Role of HCV sequence variation in liver pathology

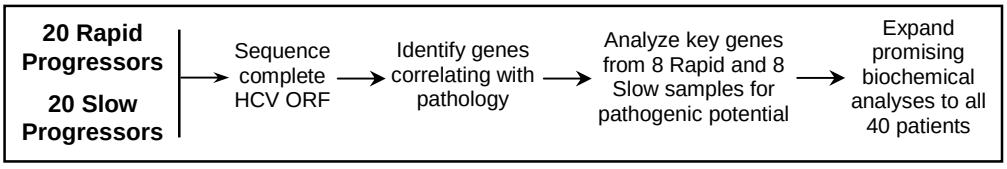
John Tavis, Ph.D., P.I.
Adrian Di Bisceglie, M.D., Co-Investigator and HALT-C representative

A. Aims.

Chronic Hepatitis C virus (HCV) infection causes slowly worsening liver disease, including hepatitis, fibrosis, cirrhosis, liver failure, and hepatocellular carcinoma. The mechanisms underlying liver disease induced by HCV are incompletely understood and the factors governing rate of disease progression are not known. HCV is highly variable genetically, and genetic variation is clearly a significant contributor to virulence in many viral diseases. However, previous analyses of how HCV genetic variation affects disease have been limited to detailed inspection of small regions of the viral genome or to characterizations in which HCV variability was defined only crudely at the level of the genotype. Therefore, a comprehensive analysis of the effect of HCV's high genetic variation on its virulence has not been performed.

HCV infection is treated with pegylated interferon α plus ribavirin, but this therapy fails to clear HCV in roughly half of genotype 1 patients, the most common genotype in the USA. To determine if progression of liver disease in patients who fail anti-HCV therapy can be slowed by long-term administration of interferon α, the NIDDK is sponsoring the HALT-C clinical study with 1050 patients at 10 clinical centers. Half of these patients are receiving long-term interferon α therapy and half are untreated controls. Because of its very large, well-characterized cohort, HALT-C presents a unique opportunity to study how HCV's genetic diversity affects the rate of disease progression.

Hypothesis: Genetic variation in HCV sequences affecting viral replication or immune evasion activities modulates the rate of progression of liver disease.



We will test this hypothesis using a two-pronged approach (Fig. 1). First, we will assess variation in the complete HCV open reading frame in viruses from patients whose disease advanced slowly or rapidly employing the genome-wide sequencing approaches we developed for the NIDDK-sponsored Virahep-C clinical study. Second, we will test how genetic variation associated with HCC development affects the function of HCV proteins implicated in modulating viral virulence.

Aim 1: Determine how HCV sequence variation is associated with progression of liver disease. HCV-induced liver disease is predicted to be influenced by viral sequence variations that modulate virulence through altering replication fitness or the ability to evade host antiviral responses. Therefore, we will sequence the complete HCV ORF in two groups of 20 HALT-C patients shortly after unsuccessful antiviral therapy and again 3.5 years later. Slow Progressors will be patients with minimal disease progression, and Rapid Progressors will be patients with substantial advancement of disease during the study, as defined by the established HALT-C endpoints. All patients will be selected from the untreated arm of HALT-C to assess natural progression of HCV disease. Sequences from viruses infecting the two groups will be compared to determine if there are viral genetic patterns correlating with disease progression and to characterize viral evolution in the context of disease progression.

Aim 2: Determine how HCV genetic variation affects the function of viral proteins implicated in modulating virulence. Variation in viral proteins that affect viral replication levels or the ability of HCV to counteract the immune system is predicted to modulate the pathogenic stimulus induced by HCV. Therefore,

variant core, E2, NS3, and NS5A genes from slow and rapid progressors analyzed in Aim 1 will be cloned and their biochemical activities proposed to promote virulence will be measured. We will also assess other HCV genes for which novel associations with disease progression may be identified in Aim 1. These results will guide interpretation of how viral genetic variation observed in Aim 1 may modulate HCV's pathogenic potential.

This study will provide a comprehensive analysis of the role of variation in the HCV coding region on progression of HCV-induced liver disease. This will yield insight into the mechanism of pathology by identifying viral genes whose variability is associated with differences in disease progression. Finally, these studies may identify viral motifs predictive of rapid development of cirrhosis or hepatic decompensation.

Interaction with the HALT-C study: We will follow all HALT-C policies, including (but not limited to) submitting all data to the HALT-C archives, collaborating with the HALT-C Data Coordinating Center for data analyses, and adhering to the Publication and Presentation Committee's procedures. We understand that in this multi-center collaborative research project, the overall authority for conduct of the study rests with the HALT-C Steering Committee.

