Characterization of Serum Proteins Associated with IL28B Genotype among Patients with Chronic Hepatitis C

Derek D. Cyr, Joseph E. Lucas, J. Will Thompson, Keyur Patel, Paul J. Clark, Alexander Thompson, Hans L. Tillmann, John G. McHutchison, M. Arthur Moseley, Jeanette J. McCarthy

1 Institute for Genome Sciences and Policy, Duke University Medical Center, Durham, North Carolina, United States of America, 2 Duke Clinical Research Institute and the Division of Gastroenterology, Duke University Medical Center, Durham, North Carolina, United States of America, 3 Division of Community and Family Medicine, Duke University Medical Center, Durham, North Carolina, United States of America

Abstract

Introduction: Polymorphisms near the IL28B gene (e.g. rs12979860) encoding interferon λ3 have recently been associated with both spontaneous clearance and treatment response to pegIFN/RBV in chronic hepatitis C (CHC) patients. The molecular consequences of this genetic variation are unknown. To gain further insight into IL28B function we assessed the association of rs12979860 with expression of protein quantitative traits (pQTL analysis) generated using open-platform proteomics in serum from patients.

Methods: 41 patients with genotype 1 chronic hepatitis C infection from the Duke Liver Clinic were genotyped for rs12979860. Proteomic profiles were generated by LC-MS/MS analysis following immunodepletion of serum with MARS14 columns and trypsin-digestion. Next, a latent factor model was used to classify peptides into metaproteins based on co-expression and using only those peptides with protein identifications. Metaproteins were then analyzed for association with IL28B genotype using one-way analysis of variance.

Results: There were a total of 4,186 peptides in the data set with positive identifications. These were matched with 253 proteins of which 110 had two or more associated, identified peptides. The IL28B treatment response genotype (rs12979860_CC) was significantly associated with lower serum levels of corticosteroid binding globulin (CBG; p = 9.2 x 10^{-6}), a major transport protein for glucocorticoids and progestins. Moreover, the CBG metaprotein was associated with treatment response (p = 0.0148), but this association was attenuated when both IL28B genotype and CBG were included in the model, suggesting that CBG association may be independent of treatment response.

Conclusions: In this cohort of chronic hepatitis C patients, IL28B polymorphism was associated with serum levels of corticosteroid binding globulin, a major transporter of cortisol, however, CBG does not appear to mediate the association of IL28B with treatment response. Further investigation of this pathway is warranted to determine if it plays a role in other comorbidities of HCV-infection.

Introduction

Hepatitis C virus (HCV) infection is a global health concern with an estimated prevalence of up to 170 million people infected worldwide [1]. Complications from liver disease as a consequence of chronic hepatitis C (CHC) infection is the leading cause for liver transplantation in the United States [2]. Response to current standard of care interferon-based treatment regimens is effective in less than one-half of infected individuals. Recently, polymorphisms in the vicinity of the IL28B gene, encoding interferon λ3, have been associated with both spontaneous clearance and virologic response to pegylated interferon-alpha and Ribavirin therapy (pegIFN/RBV) among patients with CHC [3–7]. The underlying functional genetic variant(s) responsible for this observed association remains to be identified, yet several SNPs, including rs12979860, appear to be robust markers of its effect.

CHC patients who respond to pegIFN/RBV tend to have lower pre-treatment expression levels of interferon stimulated genes (ISGs) in the liver [8]. Honda and colleagues carried out expression quantitative trait locus (eQTL) analysis, examining the IL28B SNP association with liver gene expression levels of ISGs and found that hepatic ISGs were significantly lower in HCV patients with the IL28B ‘responder’ genotype compared to patients without this genotype [9]. These results were independently replicated in another study [10], supporting the idea that a lower endogenous interferon response increases the likelihood of treatment response. However, little else is known about the mechanism through which IL28B influences viral clearance.
We recently assessed a cohort of CHC patients and reported a proteomic signature of HCV treatment response measured in pretreatment serum samples that was highly predictive of response to peg-interferon/Ribavirin [11], with equivalent performance to IL28B genotype. Interestingly, the proteomic signature and IL28B genotype demonstrated only partial correlation, suggesting unique pathways to predict response. Serum proteomic data from that investigation were utilized in the current study to explore the correlation of genetic variation with downstream protein expression. This protein quantitative trait locus (pQTL) analysis builds on the success of QTL analysis, where correlation of genetic variation with downstream mRNA expression has provided insights into the functional consequence of genetic variation. Prior studies employing similar methods have demonstrated the effectiveness of such an approach for identifying genetic variants controlling protein expression in yeast [12]. In humans, large-scale pQTL analysis of proteins measured by standard laboratory assays have also been performed [13]. Both studies confirm the utility of identifying both cis-acting (located close to the gene encoding the protein) and trans-acting (located elsewhere in the genome) loci. In the current study we use an unbiased, label-free proteomic platform to characterize serum proteins associated with an IL28B polymorphism that has been strongly associated with HCV clearance.

