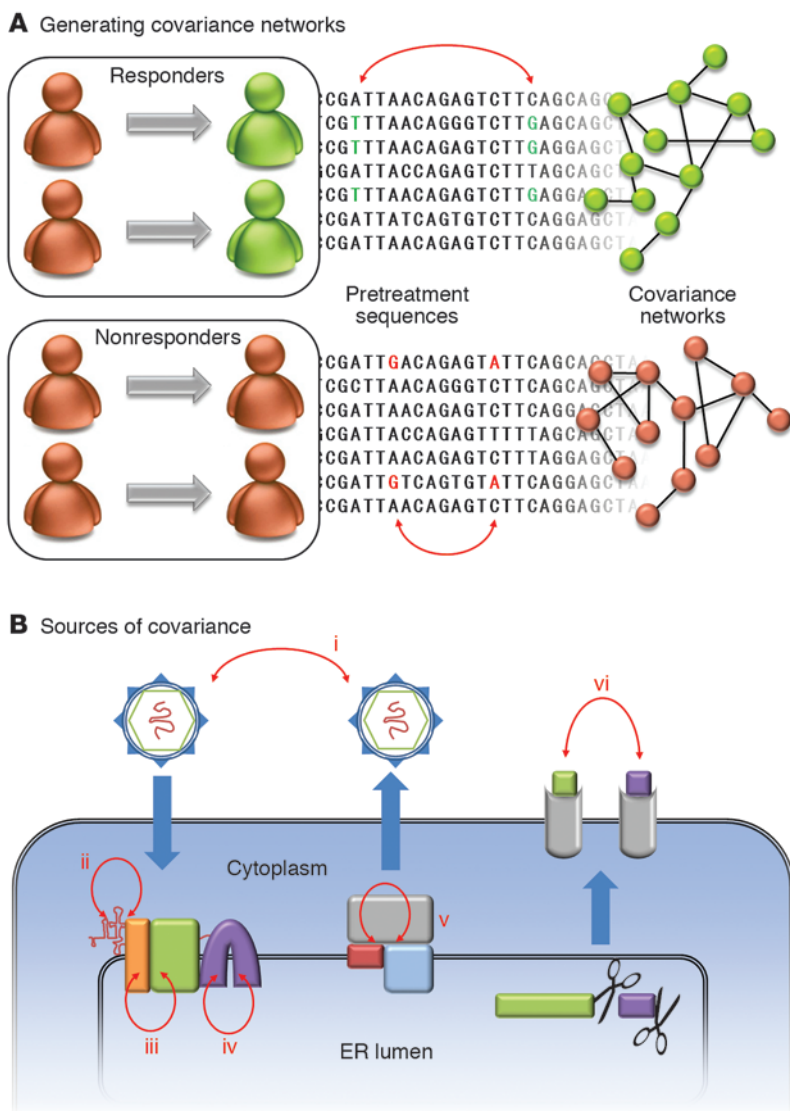


Figure 1

Covariation in HCV. **(A)** In the study reported in this issue of the *JCI* by Aurora et al. (8), patients were grouped according to their treatment response. The sequences of the complete HCV open reading frame obtained from each group of patients prior to treatment were aligned and analyzed for covariance. An example covariant pair is shown in each alignment (red arrows). The set of all covariant pairs forms a network in which each node is an amino acid position and each connecting line represents covariance between 2 positions. The networks differ by treatment response class and may be used to generate markers indicative of HCV treatment outcome. **(B)** Various causes of covariance in HCV (red arrows). (i) Phylogenetic covariance is an artifact of a shared ancestry, but does not reflect any functional relationship. (ii) RNA secondary structure gives rise to nucleotide-level covariance. (iii) Protein-protein interaction residues covary. (iv) Intraprotein covariance may indicate direct residue contact or indirect interaction through the protein. (v) Variation in a shared interaction partner (host or viral) may result in coordinated variation in a pair of residues. (vi) MHC epitopes will covary across hosts with different HLA alleles.

ever, this set of 2,000 pairs is composed of only about 200 unique residues. Clearly a residue may appear multiple times; in fact, each residue in the set was connected to anywhere between 1 and 100 other residues. The resulting networks are shown in detail in ref. 8.

