APPENDIX A: HALT-C Ancillary Study PROPOSAL

Part I (1 page)
Proposal Name: **ROLE OF THE SYSTEMIC METABOLOME IN HEPATITIS C DISEASE PROGRESSION**

Proposal PI: Dr. Arun Sanyal

Co-Investigators: Dr. Richard Sterling

HALT-C PI: Dr. Richard Sterling

Funding Agency and Review Body (e.g., NIDDK; my university/GAC): NIDDK with left over funds to Virginia Commonwealth University

I agree to follow HALT-C Policies and Procedures when conducting this study. I acknowledge that the data obtained from this study will belong to the NIH and will be placed in the HALT-C database for use by other investigators. I understand that I cannot begin experiments using HALT-C specimens/data until I receive approval from the HALT-C Ancillary Studies Committee and funding from the Scientific Review Body for my proposal. I also understand that the data analysis for this proposal will be performed by NERI (unless otherwise approved by the HALT-C study) and that Protocols approved by the HALT-C Ancillary Studies Committee will be placed on the HALT-C Restricted Website.

__Arun Sanyal, MD_____ _____
Proposal Principal Investigator Date

__Richard Sterling_____ _____
HALT-C Principal Investigator Date

Protocol Part II (4 page limit, single space)

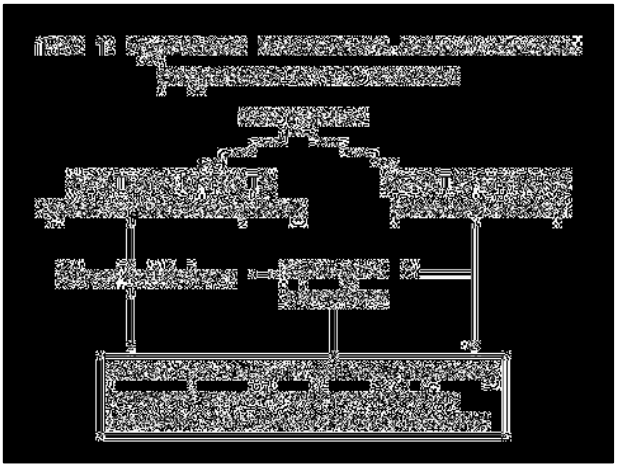
1. Aims/hypotheses:

SPECIFIC AIM: **To define the role of metabolomic changes as drivers of disease progression in hepatitis C virus infection**

Hypothesis: *HCV virus infection is associated with changes in the systemic metabolome which in turn affect inflammatory, apoptotic, oxidative stress and fibrosis related pathways in the liver and contribute to disease progression and also to the propensity for development of diabetes mellitus*.

2 & 3. Background/rationale/and relations to aims of HALT-C study:

Summary of activities of parent grant (HALT-C) and how they encompass this proposal: The HALT-C trial focused on disease progression in hepatitis C virus (HCV) infection and evaluated the effects of PEG-interferon for retarding this process. The clinical outcomes of this trial were negative and there are currently no established methods to slow disease progression in subjects with HCV infection other than achieving sustained virologic response to anti-viral therapy. A major factor limiting the development of such therapies is the relative lack of understanding of the mechanisms involved in disease progression in cases where standard anti-viral therapy fails. The current proposal will attempt to bridge this gap in knowledge by using a systems biology approach to evaluate the interactions between changes in the systemic metabolome and the expression of inflammatory, apoptotic and fibrotic pathways in the liver in subjects with HCV using samples obtained in the course of the HALT-C trial. The objective is to define the role of metabolomic changes as drivers of disease progression in HCV infection. We *expect* to define the key pathogenic targets that are up- or down-regulated in HCV infection and contribute to disease progression. The *rationale* for these studies is that it will: (1) provide novel insights in to the changes in the systemic metabolome in subjects with HCV, and (2) provide information on the impact of changes in the metabolome on the expression of inflammatory, apoptotic, oxidative stress and fibrosis related pathways in the liver. By providing data on pathways that can potentially contribute to disease progression in subjects with HCV infection and advanced fibrosis, this proposal aligns itself with the larger goals of the HALT-C trial.



