

RESEARCH ARTICLE

SRM targeted proteomics in search for biomarkers of HCV-induced progression of fibrosis to cirrhosis in HALT-C patients

Shizhen Qin¹, Yong Zhou¹, Anna S. Lok², Alex Tsodikov³, Xiaowei Yan¹, Li Gray¹, Min Yuan¹, Robert L. Moritz¹, David Galas^{1*}, Gilbert S. Omenn^{1,4*} and Leroy Hood¹

¹Institute for Systems Biology, Seattle, Washington, USA

²Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA

³Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA

⁴Department of Internal Medicine and Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA

The current gold standard for diagnosis of hepatic fibrosis and cirrhosis is the traditional invasive liver biopsy. It is desirable to assess hepatic fibrosis with noninvasive means. Targeted proteomic techniques allow an unbiased assessment of proteins and might be useful to identify proteins related to hepatic fibrosis. We utilized selected reaction monitoring (SRM) targeted proteomics combined with an organ-specific blood protein strategy to identify and quantify 38 liver-specific proteins. A combination of protein C and retinol-binding protein 4 in serum gave promising preliminary results as candidate biomarkers to distinguish patients at different stages of hepatic fibrosis due to chronic infection with hepatitis C virus (HCV). Also, alpha-1-B glycoprotein, complement factor H and insulin-like growth factor binding protein acid labile subunit performed well in distinguishing patients from healthy controls.

Received: November 14, 2011

Revised: January 6, 2012

Accepted: January 16, 2012

**Keywords:**

Biomedicine / Fibrosis / Hepatitis C / Liver-specific blood biomarkers / Selected reaction monitoring / Quantitation

1 Introduction

Human liver is the primary target for damage by hepatitis viruses, many chemicals and drugs. Liver biopsy is recommended in the management of patients with chronic hepatitis C (CHC) to provide important information about fibrosis stage and disease prognosis [1]. As an invasive procedure, liver biopsy is frequently accompanied by transient pain and may occasionally be associated with serious complications.

The accuracy of liver biopsy in staging liver disease is limited by the size and quality of the samples and sampling error [2].

In recent years, intensive research in the field of non-invasive tests of liver fibrosis has yielded a few laboratory markers, which enabled the assessment of some aspects of the severity of hepatitis C virus (HCV)-induced liver disease. For example, the FibroTest, combines six serum markers (Alpha-2-macroglobulin, Haptoglobin, Apolipoprotein A1, Gamma-glutamyl transpeptidase, Alanine transaminase, and total bilirubin) with the age and gender of the patient to generate a score that correlates with stage of fibrosis in patients with a variety of liver diseases [3]. Platelet counts [4], AST/ALT ratio [5], and AST-platelet ratio index (APRI) [6] have been reported as predictors of degree of fibrosis in CHC patients.

Correspondence: Dr. Leroy Hood, Institute for Systems Biology, 401 Terry N, Seattle, WA 98109

E-mail: Leroy.Hood@systemsbiology.org

Fax: +206-732-1299

Abbreviations: **A1BG**, alpha-1-B glycoprotein; **CFH**, complement factor H; **CHC**, chronic hepatitis C; **HALT-C**, Hepatitis C Antiviral Long-term Treatment against Cirrhosis; **HCV**, hepatitis C virus; **IGFALS**, insulin-like growth factor binding protein acid labile subunit; **PROC**, protein C; **RBP4**, retinol binding protein 4

*Additional corresponding authors: Dr. David Galas, E-mail: David.Galas@systemsbiology.org; Dr. Gilbert Omenn, E-mail: gomenn@med.umich.edu

Colour Online: See the article online to view Figs. 1 and 2 in colour.

In the hepatitis C antiviral long-term treatment against cirrhosis (HALT-C) Trial, a model based on a combination of standard laboratory tests comprising platelet count, AST/ALT ratio, and INR (international normalized ratio of prothrombin time) predicted histological cirrhosis with high accuracy in 50% of patients with CHC [7].

Molecular signatures specific for the liver and/or specific for particular causes of liver damage would be very useful for experimental, clinical, and epidemiology studies of liver diseases. Through development of targeted proteomic assays utilizing SRM incorporating heavy-isotope doping of labeled matched peptides [8, 9], and use of the Human SRMAtlas with optimized transitions associated with typically six different peptides for nearly all of the 20,300 human protein-coding genes (Kusebauch, manuscript in preparation), accurate quantitation of target proteins can be achieved for most human proteins present at levels that can be detected by targeted mass spectrometry. Here, we report application of SRM proteomics to identify many liver-specific proteins and to characterize the critical progression of fibrosis of the liver to cirrhosis in CHC patients. This objective is quite challenging in the population we studied as liver biopsies used to stage fibrosis may not be accurate for all the patients and the states being compared represent a continuum of liver damage.

2 Materials and methods

2.1 Serum samples

Serum samples were obtained from patients who participated in the HALT-C Trial [10]. Typical clotting time is 30 min to 2 h. This trial enrolled patients with CHC who had liver biopsies showing Ishak stages 2–6 (range 0–6) fibrosis at enrollment. Blood samples at enrollment were studied. Patient information at enrollment is listed in Supporting Information Table S1. Control sera from normal female and male donors ages 30–50 years were collected at FDA-regulated blood facilities with a clotting time of about 60 minutes and were nonreactive for HCV antibody (ProMedDx). Pooled plasma from ten normal donors was obtained using EDTA as anticoagulant from Innovative (Novi, Michigan). Collection and use of control and patient samples were approved by institutional review boards. Samples were stored at -80°C .