Results

The study included 41 genotype 1 CHC patients that were predominantly Caucasian (78%), male (61%), treatment responders (63%) with moderate to advanced fibrosis (85%) and a mean age of 47.4 years. The favorable response IL28B CC genotype was observed in close to one-half of our cohort. Table 1 summarizes this patient cohort.

Metaproteins associated with IL28B genotype

There were a total of 4,186 peptides in the data set with identifications by MS/MS at a 1% false discovery rate. Latent factor modeling grouped the individual peptides into 110 metaproteins. P-values generated from the association analysis of each of the 110 metaproteins and rs12979860 genotype are shown in the quantile-quantile (QQ) plot in Figure 1. A single metaprotein showed significant association with IL28B genotype after Bonferroni correction for multiple hypothesis testing; the liver protein corticosteroid binding globulin (CBG; \( p = 9.2 \times 10^{-6} \)). The rs12979860 'C' (responder) allele was associated with lower levels of CBG metaprotein and explained 46% of the variance in CBG metaprotein (Figure 2). These results remained significant when testing a recessive model (CC versus TC/TT; \( p = 7.45 \times 10^{-4} \)).

If the previously reported association between IL28B and SVR is mediated through CBG, we would expect to see an association between CBG and SVR in this cohort as well. Indeed, we found a statistically significant association between the CBG metaprotein expression and SVR \( (p = 0.0148) \). We also found a statistically significant association \( (p = 0.0021) \) between the IL28B genotype and SVR, as reported previously in patients from the Duke Hepatology Clinic cohort [14]. However, when including both IL28B genotype and CBG in a model to predict SVR, the association with CBG was completely attenuated \( (p = 0.46) \) while the IL28B association remained significant \( (p = 0.0145) \), suggesting that CBG is probably not mediating the association between IL28B and SVR. These results are summarized in Table 2.

The CBG metaprotein included all (5 of 5) of the CBG peptides generated, as well as individual peptides from three other proteins (Figure 3). As expected, the correlation between all five of the CBG peptides is high \( (0.84 \leq r < 0.95) \) and correlations between CBG peptides and the AFAM, HEMO and APOB peptides are also high \( (0.57 \leq r < 0.94) \). Each of these component peptides was tested for association with IL28B genotype to determine whether all or only some of the peptides contributed to the association observed with the metaprotein (Table 3). All eight component peptides were found to be statistically significant, even with correction for multiple hypotheses \( (p \leq 0.006) \).

Discussion

We recently reported the successful application of an unbiased LC-MS/MS platform for proteomic expression signatures associated with HCV treatment response [11]. Using this proteomic data, we were able to demonstrate a relationship between protein expression and genetic variation at the IL28B gene locus, a strong determinant of treatment-induced HCV clearance. Corticosteroid binding globulin peptides were strongly associated with IL28B genotype, accounting for over a third of the variance of the CBG metaprotein and thus, IL28B may be an important trans-pQTL for serum CBG levels among HCV-infected individuals. We were unable to show that CBG mediates the relationship between IL28B and treatment response, indicating possible pleiotropic effects of this locus.

CBG, also known as Transcortin, is a major high-affinity plasma transport protein for glucocorticoids and progestins and is expressed primarily in the liver. Serum cortisol and CBG levels are inversely related [15,16], thus, we would expect rs12979860-CC to associate with not only lower CBG but higher cortisol levels as well. Cortisol modulates the inflammatory response, and the primary function of CBG in proinflammatory states such as CHC infection is to deliver cortisol to sites of inflammation leading to down regulation of neutrophil activity and the cell mediated immune system. Treatment of CHC patients with interferon-beta has been linked to increases in both inflammatory cytokines and cortisol, but these effects appear to be independent [17]. Similarly, administration of interferon-alpha has been shown to alter plasma cortisol levels and immune cytokines, but these appear to be
independent effects [18]. Based on our research, we postulate that the link between interferon and cortisol levels may hold for endogenous interferon lambda as well. Moreover, this relationship is unrelated to treatment response and thus, supports pleiotropic effects of interferons on the steroid hormone system and treatment response.