4. Study design, experimental groups:

APPROACH: In this proposal, we will test the hypothesis that HCV infection produces alterations in the systemic metabolome that reflect the inflammatory state produced by HCV. It also has direct effects on the hepatic genome and indirect effects due to intrahepatic inflammation and oxidative stress. These changes lead to complex alterations in hepatic gene expression and result in altered hepatic metabolism, activation of inflammation, apoptosis and fibrotic pathways. This will be studied by an analysis of the systemic metabolome and hepatic gene expression in subjects from the control arm of the HALT-C trial who developed liver failure (defined outcome defined below) within 24 months of randomization or of month 24 biopsy. The potential impact of prior anti-viral therapy will be defined by analysis of those of another control group of treatment naïve subjects with or without advanced fibrosis.

Innovative aspects of this proposal and why VCU is highly suited to perform the proposed studies:

- Cutting edge technology platform for metabolomics (www.metabolon.com)
- Systems biology approach and bioinformatics expertise (www.venebio.com)
- Expertise in hepatic metabolism [1-4]
- Expertise in the metabolomics [2, 5]
- Expertise in micro-array profiling studies [6]

- Expertise in clinical aspects of hepatitis C and metabolic changes [7-9]

Inclusion criteria:

Test group: Subjects receiving no active treatment in the HALT-C cohort who developed liver failure (defined below) within 24 months of randomization or month 24 biopsy.

Control group 1: Subjects receiving no active treatment in the HALT-C cohort who did not develop liver failure within 24 months of randomization or month 24 biopsy.

Control group 2: Treatment naïve subjects with hepatitis C with advanced fibrosis

Control group 3: Treatment naïve subjects with hepatitis C who do not have advanced fibrosis

Exclusion criteria: Subjects in the HALT-C cohort who received PEG-interferon will be excluded.

JUSTIFICATION OF INCLUSION AND EXCLUSION CRITERIA: A key objective of this study is to determine the factors that drive development of liver failure in subjects with HCV infection. The inclusion of subjects with HCV who develop liver failure is therefore a logical study group. Comparison with those who do not develop liver failure but are otherwise similar in every respect will provide information on changes that are specifically seen in those who develop liver failure. Comparison of control group 1 to those who are treatment naïve but otherwise similar will provide information on the potential impact of prior anti-viral therapy on changes seen in control group 1. Comparison of data from control groups 2 and 3 versus treatment naïve subjects without advanced fibrosis will provide important new information on the role of metabolomic changes that are associated with advanced fibrosis in subjects with HCV infection. Subjects who received PEG-interferon will be excluded because interferon therapy will confound analysis.

Sources of subjects:

Test and control group 1: These samples will be obtained from the plasma and tissue repository of the HALT-C trial.

Control groups 2 and 3: These samples have already been collected and are available to the investigators. Matching will be done for age (± 5 yrs), gender, race, BMI (± 3) and use of statins. Please note that there will be no need to prospectively consent and collect any samples.

5. Methods, data usage

SAMPLES TO BE USED AND PROCEDURES TO BE PERFORMED:

1. **Plasma (500 µl each from baseline): The following measurements will be made:**
   - Metabolomic profile: This will be performed using LC:MS as previously described [2].
   - Cytokine profile: These will be measured independently using a multiplex assay previously used by the investigators.
2. **Snap-frozen liver tissue:** This will be obtained from baseline biopsy samples. The HALT-C trial has already banked snap-frozen tissue for total mRNA extraction. If tissue is provided, we will extract total mRNA using Trizol reagent and return cDNA to the repository for other investigators to use. If cDNA has already been extracted, we will use the available cDNA for the proposed studies.

mRNA isolation and expression analysis: This will be done using a standard Illumina gene expression chip SA Biosciences (CS550). Key differentially expressed mRNA will be confirmed by real time quantitative PCR as previously described [6].

PLAN OF ANALYSIS:

Definition of liver failure: This will be defined by a CTP≥ 7 on 2 successive visits 3 months apart within 24 months after SOO or M24 liver biopsy. Although these subjects may also have ascites or development of hepatic encephalopathy, these are more related to portal hypertension rather than liver failure and will not be required.