2.2 Sample preparation for SRM

To reduce the complexity of samples, the top 14 highly abundant proteins were depleted using an AKTA FPLC system (GE Healthcare, Piscataway, NJ, USA) coupled with a Seppro[®] IgY14 human LC2 depletion column (Sigma-Aldrich, St. Louis, MO, USA). We observed significant sample-to-sample variations with the Seppro[®] IgY14 spin column. In contrast, the LC system coupled with an IgY14 human LC2 depletion column dramatically improved the reproducibility (Sup-

porting Information Fig. S1). All 40 HALT-C and 17 normal serum samples were processed similarly; about 95% of the total protein was depleted. Proteins in depleted sera were denatured in 50% (v/v) 2,2,2-trifluoroethanol (TFE, J.T. Baker, Phillipsburg, NJ, USA) for 30 min at 55°C , then cysteine residues were reduced and alkylated. Samples were diluted ten-fold with 100 mM NH_4HCO_3 (pH 8.3) before adding trypsin (1:25 trypsin versus serum protein) for overnight digestion at 37°C . Peptides were desalted with Oasis MCX cartridges (Waters, Milford, MA, USA).

2.3 Building the liver-specific and liver-enriched proteins list

We used a targeted approach focusing on organ-specific proteins to increase the likelihood of identifying protein biomarkers in blood that may reflect pathology of a particular organ. Our list of liver-specific or liver-enriched proteins (liver proteins) was created by mining multiorgan transcriptomic data generated through Massively Parallel Signature Sequencing (MPSS). The MPSS dataset contains transcriptomes of 34 pooled normal (Caucasian) human tissues [11]. Signatures with their expression levels in liver either five-fold higher than any other organs or two-fold greater than the sum of all other organs were selected as liver protein candidates. We also performed organ-specific protein search with Gene Atlas Interface analysis. The databases searched against were three datasets from NCBI-GEO (Gene Expression Omnibus) with a total of 180 human tissues from multiple donors [12–14]. We also included 21 enzymes and other proteins used in clinical practice or previously reported as liver biomarker candidates.

2.4 Peptide selection from the liver protein list

Two to three peptides were selected for each liver protein based on the sequence of individual liver proteins that were previously detected by MS. Peptide selection criteria are as follows [15]: (i) length 8–20 amino acid residues; (ii) no chemically unstable residues (single letter notation; M, NG, DG, QG, N-terminal N, and N-terminal Q); (iii) LC-compatible; (iv) avoid cysteine residue if possible; and (v) sequence specific for the target protein (e.g. proteotypic peptides). Peptides previously identified in PeptideAtlas [16] were preferentially chosen. All peptides used in this study were checked by BLAT (<http://genome.ucsc.edu/>) and Protein BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) searches to ensure that they are unique to the target protein at both proteomic and genomic levels. Finally, the uniqueness of every Q1/Q3 pair from the target peptides was confirmed by an SRM theoretical collision calculator tool (<http://proteomics.washington.edu/cgi-bin/srmcalc.cgi>). Mass tolerances for searching both Q1 and Q3 were ± 0.35 Da, which matches the Agilent QQQ settings for SRM.

2.5 MS and HPLC

All SRM analyses were performed on an Agilent 6460A triple quadrupole (QQQ) mass spectrometer with a ChipCube nano-electrospray ionization source coupled with an Agilent 1200 nanoFlow HPLC system. Serum samples were eluted over a 60-min gradient with 0.66% per minute acetonitrile slope in the presence of 0.1% formic acid using a large capacity Agilent HPLC chip (Cat number G4240–62101, 160 nL trap, 150 mm C18 column). Spray voltage was set at 1900 V. The scheduled SRM were performed with 5-min retention time windows and an instrument cycle time of 2000 ± 500 ms. Dwell times varied depending on the number of concurrent transitions; in all cases they were at least 10 ms.

2.6 Monitoring liver-specific proteins in blood by SRM

Crude unpurified peptide standards that correspond to the detected natural counterparts (light peptides) were synthesized with heavy isotopic Lysine ($^{13}\text{C}_6^{15}\text{N}_2$) or Arginine ($^{13}\text{C}_6^{15}\text{N}_4$) at the C-termini (heavy peptides) (Thermo-Fisher Scientific, Ulm, Germany or Sigma-Aldrich). Collision energies (CE) were determined using the default formula from Agilent ($0.036 \times \text{precursor mass } m/z - 4.80$) and then optimized with four additional CE steps (± 5 V, ± 10 V). The best four transitions were selected. Detected heavy peptides were titrated at six concentrations in a normal human serum background to build a titration curve and to determine the proper amount of each peptide standard to spike-in. (Supporting Information Fig. S2).

2.7 SRM data analysis

All SRM data were processed using the Skyline Targeted Proteomics Environment (v1.1) [17]. The setting of 0.055 Th match tolerance m/z was used. The default peak integration and Savitzky–Golay smoothing algorithm were applied. All data were manually inspected to ensure correct peak detection and accurate integration. Peptides with at least three-fold signal-to-noise ratio were considered detectable. The total peak area and light/heavy ratio of each peptide were exported for statistical analysis.