Because covariant pairs by definition vary, any one pair will appear in only a fraction of sequences. Similarly, a combination of residues correlated with one outcome can appear in a sequence of the opposite outcome, not because the residues are functionally linked, but simply by chance. For this reason, the authors searched for small collections of interconnected pairs, or subnetworks, which were correlated with outcome. By means of exhaustive search, they identified several hundred such subnetworks, which appeared in greater than

95% of sequences of one therapeutic outcome and never appeared in sequences of the opposite outcome (8).

The attentive reader will note — and the authors are quick to point out — that the sequences for which the markers are evaluated are the same sequences used to generate the markers. This is attributed to the unavailability of other sequence sets for which the treatment outcome is known. Nevertheless, the authors provide evidence that the differences observed in the covariance networks are real and will translate into markers that will hold up outside the initial data set. First, the difference in the covariance sets between the 2 possible outcomes is quite large, as much as 90%. Second, the subnetwork analysis yielded not a handful of potential markers, but hundreds of subnetworks with 100% corre-

lation to treatment outcome. Finally, and most interestingly, the chemical makeup of the covariant pairs is significantly different; the nonresponsive sequences contain 3 times as many hydrophobic covariant amino acid pairs as the responsive sequences. This unexpected result implies that the differences in the covariance networks are directly reflective of an underlying physical phenomenon. The authors suggest that the higher fraction of correlated hydrophobic residues is evidence for more stable protein-protein complexes in the nonresponsive strains. This could be envisioned to result in viral replication complexes that are more resistant to antiviral effectors, or even to alter interactions of immunomodulatory HCV proteins with their target host factors. Analysis of covariance networks may therefore not only reveal biomarkers for



therapeutic outcome, but also shed light on the mechanistic bases for resistance to treatment and even identify novel targets for antiviral drugs.

Conclusions

Although it still remains for these markers to be validated, the early results presented in this study are promising (8). It is interesting to speculate on the relationship between these markers and other markers, particularly those based on host characteristics. The circulating virus is not an independent entity, but is continually shaped by host selective pressures even as it in turn modulates its host environment. Viral sequences observed prior to treatment may very well represent the success or failure of the host in selecting against the most treatment-resistant variants. Covariance networks may serve as an exciting new tool in further studies along this avenue; networks generated from viral sequences obtained during acute viral infection should be particularly informative.

With the sustained and rapid growth of both computational power and sequencing capabilities, we expect covariation analyses to become increasingly common as a tool to study different aspects of HCV biology (14). The high mutation rate of RNA viruses and the intense competition within the

quasispecies makes them particularly amenable to this technique. We look forward to seeing further application of covariance networks to questions ranging from protein structure and protein-protein interactions to drug resistance, host selection pressures, and viral evolution.

Acknowledgments

The authors acknowledge Catherine Murray for her assistance in the preparation of this manuscript. Work in the laboratory of C.M. Rice is supported by the Greenberg Medical Research Institute and the Starr Foundation.

Address correspondence to: Charles M. Rice, Laboratory of Virology and Infectious Disease, The Rockefeller University, Box 64, 1230 York Avenue, New York, New York 10065, USA. Phone: (212) 327-7046; Fax: (212) 327-7048; E-mail: ricec@rockefeller.edu.

1. Wasley, A., and Alter, M.J. 2000. Epidemiology of hepatitis C: geographic differences and temporal trends. *Semin. Liver Dis.* **20**:1–16.
2. Armstrong, G.L., et al. 2006. The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. *Ann. Intern. Med.* **144**:705–714.
3. Moreau, I., Levis, J., Crosbie, O., Kenny-Walsh, E., and Fanning, L.J. 2008. Correlation between pretreatment quasispecies complexity and treatment outcome in chronic HCV genotype 3a. *Virology*. **5**:78.
4. Torres-Puente, M., et al. 2008. Genetic variability in hepatitis C virus and its role in antiviral treatment

response. *J. Viral Hepat.* **15**:188–199.