Analysis of the lipid metabolome (Quantitative analysis of plasma lipids in HCV): ANOVA and pair-wise t-tests on groups of subjects will be used to define lipid profiles. Simple linear regression will be used to find associations with other continuous variables over the study population e.g. cytokine levels etc. Z scores will be used to identify changes in metabolites. Metabolon's custom-designed software will be used to take quantitative metabolite data to estimate steady-state flux of lipid and other metabolites through key metabolic pathways. This is based on creation of a correlation matrix between metabolites which is mapped on to metabolic pathways based on the Kyoto encyclopedia of genes and genomes (KEGG) and represented graphically [10, 11]. Pathway maps will show estimated bulk flow of metabolites through specific pathways and

changes in either the concentration or relative flux of metabolites between the study groups. These will provide novel information on metabolic pathways in subjects with hepatitis C. Specifically, novel data will be obtained regarding changes in the metabolome in those with or without advanced fibrosis. Any differences in the lipid profile of those with advanced fibrosis who develop liver failure in the next 12 months versus those who do not will also be ascertained by this approach.

Characterization of the relationship between circulating lipids versus liver histology etc: The relationship between liver histology (none to early fibrosis vs advanced fibrosis, high inflammatory scores vs low scores, steatosis etc) and changes in metabolites especially those reflecting lipogenesis (MUFA:SFA, oleic:stearate and palmitoleic:palmitate ratio), cycloxygenase activation (thromboxanes), lipoxygenase activation (5, 8 or 15-HETE), peroxisomal dysfunction (DPA:DHA ratio) and oxidative stress (11-HETE) will be used for this analysis [2]. Analyses will also provide information on differences between those with or without advanced fibrosis and those who do or do not develop liver failure. Clustering and discriminant analyses will be performed to identify relationships between variables and presence/absence of specific histologic features. In separate analyses, the relationship of changes in various histologic parameters of HCV with the lipid metabolome will be studied in the various study groups. The impact of potential confounders e.g. diabetes, timing of when the sample was obtained (at time of biopsy versus at a distant time) will be studied using logistic regression analyses. The biologic and biochemical plausibility of these relationships will be examined and will lead to a smaller set of variables that will serve as an input to classifier methods to identify responders and also for hypothesis generation about the pathogenesis of disease progression in HCV.

What is the impact of prior anti-viral therapy on the metabolome? This will be evaluated by comparison of subjects with advanced fibrosis who are in the HALT-C cohort and are non-responders to antiviral therapy to those with advanced fibrosis who are treatment naïve (control group 2). A priori, we do not anticipate seeing any substantive changes based on prior exposure to antiviral therapy.

Are there specific changes in the metabolome associated with advanced fibrosis? This will be evaluated by comparison of subjects with advanced fibrosis to those without advanced fibrosis (control group 3). If prior analyses indicate that there are no differences in the metabolome in those who have or have not received prior antiviral therapy, the data from control groups 1 and 2 will be pooled. On the other hand, if differences are noted, then data from those who are treatment naïve and have advanced fibrosis (control group 2) will only be compared to those without advanced fibrosis because the latter group is also treatment naïve.

Chemometric modeling to identify impact of changes in the metabolome on specific disease pathways: It will utilize both the metabolomic data obtained by the studies proposed above but also transcriptomic data obtained from total mRNA samples obtained from the baseline liver biopsy for the HALT-C cohort. The plan of analysis will be similar to that described by Laaksonen et al [12]. Chemometric modeling of the metabolomic data will be performed using partial least square discriminant analysis as a supervised modeling method [13]. Venetian blinds cross validation method and Q^2 methods will be used to optimize the model [14]. Top loading for latent variables associated with disease states (advanced disease) will be reported and the variable importance in the projection values calculated to identify the most important molecular species for clustering of specific groups using Matlab (ver. 7.2) software and the PLS Toolbox (Eigenvector Research Inc). This method will use just the metabolomic data (or combined metabolomic and a gene expression data subset) to distinguish between groups (with or without advanced fibrosis or those who do or do not develop liver failure). The regression of lipidomic data on hepatic gene expression will be performed using the lasso method [15, 16]. The lasso regression coefficients will be calculated with the least angle regression method implemented in the R statistical language (package LARS) [16]. The corrected R^2 value and the Schwarz criterion will be reported along with the measured and predicted gene expression values [17]. This method will link metabolomic changes with specific gene expression changes. Additional analyses using random forest methodology will provide information on whether specific changes can be used to classify pathways as significantly changed.