2.8 Statistic analysis

The exported SRM results were analyzed using R scripts generated for this data set, using standard methods of supervised analysis from R Bioconductor [18]. The key challenge for the present analysis is a high dimension of the feature vector (large number of potentially predictive proteins). Our preprocessing selection included proteins that have no more than 30% missing data. A similar criterion was applied to samples.

Missing data were handled using k -nearest neighbor imputation algorithms ($k = 10$) [19]. Repeated (duplicate) measurements for the same protein-peptide- m/z combination were averaged. Platelet level and gender were included as a clinical predictor of liver damage. Regularization methods based on logistic regression were used to reduce overfitting: LASSO (least absolute shrinkage and selection operator) and Elastic Nets penalty [20] were applied. The choice of the optimal regularization parameter was done using the Area Under the Receiver-Operating-Characteristic (AUROC) curve as a criterion. Such regularized regression methods make it possible to handle high dimension within the model-based (regression) classification framework and to improve the multivariate correlation of the panel under design with the disease risk. In order to obtain an approximately unbiased assessment of the performance of predictive signatures, ten-fold cross validation was used to correct for potentially overoptimistic model building and signature optimization bias. An average over cross-validation runs is reported as the final optimal area under curve (AUC) characterizing the predicted performance of the biomarker signature. LASSO penalty was preferred for its ability to drop nonessential biomarkers from the signature by explicitly assigning them zero weights. All other analyses including calculation and graphics were generated by Prism 5 (GraphPad software, La Jolla, CA, USA).

3 Results

We adopted an organ-specific protein-based strategy for blood biomarker discovery. The approach is centered on the idea that concentration of organ-specific proteins in the blood can be used to monitor the health status of a specific organ because changes in blood concentrations reflect the normal as opposed to disease-perturbed status of their cognate biological networks. We mined comprehensive transcriptomic databases to identify organ-specific proteins. To demonstrate the power of this approach, we undertook a study to identify biomarkers related to HCV infection. We have found several liver proteins that can be used to separate HCV-infected patients from healthy individuals. In addition, the levels of two proteins showed strong correlations with progressive stages of the disease.

3.1 The identification of liver proteins

Using strategies described in Section 2.3, we identified 109 liver proteins that passed GeneCards verification. GeneCards summarized each gene's expression in normal and diseased human tissues by three categories: (i) mRNA expression data from GeneNote and GNF BioGPS, (ii) UniGene electronic Northern, and (iii) SAGE (Serial Analysis of Gene Expression). In combination with the 21 proteins that have been used in clinic practice or reported as liver biomarker candidates, a list composed of 130 proteins was created.

3.2 Proteins detected by SRM

After suitable peptides and transitions for each liver protein were selected, we used control plasma and serum to determine how many liver proteins can be detected by SRM. From the 89 liver proteins previously observed in MS/MS experiments, we detected 100 peptides derived from 54 proteins in pooled control plasma (Supporting Information Table S2). However, the HALT-C samples were in the form of serum, which is not ideal for mass spectrometry-based blood protein measurements due to variation in proteolysis derived from the coagulation cascade, resulting in decreased concentrations of proteins compared to plasma [21]. In order to determine how many of our proteins detected in plasma can be detected in serum; we performed the same SRM analyses against pooled healthy human serum samples. Altogether 38 proteins (represented by 65 peptides) were detected in sera (Supporting Information Table S3). All SRM parameters and results are deposited in the SRM chromatographic repository at ISB and are publicly available (<http://www.srmatlas.org> and <http://www.peptideatlas.org/passel/>). Typical SRM chromatograms of five target peptides are presented in Supporting Information Fig. S3.

3.3 Consistency and accuracy of SRM data

3.3.1 Duplicate SRM runs are well correlated

Duplicate runs were performed for each sample, technical variations between the two runs were generally small (Fig. 1A). Pearson tests showed good correlations between runs (≥ 0.9 in most cases; e.g. PROC = 0.90, RBP4 = 0.90, CFH = 0.95, A1BG = 0.94). Samples from two HALT-C patients, DA000739 and DJ000004, were eliminated from the analysis due to sample degradation.

3.3.2 Protein levels measured by multiple features are consistent

When a protein level is measured by more than one feature (i.e. multiple peptides or same peptide with differently charged precursor ions), close agreement in quantification was observed; an example is the set of three features for the protein of A1BG (Fig. 1B). This observation gave us reasonable confidence that protein levels in samples estimated even from a single peptide can be reliable.