Blinding of samples: We appreciate the extreme importance of maintaining the integrity of the data in a long-term multi-center clinical trial to maximize statistical power during data analysis. We have proposed to stratify our patient groups based on the primary outcomes of HALT-C, and hence it is essential that the samples be provided to us in a strictly blinded manner. We therefore request that (i) the Data Coordinating Center identify the samples based upon our criteria but without our participation, and (ii) provide them to us identified *only* by a special code used exclusively for this proposed study. We suggest the code be HC-###-* in which HC stands for HALT-C, the sample number is arbitrarily assigned by the DCC, and * is a letter arbitrarily representing the outcome group (A or B). The code would be maintained only at the DCC and would be broken only upon approval by the Steering Committee following the end of the treatment phase of HALT-C. This would avoid revealing the HALT-C specimen or patient ID numbers to our site and would preclude accidentally compromising the integrity of the data.

Coding the samples in this manner will present no problems to the conduct of the proposed study because the earliest possible funding date (July 2007) is after patient treatment terminates in January 2007.

Sample Utilization: Power analyses for HCV genetic assessments are difficult to perform because the pattern of variation needed to separate one group from another cannot be predicted and may vary for different biological parameters. We have discussed this issue with the biostatisticians at Saint Louis University, and they agree that formal power analyses will not be informative in this case. Therefore, we will rely on our experience in Virahep-C to justify our sample size of 20 per group. In Virahep-C there were 16 patients per group, and we detected highly significant diversity differences correlating with response to therapy in NS3 and NS5A ($p \leq 0.001$). We increased the group size to 20 in this proposed study in case differences correlating with pathology prove less obvious than differences associated with response to therapy.

Funds to Support Additional HALT-C Efforts: Funds will be requested to support the additional effort of the HALT-C DCC needed for this proposed study and for BBI to ship the samples to Dr. Tavis' lab.

IRB Approval and Safety Oversight: This proposed study will be conducted in accordance with Saint Louis University policies under IRB approval (approved project #14138). Safety oversight of this study will fall under the HALT-C Data and Safety Monitoring Board.

Table 1. HALT-C samples required.

Visit	Liver # patients, mm*	Blood # patients, ml	DNA # patients, ug	Patient Time # patients, min	Other (describe) # pts, amount
Screen 1					
Screen 2					
Baseline					
Lead in Week 4					
Week 8					
Week 12					
W16					
Week 20					
Week 24					
Randomized Month 9		60^, 1.0 ml*			
Month 12					
Month 15					
Month 18					
Month 21					
Month 24					
Month 27					
Month 30					
Month 33					
Month 36					
Month 39					
Month 42					
Month 45					
Month 48		40, 1.0 ml*			
Post-treatment					
Responders W30					
W36					
W42					
W48					
W60					
W72					

^ 20 of these patients will be fibrotic progressors whose subgenotype is not known (1, no subtype). We will subtype these by sequencing until the 20 1a fibrotic progressors needed for this study are identified. In Virahep-C, 93% of the "genotype 1-no subtype" samples proved to be 1a upon sequencing, therefore, this is an upper limit for the number of patients we will need.

* Serum or plasma are equally acceptable to us. We will use whichever is in greater supply, at HALT-C's discretion.

Role of HCV sequence variation in liver pathology

John Tavis, Ph.D., P.I.
Adrian Di Bisceglie, M.D., Co-Investigator and HALT-C representative

Revision Request (approved 2/29/08).

We request two modifications to our approved and funded ancillary study:

Modification 1) We propose to sequence the HCV genomes through a collaboration with the Broad Institute at the Massachusetts Institute of Technology rather than in-house.

Modification 2) We propose to expand the sample sizes from 20 rapid progressors and 20 matched slow progressors to 29 rapid and 29 matched slow progressors.

Justifications.

Modification 1) Sequencing full HCV genomes is difficult due to the RNA form of the genome, the relatively low titre of the virus in the bloodstream, and the high genetic variability of the viral sequences which prevents uniform binding of the amplification primers during PCR. Our lab developed a standardized process to do this on moderate numbers of genomes (3) and used this process to analyze 94 full-length HCV open reading frames (~9,000 nt each) (2) as part of the Virahep-C study (1). However, despite the improvements that we pioneered, this process remains relatively laborious and slow. We anticipate that it will take two scientists working ~1.5-2 years to sequence the 80 HALT-C genomes approved for our project (40 subjects, each at early and late time points). Furthermore, this process is expensive because of the large amount of labor it requires.