What specific pathways worsen in those who develop liver failure? The relationships between changes in expression of inflammatory, apoptotic and fibrotic pathways and development of liver failure will be assessed

from data from the test group versus control group 1. Also, the interaction between changes in the metabolome and cytokine profile on one hand versus the transcriptome on the other and changes in bilirubin, albumin and coagulant factor synthetic pathways will be evaluated. We anticipate that these proposed analyses will lead to hypothesis generation for future focused studies of changes in specific metabolites and their impact on pathways leading to disease progression and liver failure.

Is hepatic biologic pathway activation related by circulating metabolites that reflect oxidative stress or anti-oxidant response? Gene Set Enrichment Analysis (GSEA) [18] of the mRNA expression data will be used to identify activated hepatic biologic pathways related to development of steatosis, inflammation, fibrosis, apoptosis, bilirubin, albumin and coagulant factor expression that are affected in those with advanced fibrosis versus those without and in those who develop liver failure versus those who do not. Regression of oxidative stress-associated metabolites on the genes that comprise the activated pathways will be used to determine whether their expression is potentially driven by oxidative stress or anti-oxidant response.

Is there evidence for gene-gene interactions in HCV modulation of the biologic pathways that drive disease progression? We will use the method of Schafer & Strimmer [19], to identify gene-gene interactions driven by HCV. This analytical method uses mRNA expression data to infer gene association networks and will thus identify gene transcription cascades in which a gene directly modulated by HCV results in a broader transcriptional response by way of its regulation of downstream target genes.

Sample size: For the primary analysis, differences in mean change in key metabolites and mRNA levels in specific pathways in those who develop liver failure will be compared to those who do not in the HALT-C cohort. The hypothesis to be tested is that changes in key pathways in those develop liver failure are significantly different from those who do not. Using a relatively conservative *p* value cutoff of 0.025, a sample size of 9 subjects per group will identify a 25% difference between groups, assuming a standard deviation of 15% for the metabolite or transcript in question, with a power of 80%. The error rate will be further reduced by avoiding inter-individual variability by using each subject as their own control. Based on the available samples (10/test group) and equal number for test groups, the sample size is adequate according to traditional power analysis for single targets, and will likely be sufficient even when correcting for multiple tests.

Statistical Support: All analysis will be done at VCU from the data set provided by the DCC. We will request the dataset to include demographic, medication, laboratory (including HCV RNA and lipid profile and HOMA), histology (including steatosis grade), anthropometric (BMI etc) data.

Financial support: This will come from left over funds at our site after approval of NIDDK.

6. Anticipated results: We *expect* to define the key pathogenic targets that are up- or down-regulated in HCV infection and contribute to disease progression.

Potential Pitfalls and alternate approaches:

- Micro-array, and to a lesser extent metabolomic, data can suffer from the problems of sensitivity and issues related to false discovery. We have reduced a major source of error by using each subject as their own control. Multiple testing will be controlled for using contemporary false discovery rate (FDR) approaches [20]. Also, the GSEA procedures for pathway analysis using the gene expression data are less susceptible to these false discovery issues and will likely increase our power to detect meaningful changes [18]. Furthermore, key genes that are differentially expressed across groups will be independently validated using real time PCR.
- mRNA isolation will be done first. We will first extract RNA from 4 biopsies for quality control. If these are good, we will proceed to the rest. If not, we will limit the studies to plasma metabolomics only and use the liver tissue for metabolomic analysis and combine the two to obtain novel integrative information about the metabolic response to HCV and disease progression.
- The changes in pro-fibrotic pathways may be specifically present in stellate cells. In this study, we will have total mRNA from the liver tissue which will mainly reflect changes in hepatocytes. Thus, potential

changes in stellate cells may be missed. In future studies, we will use laser capture micro-dissection to isolate fibrotic areas which are enriched in stellate cells to try to get around this issue. We will also focus on pathways such as epithelial-mesenchymal transition pathways that are more likely to be altered in hepatocytes and also are involved in fibrosis in the proposed studies.

- Ideally a fasting plasma sample would be best. However, given the limited availability of specimens and the low probability of changes in phospholipids and eicosanoids in plasma due to diet, we believe the overall limitation will be relatively minor.

Future Directions: We believe these studies will provide data for more detailed hypothesis-driven future studies of the complex interactions between the metabolome and transcriptome in driving disease progression in HCV infection. We anticipate future RO1 applications focused around these areas.