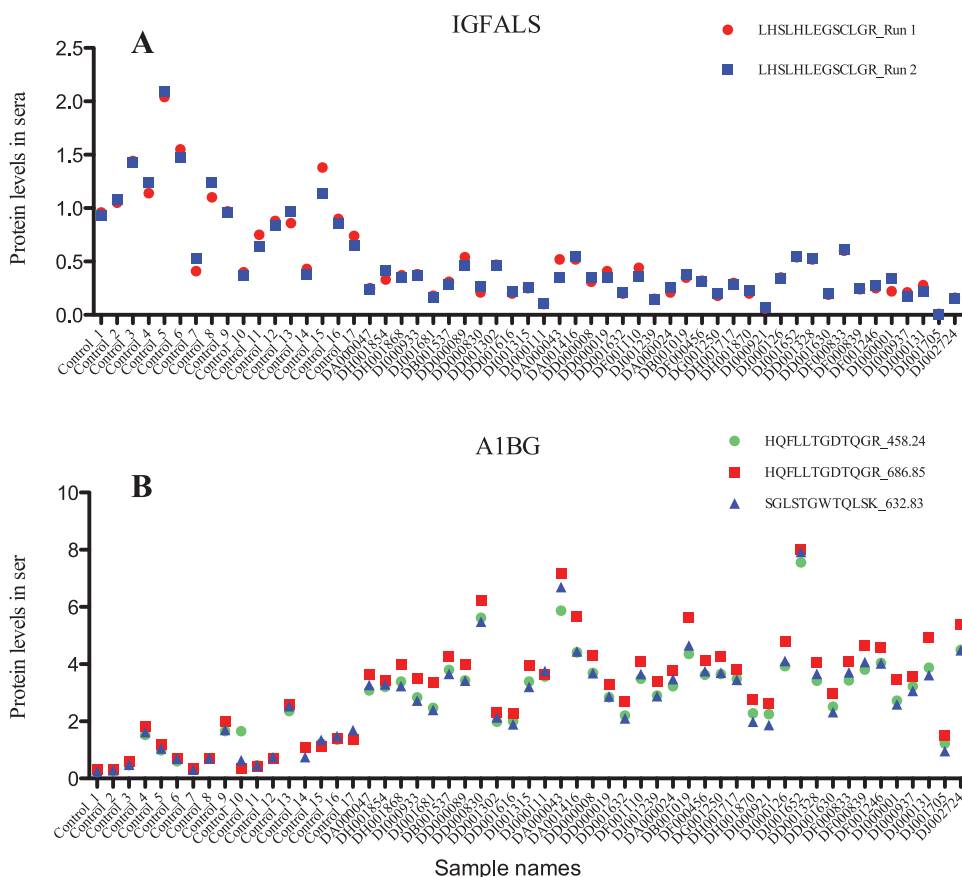


Table 1. Five informative protein concentrations in control sera measured by crude heavy peptides in SRM

NCBI official symbol	Peptide sequence	Average conc. in SRM assay (pg/mL)	Conc. range in SRM assay (pg/mL)	Ref_1 conc. (pg/mL) [22, 23]	Ref_2 conc. (pg/mL) [24]
A1BG	SGLSTGWTQLSK	5.53E07	1.23E07–1.62E08	2.20E8	5.00E + 07
CFH	CTSTGWIPAPR	5.51E07	2.35E07–1.07E08		5.70E + 07
IGFALS	VAGLLEDTFPGLLGLR	6.52E05	1.64E05–1.36E06		1.50E + 06
PROC	TFVLNFIK	1.00E05	2.61E04–1.72E05	3.70E06	6.50E + 04
RBP4	YWGVASFLOK	3.22E06	1.72E06–5.79E06	3.17E07	5.80E + 08

Assuming purity of crude peptide is 80% for 10-mers and 70% for 15-mers, respectively.

3.3.3 Absolute protein levels in sera measured by SRM are similar to prior report

In this study, we did not aim to quantify absolute protein concentrations. The crude heavy peptides synthesized by rapid process were not appropriate for absolute quantitation of protein levels in specimen due to the wide range of purity (~50–80%). In addition, the immunodepletion procedure during sample preparation induces additional variation in protein concentrations. Nevertheless, as summarized in Table 1, levels of five informative proteins in depleted control sera measured by SRM are close to concentrations reported in pub-

lished literature. The only exception is RBP4; the concentration in our SRM study is ten-fold lower than the studies of Gahne et al. and Polanski et al. [22, 23] and 100-fold lower than the study of Farrah et al. [24].

3.4 Five informative proteins were found to distinguish normal controls and HCV-infected patients or CHC patients at different stages of fibrosis

We identified five proteins, namely A1BG, CFH, IGFALS, PROC, and RBP4, that are able to separate HCV-infected