The Broad Institute at the Massachusetts Institute of Technology is a cutting-edge high-throughput sequencing center. It has recently applied its expertise in sequencing automation and process design to determination of the consensus sequence for HCV genotypes 1a and 1b, and has successfully sequenced over 400 genomes to date (unpublished data). Conceptually, their approach is identical to ours: the HCV genome is reverse transcribed, amplified in overlapping fragments, and then directly sequenced. The difference between our approaches is that our process was designed for a typical molecular biology laboratory, whereas the Broad's approach was designed for a high-throughput sequencing facility.

There are two very significant advantages to collaborating with the Broad Institute. First, the sequence data will be obtained much faster with much less labor on our part. We will isolate the viral RNAs, reverse transcribe them, and perform quality control assays prior to sending the cDNA to the Broad Institute. They will then perform the PCR, sequencing, assembly, and annotation steps. One post-doc in our lab will be able conduct the cDNA syntheses and control assays in a few months, and then the Broad Institute will take only a month or so to fully sequence the samples. The outcome is that we will have the sequences in a few months rather than 1.5-2 years. The second major advantage is cost. The Broad Institute can perform its share of the work for $450 per genome. This will result in a major cost savings to us, as we estimate it costs us nearly $800-$900 per genome in supply costs, plus a much larger sum in salary expenses.

Modification 2) The original sample size of 20 rapid and 20 slow progressors was determined through a combination of practical concerns (cost, labor, sequencing speed) and experience from working with the Virahep-C data set. Formal power analyses were not employed because the relationship between viral sequence variation and disease progression is not understood, and hence we had no good model upon which to base proper power calculations. Therefore, increasing the sample size as much as is possible is important to maximize the power of the study, and consequently its chances of uncovering meaningful answers. We have

discussed this problem extensively with Ms Curto at the HALT-C DCC, and she has identified 29 genotype 1a rapid progresssors (e.g., had a HALT-C clinical outcome) suitable for this analysis and for which sufficient serum is available at the time points we will study. Because there are many more slow progressors in the HALT-C cohort, finding matched slow progressors for the control group will be straight forward.

Other Considerations Relative to the Collaboration with the Broad Institute.

There are a number of additional issues related to the proposed collaboration with the Broad Institute that Dr. Matthew Henn (Research Scientist II, Microbial Genome Sequencing Project) and I have negotiated. Please see the attached letter of support from Dr. Henn.

Issue 1) We require two products from the sequencing procedure to conduct our study: the sequences themselves and the amplified HCV DNAs that were the templates for the sequencing reactions (to be cloned for use in functional assays). Dr. Henn has assured us that the amplified HCV DNAs will be saved and returned to us. In the event that these prove to be unsuitable for our purposes, we will retain a portion of the cDNA used for the sequencing and amplify the regions to be cloned directly from the cDNA employing primers designed based on the sequence of each isolate.

Issue 2) The Broad Institute is a scientific center, not a sequencing service, and hence a condition of their participation is that they be involved in the data analysis. We consider this to be a major advantage to the collaboration. We have extensive sequence analysis capabilities, but HCV sequence analysis is an evolving field, and any additional expertise that can be brought to bear on the problem can only be an asset. This collaboration will be a two-way exchange, as some of the analyses we have performed would be of use to other projects being conducted by the Broad Institute. In some cases, we may seek approval from the HALT-C Ancillary Studies Committee to obtain data needed to include our sequences in studies the Broad Institute is conducting independently of HALT-C. Approval from the Ancillary Studies Committee for these analyses will be sought on a case-by-case basis in the future.

Issue 3) As an NIAID-funded sequencing center, the Broad Institute is required to release all sequences to the public databases as soon as they have undergone appropriate quality control assessments. This is acceptable to us. Some details will need to be finalized concerning the patient information to be included with the sequences (none of the information will permit identification of the patients). These details will be easily settled through future discussions with the Tavis lab, the Broad Institute, and the HALT-C DCC and/or Ancillary Studies Committee.

R**eferences.**

1. **Conjeevaram, H. S., M. W. Fried, L. J. Jeffers, N. A. Terrault, T. E. Wiley-Lucas, N. Afdhal, R. S. Brown, S. H. Belle, J. H. Hoofnagle, D. E. Kleiner, and C. D. Howell**. 2006. Peginterferon and ribavirin treatment in african american and caucasian american patients with hepatitis C genotype 1. Gastroenterology **131**:470-477.

2. **Donlin, M. J., N. A. Cannon, E. Yao, J. Li, A. Wahed, M. W. Taylor, S. H. Belle, A. M. Di Bisceglie, R. Aurora, and J. E. Tavis**. 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.

3. **Yao, E., J. E. Tavis**. 2005. A general method for nested RT-PCR amplification and sequencing the complete HCV genotype 1 open reading frame. Virol. J. **2**:88.