Timelines:

Mth 1	Mth 2	Mth 3	Mth 4	Mth 5	Mth 6	Mth 7	Mth 8	Mth 9	Mth 10	Mth 11	Mth 12
Samples to be sent for analysis			Sample analysis			Complete data analyses, identify additional hypotheses, perform additional analyses as appropriate			Publications and plans for additional grant applications		
			Establishing GSEA pathways and fine tuning methodologies for analysis								

REFERENCES

1. Sanyal AJ, Campbell-Sargent C, Mirshahi F, et al. Nonalcoholic steatohepatitis: association of insulin resistance and mitochondrial abnormalities. Gastroenterology 2001;120(5):1183-92.
2. Puri P, Weist, M.M., Cheung, OP, Mirshahi, F, Sargeant, C, Min, HK, Contos, MJ, Sterling, RK, Fuchs, M, Zhou, HP, Watkins, SM, and Sanyal AJ. . The plasma lipidomic signature of nonalcoholic steatohepatitis. Hepatology 2009;in press.
3. Puri P, Sanyal AJ. Sirtuins and the hepatic regulation of energy homeostasis. Hepatology (Baltimore, Md 2009;50(5):1668-70.
4. Sanyal AJ, Contos MJ, Sterling RK, et al. Nonalcoholic fatty liver disease in patients with hepatitis C is associated with features of the metabolic syndrome. The American journal of gastroenterology 2003;98(9):2064-71.
5. Puri P, Baillie RA, Wiest MM, et al. A lipidomic analysis of nonalcoholic fatty liver disease. Hepatology (Baltimore, Md 2007;46(4):1081-90.
6. Cheung O, Puri P, Eicken C, et al. Nonalcoholic steatohepatitis is associated with altered hepatic MicroRNA expression. Hepatology (Baltimore, Md 2008;48(6):1810-20.
7. Sanyal AJ, Banas C, Sargeant C, et al. Similarities and differences in outcomes of cirrhosis due to nonalcoholic steatohepatitis and hepatitis C. Hepatology (Baltimore, Md 2006;43(4):682-9.
8. Negro F, Sanyal AJ. Hepatitis C virus, steatosis and lipid abnormalities: clinical and pathogenic data. Liver Int 2009;29 Suppl 2:26-37.
9. Sterling RK, Contos MJ, Sanyal AJ, et al. The clinical spectrum of hepatitis C virus in HIV coinfection. Journal of acquired immune deficiency syndromes (1999) 2003;32(1):30-7.
10. Katajamaa M, Miettinen J, Oresic M. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. Bioinformatics (Oxford, England) 2006;22(5):634-6.
11. Steuer R. Review: on the analysis and interpretation of correlations in metabolomic data. Briefings in bioinformatics 2006;7(2):151-8.
12. Laaksonen R, Katajamaa M, Paiva H, et al. A systems biology strategy reveals biological pathways and plasma biomarker candidates for potentially toxic statin-induced changes in muscle. PLoS ONE 2006;1:e97.
13. Barker M RW. Partial least squares for discrimination. J CHemometrics 2003;17:166-73.
14. Wise BM GN, Bro R, Shaver JM, Winding W. PLS Toolbox 3.5 for use with Matlab. . 1987.
15. Tibshirani. Regression shrinkage and selection via the lasso. J R Statist Soc 1996;B 58:267-88.
16. Efron B JI, Hastie T, Tibshirani R. Least angle regression. Ann Statist 2004;32:407-99.
17. Pindyck RS RD. Econometric models and economic forecasts. 1998.

18. Ackermann M, Strimmer K. A general modular framework for gene set enrichment analysis. BMC bioinformatics 2009;10:47.
19. Schafer J, Strimmer K. An empirical Bayes approach to inferring large-scale gene association networks. Bioinformatics (Oxford, England) 2005;21(6):754-64.
20. Pawitan Y, Michiels S, Koscielny S, Gusnanto A, Ploner A. False discovery rate, sensitivity and sample size for microarray studies. Bioinformatics (Oxford, England) 2005;21(13):3017-24.

Protocol Part III: Sample Requirements. (link to web site with actual sample availability)

Visit	Liver # patients, mm*	Blood # patients, ml	DNA # patients, ug	Liver Biopsy Slides # patients, slides/patient	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					
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					
Post-treatment					
Responders W30					
W36					
W42					
W48					
W60					
W72					

* Assume 1 mm tissue weighs about 0.75 mg (= 0.5 mm^2 X Π X density of tissue)

Data needed (please specify):
Comments (if any):