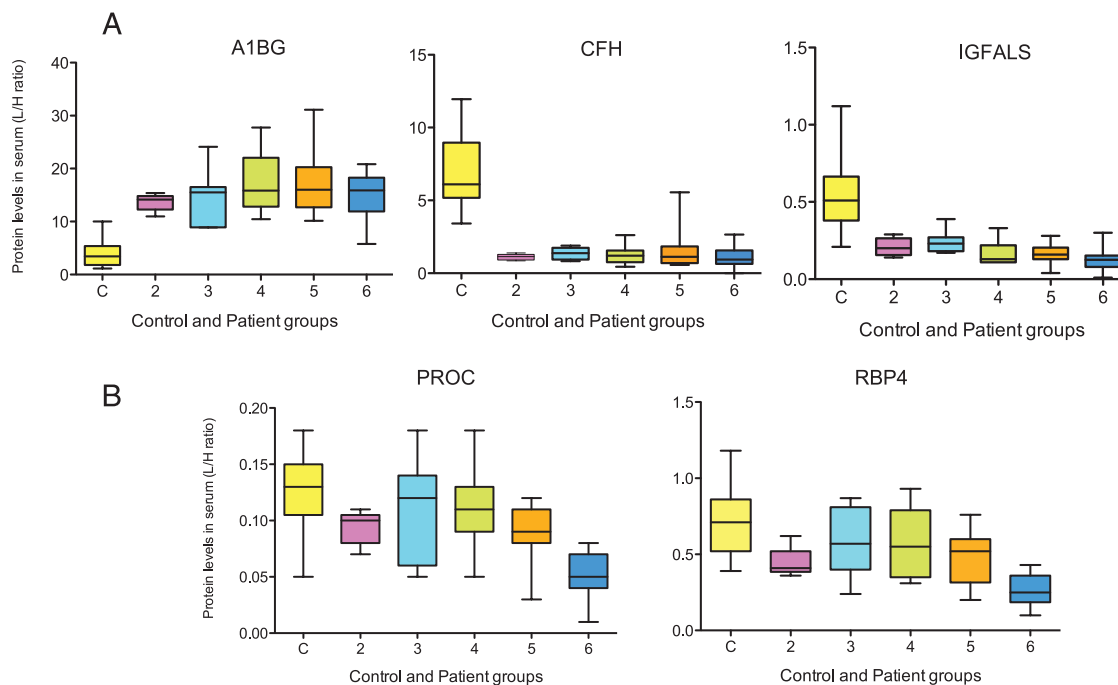


Figure 2. Box plot analysis shows significant differences when comparing serum protein levels of A1BG, CFH, and IGFALS in patients from Ishak fibrosis 2–6 against controls (C). The combined serum A1BG level of all patients is significantly elevated compared to the controls ($p = 7.4E-15$), while the combined serum CFH and IGFALS levels are significantly decreased in patients versus controls ($p = 4.5E-08$ and $5.7E-12$, respectively) (A). PROC and RBP4 levels in serum classify HCV patients in different stages of fibrosis. Box plot analysis shows strong evidence that the median serum levels of PROC and RBP4 in every Ishak fibrosis group decreased with the progression of the disease (B). Each box represents the range between upper (75th) and lower (25th) quartiles with the median level shown as an intersect line. The minimum and maximum values (whiskers) were within 1.5 IQR (the interquartile range).

patients from healthy individuals or distinguish CHC patients at different stages of fibrosis.

3.4.1 A1BG, CFH, and IGFALS can distinguish controls from patients

As shown in Fig. 2A, the average A1BG level in sera of all HCV patients in this study was significantly elevated compared with controls ($p = 7.4E-15$), while CFH and IGFALS levels were significantly decreased in patients versus in controls ($p = 4.5E-08$ and $5.7E-12$, respectively). AUROC scores for controls versus HALT-C patients were 0.99, 0.99, and 0.96 for A1BG, CFH, and IGFALS, respectively.

3.4.2 PROC and RBP4 levels in serum can further classify HCV patients

The concentrations of two proteins—PROC and RBP4—showed good correlation with disease severity. Serum concentrations of PROC and RBP4 decreased as liver disease progressed. Box-and-whisker plots and Student's *t*-test showed that each protein can distinguish different disease stages (Fig. 2B).

With Student's *t*-test, the difference in serum concentrations of PROC or RBP4 between patients with earlier stages of fibrosis (Ishak 2–4) and patients with cirrhosis (Ishak 5–6) is significant. PROC appears to be a good marker to distin-

guish cirrhosis patients from those with fibrosis ($p = 0.004$). RBP4 levels showed a similar decrease from normal control to Ishak 5–6 but the difference between Ishak 2–4 and 5–6 was not statistically significant ($p = 0.07$ with outliers and $p = 0.02$ without outliers) (Supporting Information Fig. S4). AUROC scores for PROC were 0.77 for controls versus patients, 0.75 for Ishak <5 versus Ishak ≥ 5 , and 0.83 for Ishak <5 versus Ishak 6. AUROC scores for RBP4 were 0.79 for controls versus patients, 0.68 for Ishak <5 versus Ishak ≥ 5 , and 0.80 for Ishak <5 versus Ishak 6.

3.5 Multivariate analysis

Multivariate analysis of Ishak 2–4 versus 5–6 patients using PROC and RBP4 proteins, gender, and platelets gave a cross-validated AUROC = 0.89. PROC, RBP4, and platelets distinguish advanced stages of fibrosis patients (Ishak 5–6) from patients in earlier stages with an impressive sensitivity of 95% and specificity of 84% (Fig. 3). A minor improvement (AUROC = 0.72 versus 0.69) was observed with platelets excluded in the analysis based on all available versus only PROC and RBP4 proteins.