5. Persico, M., et al. 2008. Elevated expression and polymorphisms of SOCS3 influence patient response to antiviral therapy in chronic hepatitis C. *Gut*. **57**:507–515.
6. Shirakawa, H., et al. 2008. Pretreatment prediction of virological response to peginterferon plus ribavirin therapy in chronic hepatitis C patients using viral and host factors. *Hepatology*. Online publication ahead of print. doi:10.1002/hep.22543.
7. El-Shamy, A., et al. 2007. Prediction of efficient virological response to pegylated interferon/ribavirin combination therapy by NSSA sequences of hepatitis C virus and anti-NSSA antibodies in pretreatment sera. *Microbiol. Immunol.* **51**:471–482.
8. Aurora, R., Donlin, M.J., Cannon, N.A., and Tavis, J.E., for the Virahep-C Study Group. 2009. Genome-wide hepatitis C virus amino acid covariance networks can predict response to antiviral therapy in humans. *J. Clin. Invest.* **119**:225–236.
9. Gobel, U., Sander, C., Schneider, R., and Valencia, A. 1994. Correlated mutations and residue contacts in proteins. *Proteins*. **18**:309–317.
10. Wang, Y.E., and DeLisi, C. 2006. Inferring protein-protein interactions in viral proteins by co-evolution of conserved side chains. *Genome Inform.* **17**:23–35.
11. Eyal, E., Frenkel-Morgenstern, M., Sobolev, V., and Pietrokowski, S. 2007. A pair-to-pair amino acids substitution matrix and its applications for protein structure prediction. *Proteins*. **67**:142–153.
12. Dunn, S.D., Wahl, L.M., and Gloor, G.B. 2008. Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics*. **24**:333–340.
13. Donlin, M.J., et al. 2007. Pretreatment sequence diversity differences in the full-length Hepatitis C Virus open reading frame correlate with early response to therapy. *J. Virol.* **81**:8211–8224.
14. Campo, D.S., Dimitrova, Z., Mitchell, R.J., Lara, J., and Khudiyakov, Y. 2008. Coordinated evolution of the hepatitis C virus. *Proc. Natl. Acad. Sci. U. S. A.* **105**:9685–9690.

Role for $\alpha 3$ integrin in EMT and pulmonary fibrosis

Zea Borok

Will Rogers Institute Pulmonary Research Center, Division of Pulmonary and Critical Care Medicine, Department of Medicine, University of Southern California, Los Angeles, California, USA.

Idiopathic pulmonary fibrosis (IPF) is characterized by progressive (myo)fibroblast accumulation and collagen deposition. One possible source of (myo)fibroblasts is epithelial cells that undergo epithelial-mesenchymal transition (EMT), a process frequently mediated by TGF- β . In this issue of the JCI, Kim et al. report that epithelial cell-specific deletion of $\alpha 3$ integrin prevents EMT in mice, thereby protecting against bleomycin-induced fibrosis (see the related article beginning on page 213). The authors propose a novel mechanism linking TGF- β and β -catenin signaling in EMT through integrin-dependent association of tyrosine-phosphorylated β -catenin and pSmad2 and suggest targeted disruption of this interaction as a potential therapeutic approach.

Conflict of interest: The author has declared that no conflict of interest exists.

Nonstandard abbreviations used: AEC, alveolar epithelial cell; EMT, epithelial-mesenchymal transition; IPF, idiopathic pulmonary fibrosis.

Citation for this article: *J. Clin. Invest.* **119**:7–10 (2009). doi:10.1172/JCI38084.

Idiopathic pulmonary fibrosis (IPF) is a progressive disorder of unknown etiology characterized by fibroblast accumulation, collagen deposition, and ECM remodeling leading to parenchymal destruction (1). Historically, inflammation has been

viewed as central to the pathogenesis of IPF. A recent paradigm shift proposes a model in which injury to the epithelium initiates a proinflammatory and profibrotic cascade, resulting in fibroblast expansion and progressive fibrosis reminiscent of abnormal wound healing (2). Myofibroblasts (activated fibroblasts) are key effector cells in pulmonary fibrosis, being responsible for matrix deposition and structural remodeling. The source of myofibroblasts in IPF remains the subject of debate: in addition to arising from circulating progenitors and resident fibroblasts, myofibroblasts have recently been shown to be derived from alveolar epithelial cells (AECs) through epithelial-mesenchymal transition (EMT) (3, 4).