Levels of serum CBG may be altered in chronic disease states, but are unlikely to confound our results. CBG levels may be

Figure 1. Quantile-Quantile plot showing the distribution of observed P values for IL28B rs12979860 genotype association with 110 metaproteins among chronic HCV patients to the expected distribution of P values under the null hypothesis. The red point denotes the CBG metaprotein. doi:10.1371/journal.pone.0021854.g001

Figure 2. Expression level of CBG metaprotein by IL28B rs12979860 genotype in 41 Chronic HCV patients. Filled circles represent non-responders (NR) to interferon treatment and open circles are from subjects who achieved sustained virologic response (SVR) to treatment. doi:10.1371/journal.pone.0021854.g002
from the associated protein. In the factor and % Signature is the percent of the factor that comes coverage is the percent of identified peptides from that protein that are sections of the protein that are not identified in the data set. % polypeptides from isotope groups that are in the factor and black a representation of the labeled protein. Red sections represent Figure 3. CBG metaprotein peptide composition.

from isotope groups that are in the factor and black a representation of the labeled protein. Red sections represent Figure 3. CBG metaprotein peptide composition. Each gray bar is a representation of the labeled protein. Red sections represent polypeptides from isotope groups that are in the factor and black sections represent polypeptides from the isotope groups that are identified in the data set but are not in the factor. Gray represents sections of the protein that are not identified in the data set. % Coverage is the percent of identified peptides from that protein that are in the factor and % Signature is the percent of the factor that comes from the associated protein.

doi:10.1371/journal.pone.0021854.g003

Serum Proteins Associated with IL28B Genotype

Table 2. Statistical associations.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Additional Covariates</th>
<th>Odds Ratio (95% CI) for Independent Variable</th>
<th>P-value for Independent Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVR</td>
<td>CBG</td>
<td>sex, race</td>
<td>2.62 (0.93, 8.64)</td>
<td>0.0148</td>
</tr>
<tr>
<td>SVR</td>
<td>IL28B</td>
<td>sex, race</td>
<td>0.03 (0.001, 0.19)</td>
<td>0.0021</td>
</tr>
<tr>
<td>SVR</td>
<td>IL28B</td>
<td>sex, race, CBG</td>
<td>0.05 (0.002, 0.39)</td>
<td>0.0145</td>
</tr>
<tr>
<td>CBG</td>
<td>IL28B</td>
<td>sex, race</td>
<td>1.16</td>
<td>7.45 x 10^-5</td>
</tr>
</tbody>
</table>

1For IL28B as the independent variable of interest, we are testing a recessive model, i.e., CC vs. TT/TC.

doi:10.1371/journal.pone.0021854.t002

Despite strong statistical support for the observed association between IL28B and CBG in our study, the results were based on a relatively small number of patients and should thus be considered preliminary. Post-pubertal CBG levels do not vary significantly with age or gender, and there is minimal diurnal variation, thus reducing the potential influence of normal physiological variation on our observations [25]. Moreover, the generalizability of these findings, including whether this is a mechanism specific to infection with HCV, or an association with a chronic inflammatory state is unknown. Replication in larger cohorts will be necessary to validate and extend these findings. Studies that incorporate direct measures of cortisol and CBG may provide further validation of the relationship between IL28B genotype and corticosteroids.

The current study utilized proteomic data generated by LC-MS/MS and analyzed using novel methods to group peptides into ‘metaprotein’ clusters. Thus, the intact proteins were not measured directly, but rather, specific groups of peptides representing the protein were measured. The metaprotein algorithm preferentially, but not exclusively, groups identified peptides according to their assigned “parent” protein. In addition, the algorithm permits exclusion of a peptide from the model if its expression trend does not track with the other peptides, and inclusion of peptides from different proteins or unidentified peptides which have expression in common with the metaprotein. In this way, future work correlating metaproteins with genes identified in pQTL analysis may assist in identification of currently unidentified proteins, or at least functional correlation of these unknown peptides with biological mechanisms.

In summary, the current study used an open-proteomic platform to characterize protein expression associated with a polymorphism strongly associated with HCV clearance and have identified a strong trans-acting pQTL, providing insight into the biological consequences of this variation. We have found the CBG metaprotein and its constituent peptides to be strongly associated with IL28B genotype, but not mediating the effect of IL28B genotype on treatment response. Our research showcases the power of pQTL analysis for the functional annotation of genetic variants.