Classifying stages of liver fibrosis by biopsy has an accuracy of about 80% [25]. Other studies have suggested that there can be up to a 33% error in the diagnosis of cirrhosis by biopsy [26]. When the value of biomarkers is validated against biopsy, it is improbable to have a discrimination power that exceeds biopsy. In fact, reported noninvasive methods intended for

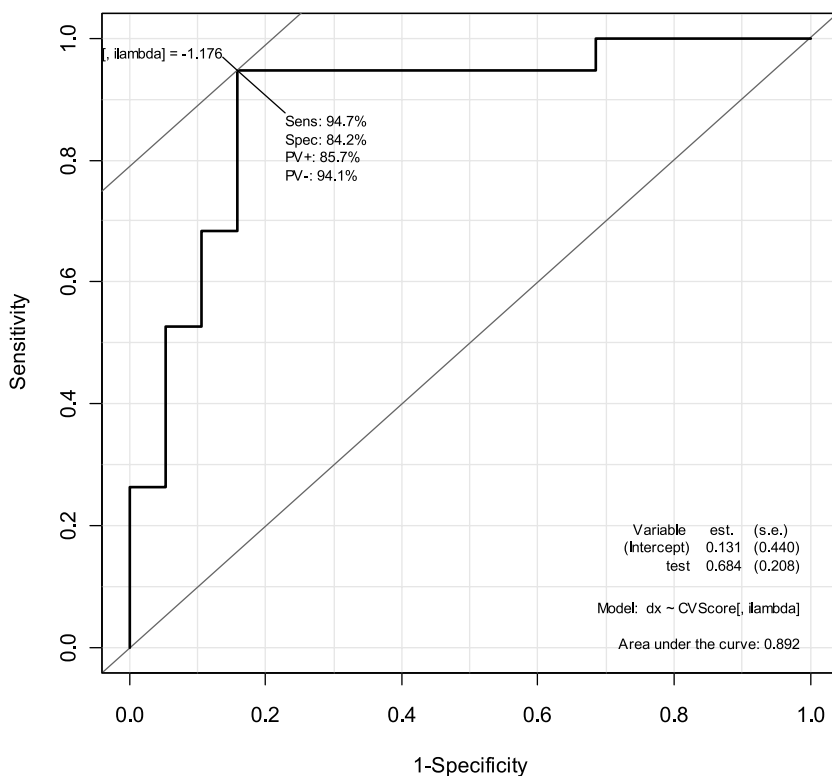


Figure 3. Protein markers PROC and RBP4 and platelets distinguish patients with cirrhosis (Ishak 5–6) from patients in other stages with sensitivity of 95% and specificity of 84%. Inclusion of additional proteins in this study did not result in meaningful further improvement.

discriminating hepatic fibrosis rarely have an AUROC (accuracy) exceeding 0.8–0.9 [27].

4 Discussion

As shown by multivariate analyses, people with the disease and patients at different disease stages may be classified better by employing a combination of new markers with existing markers to increase the sensitivity and specificity. In this study, a three-protein panel of A1BG, CFH, and IGFALS performed well to distinguish healthy controls from HCV-infected patients with fibrosis. An integrated analysis of A1BG, CFH, and IGFALS or A1BG with either of the other two markers resulted in an AUROC score of 1.0 for discriminating healthy controls and HCV-infected patients (Fig. 4A and B). We have observed from this and an independent study of chemical hepatotoxicity that sets of markers for discriminating acute and chronic liver diseases may be very different (unpublished observation). The conventional liver markers AST and ALT may be better markers for acute liver injuries but not as good for chronic liver diseases. This three-protein panel of A1BG, CFH, and IGFALS may potentially prove useful for chronic liver diseases. Apparently, these proteins have not previously been reported as liver disease markers. Evaluation will require sampling a larger patient population. Although sera obtained from ProMedDx and HALT-C trial were prepared under similar conditions, these specimens were ob-

tained from separate sources. Thus, the risk of batch effects generated from sera preparation and storage on the differences for these three proteins between controls and patients cannot be excluded.

PROC and RBP4 have been related to liver diseases. Kloczko et al. reported that the plasma PROC level of patients with liver damage due to chronic alcohol consumption was decreased and correlated with clinical performance of the patients [28]. Our results with different methods confirmed that serum or plasma PROC levels might constitute a useful marker of advanced liver disease. Romanowska et al. reported serum concentrations of RBP-4 were significantly higher in obese children with nonalcoholic fatty liver disease (NAFLD) compared to controls and proposed RBP4 as a serum marker of intrahepatic lipid content in obese children [29]. A contradictory report showed serum RBP4 levels were not different between the steatosis group and controls as well as between subgroups with high and normal ALT, indicating that serum RBP4 may not be a predictive factor in NAFLD [30]. In this study, we showed a significant decrease of serum RBP4 in patients with advanced stages of disease due to HCV infection. Our findings are the first to show PROC and RBP4 may be used for discriminating fibrosis versus cirrhosis in these patients.

Correct classification of different stages of liver fibrosis needs a multiparameter test. Our preliminary results related to staging the progression of fibrosis do not constitute a new biomarker panel, but do provide additional markers

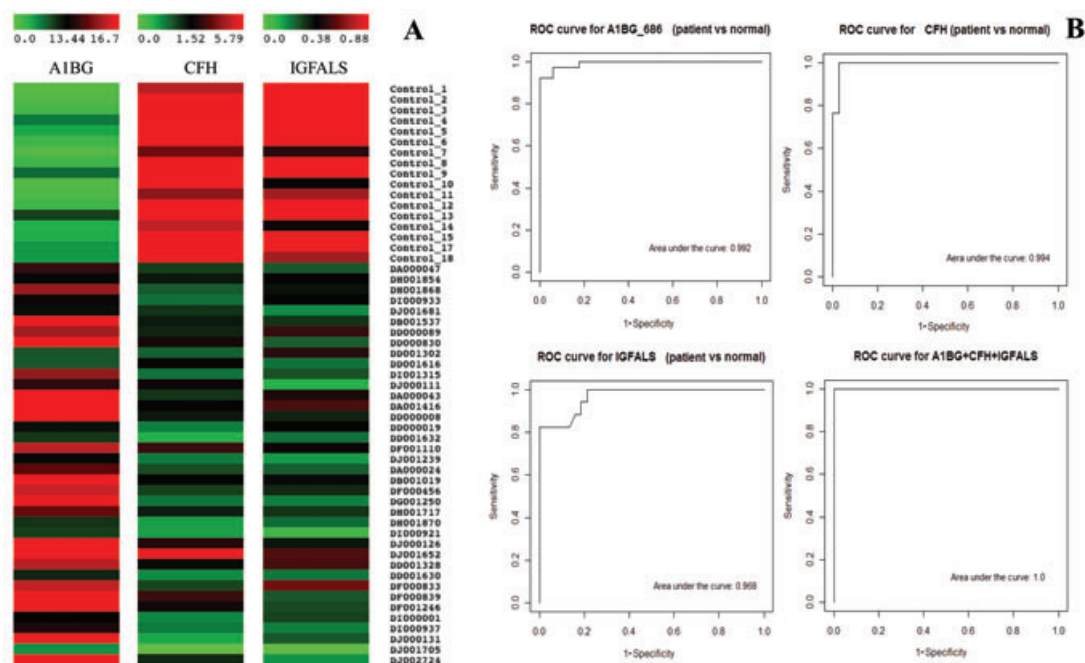


Figure 4. SRM data of serum A1BG and CFH levels in both patients DJ001705 and DJ001652 (arrows) are of good quality; their relative protein levels are probably truly different from other patients. However, this discrepancy can be resolved by examining simultaneously A1BG and CFH or A1BG and IGFALS levels in these patients (A). A combination of A1BG, CFH, and IGFALS is able to discriminate healthy controls from HCV patients from Ishak scores 2 to 6 with predicted high sensitivity and specificity (B).

to strengthen the existing noninvasive methods, including platelet counts. Since there are many other causes of low blood platelet levels, addition of liver-specific proteins to a diagnostic test panel can increase the diagnostic relevance. Moreover, protein analytes can be good targets for developing antibodies or synthetic capture agents that can be integrated into microfluidic chips (Integrated Blood-Barcode Chip) – devices that have the potential to analyze large numbers of patient samples rapidly (in a few minutes), inexpensively, and in a highly multiplexed format (100s or even 1000s of different assays investigating many different diseases) employing blood from a pinprick [31]. Such microfluidic devices are likely to constitute an important foundation for P4 (Predictive, Preventive, Personalized, and Participatory) Medicine with Point-of-Care Diagnosis [32].

In summary, by employing a liver-specific protein strategy and targeted quantitative SRM proteomics technology, we have analyzed 38 liver-specific protein levels in sera of 17 healthy controls and of 38 HCV patients at Ishak fibrosis stages from 2 to 6. In this pilot study, we found that the measured levels of two proteotypic peptides derived from Protein C and retinol-binding protein 4, were present at lower levels in patients than in controls. With AUC statistical analyses, these two proteins represented by their corresponding peptides together appear promising for distinguishing fibrosis versus cirrhosis among HALT-C patients. We also report three proteins, A1BG, CFH, and IGFALS, as measured with their unique peptides, that performed well in distinguishing HCV-infected patients from healthy controls, with an individual AUROC score >0.96 for each marker. These five proteins may contribute to multiparameter panels that will have useful diagnostic potential for CHC and other liver diseases.

The authors would like to thank Drs. Gustavo Glusman and Juan Caballero for their contributions in generating the liver-specific protein list. The authors are grateful to Dr. Kai Wang for valuable discussions throughout this project. This study was partially supported by the Luxembourg Centre for Systems Biomedicine and the University of Luxembourg for Leroy Hood, federal funds from the National Institutes of Health-National Human Genome Research Institute (HG005805 for Robert Moritz) and the NIH National Center for Integrative Biomedical Informatics (U54 DA021519, U54 ES017885 for Gilbert Omenn).

The authors have declared no conflict of interest.

5 References

- [1] Ghany, M. G., Strader, D. B., Thomas, D. L., Seeff, L. B., Diagnosis, management, and treatment of hepatitis C: an update. *Hepatology* 2009, *49*, 1335–1374.
- [2] Regev, A., Berho, M., Jeffers, L. J., Milikowski, C., et al., Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. *Am. J. Gastroenterol.* 2002, *97*, 2614–2618.
- [3] Munteanu, M., Ratziu, V., Morra, R., Messous, D., et al., Non-invasive biomarkers for the screening of fibrosis, steatosis and steatohepatitis in patients with metabolic risk factors: FibroTest-FibroMax experience. *J. Gastrointestin. Liver Dis.* 2008, *17*, 187–191.
- [4] Everson, G. T., Hoefs, J. C., Seeff, L. B., Bonkovsky, H. L., et al., Impact of disease severity on outcome of antiviral therapy for chronic hepatitis C: lessons from the HALT-C trial. *Hepatology* 2006, *44*, 1675–1684.
- [5] Sheth, S. G., Flamm, S. L., Gordon, F. D., Chopra, S., AST/ALT ratio predicts cirrhosis in patients with chronic hepatitis C virus infection. *Am. J. Gastroenterol.* 1998, *93*, 44–48.
- [6] Wai, C. T., Greenson, J. K., Fontana, R. J., Kalbfleisch, J. D., et al., A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology* 2003, *38*, 518–526.
- [7] Lok, A. S., Everhart, J. E., Chung, R. T., Kim, H. Y., et al., Evolution of hepatic steatosis in patients with advanced hepatitis C: results from the hepatitis C antiviral long-term treatment against cirrhosis (HALT-C) trial. *Hepatology* 2009, *49*, 1828–1837.
- [8] Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B., Aebersold, R., Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell* 2009, *138*, 795–806.
- [9] Picotti, P., Rinner, O., Stallmach, R., Dautel, F., et al., High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat. Methods* 2010, *7*, 43–46.
- [10] Ghany, M. G., Kim, H. Y., Stoddard, A., Wright, E. C., et al., Predicting clinical outcomes using baseline and follow-up laboratory data from the hepatitis C long-term treatment against cirrhosis trial. *Hepatology* 2011, *54*, 1527–1537.
- [11] Lin, B., Madan, A., Yoon, J. G., Fang, X., et al., Massively parallel signature sequencing and bioinformatics analysis identifies up-regulation of TGFBI and SOX4 in human glioblastoma. *PLoS One* 2010, *5*, e10210.
- [12] Ge, X., Yamamoto, S., Tsutsumi, S., Midorikawa, Y., et al., Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues. *Genomics* 2005, *86*, 127–141.
- [13] Roth, R. B., Hevezi, P., Lee, J., Willhite, D., et al., Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. *Neurogenetics* 2006, *7*, 67–80.
- [14] Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., et al., A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. USA* 2004, *101*, 6062–6067.
- [15] Lange, V., Picotti, P., Domon, B., Aebersold, R., Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* 2008, *4*, 222.
- [16] Deutsch, E. W., Lam, H., Aebersold, R., PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Rep.* 2008, *9*, 429–434.
- [17] MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., et al., Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 2010, *26*, 966–968.

- [18] Gentleman, R., Irizarry, R. A., Carey, V. J., Dudoit, S., Huber W, *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, Springer, New York 2005.
- [19] Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., et al., Missing value estimation methods for DNA microarrays. *Bioinformatics* 2001, 17, 520–525.
- [20] Zou, H., Hastie, T., Regularization and variable selection via the elastic net. *J. Roy. Stat. Soc. B* 2005, 67, 301–320.
- [21] Rai, A. J., Gelfand, C. A., Haywood, B. C., Warunek, D. J., et al., HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. *Proteomics* 2005, 5, 3262–3277.
- [22] Gahne, B., Juneja, R. K., Stratil, A., Genetic polymorphism of human plasma alpha 1B-glycoprotein: phenotyping by immunoblotting or by a simple method of 2-D electrophoresis. *Hum. Genet.* 1987, 76, 111–115.
- [23] Polanski, M., Anderson, N. L., A list of candidate cancer biomarkers for targeted proteomics. *Biomark Insights* 2007, 1, 1–48.
- [24] Farrah, T., Deutsch, E. W., Omenn, G. S., Campbell, D. S., et al., A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol. Cell. Proteomics* 2011, 10, M110.006353.
- [25] Grigorescu, M., Noninvasive biochemical markers of liver fibrosis. *J. Gastrointestin. Liver Dis.* 2006, 15, 149–159.
- [26] Afdhal, N. H., Nunes, D., Evaluation of liver fibrosis: a concise review. *Am. J. Gastroenterol.* 2004, 99, 1160–1174.
- [27] Ray S, T. D., *Hepatitis C*, Churchill, Livingston, Elsevier., Philadelphia, PA 2009.
- [28] Kloczko, J., Mian, M., Wojtukiewicz, M. Z., Babiuch, L., et al., Plasma protein C as a marker of hepatocellular damage in alcoholic liver disease. *Haemostasis* 1992, 22, 340–344.
- [29] Romanowska, A., Lebensztejn, D. M., Skiba, E., Tarasow, E., Kaczmarek, M., Retinol binding protein-4 as a serum biomarker of intrahepatic lipid content in obese children—preliminary report. *Acta Biochim. Pol.* 2011, 58, 35–38.
- [30] Cengiz, C., Ardicoglu, Y., Bulut, S., Boyacioglu, S., Serum retinol-binding protein 4 in patients with nonalcoholic fatty liver disease: does it have a significant impact on pathogenesis? *Eur. J. Gastroenterol. Hepatol.* 2010, 22, 813–819.
- [31] Fan, R., Vermesh, O., Srivastava, A., Yen, B. K., et al., Integrated barcode chips for rapid, multiplexed analysis of proteins in microliter quantities of blood. *Nat. Biotechnol.* 2008, 26, 1373–1378.
- [32] Hood, L., Friend, S. H., Predictive, personalized, preventive, participatory (P4) cancer medicine. *Nat. Rev. Clin. Oncol.* 2011, 8, 184–187.