Materials and Methods

Patient population and ethics statement

Patient samples were selected from the Duke Hepatology Clinical Research (DHCR) database and were included in a prior validation study of proteomic signatures of virologic response to current IFN-based standard-of-care therapy [11]. The cohort included 41 patients with chronic genotype 1 HCV infection who had serum samples available prior to treatment, known treatment response, and DNA available for genetic analysis. All patients provided written informed consent for genetic testing and all study procedures were approved by the Duke University Institutional Review Board.

Figure 3. CBG metaprotein peptide composition. Each gray bar is a representation of the labeled protein. Red sections represent polypeptides from isotope groups that are in the factor and black sections represent polypeptides from the isotope groups that are identified in the data set but are not in the factor. Gray represents sections of the protein that are not identified in the data set. % Coverage is the percent of identified peptides from that protein that are in the factor and % Signature is the percent of the factor that comes from the associated protein.

doi:10.1371/journal.pone.0021854.g003
**Table 3.** Association of component peptides from CBG metaprotein with *IL28B* genotype.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Monoisotopic M/Z</th>
<th>Peak Centroid</th>
<th>Charge State</th>
<th>Peak Time Score</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBG</td>
<td>740.3704</td>
<td>68.17205,</td>
<td>2</td>
<td>0.8491636,</td>
<td>1.46×10⁻⁵</td>
</tr>
<tr>
<td></td>
<td></td>
<td>68.173,</td>
<td></td>
<td>0.8916217,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>68.17549,</td>
<td></td>
<td>0.9133328,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>68.17624,</td>
<td></td>
<td>0.9575863,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>68.19092,</td>
<td></td>
<td>0.9591282,</td>
<td></td>
</tr>
<tr>
<td>CBG</td>
<td>717.356</td>
<td>58.446613,</td>
<td>3</td>
<td>0.91022366,</td>
<td>1.50×10⁻⁴</td>
</tr>
<tr>
<td></td>
<td></td>
<td>58.451298,</td>
<td></td>
<td>0.91063344,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>58.462116,</td>
<td></td>
<td>0.91442996,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>58.475937,</td>
<td></td>
<td>0.918305,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>58.490663,</td>
<td></td>
<td>0.93158853,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>58.50195,</td>
<td></td>
<td>0.92290806,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>58.507416,</td>
<td></td>
<td>0.9367805,</td>
<td></td>
</tr>
<tr>
<td>AFAM</td>
<td>553.2891</td>
<td>32.656315,</td>
<td>1</td>
<td>0.9143975,</td>
<td>1.74×10⁻⁴</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32.71019</td>
<td></td>
<td>0.9606747</td>
<td></td>
</tr>
<tr>
<td>CBG</td>
<td>538.3050</td>
<td>59.005672,</td>
<td>2</td>
<td>0.9508445,</td>
<td>2.42×10⁻⁴</td>
</tr>
<tr>
<td></td>
<td></td>
<td>59.028744</td>
<td></td>
<td>0.9678301</td>
<td></td>
</tr>
<tr>
<td>APOB</td>
<td>461.2762</td>
<td>32.52292,</td>
<td>2</td>
<td>0.8160564,</td>
<td>3.40×10⁻⁵</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32.57379</td>
<td></td>
<td>0.9644853</td>
<td></td>
</tr>
<tr>
<td>CBG</td>
<td>432.76743</td>
<td>32.560402,</td>
<td>1</td>
<td>0.7805042,</td>
<td>5.17×10⁻⁵</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32.605553</td>
<td></td>
<td>0.9047255</td>
<td></td>
</tr>
<tr>
<td>CBG</td>
<td>864.5479</td>
<td>32.484024,</td>
<td>3</td>
<td>0.8152123,</td>
<td>6.36×10⁻⁴</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32.5972</td>
<td></td>
<td>0.9661816</td>
<td></td>
</tr>
<tr>
<td>HEMO</td>
<td>1238.0748</td>
<td>66.4398,</td>
<td>2</td>
<td>0.80079937,</td>
<td>7.37×10⁻⁵</td>
</tr>
<tr>
<td></td>
<td></td>
<td>66.45684</td>
<td></td>
<td>0.80459476</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>66.46535</td>
<td></td>
<td>0.8478435</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>66.47275</td>
<td></td>
<td>0.88555866</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>66.47417</td>
<td></td>
<td>0.8909334</td>
<td></td>
</tr>
</tbody>
</table>

**IL28B polymorphism**

The genomic region associated with HCV response in [4] contains several highly correlated SNPs around the *IL28B* gene. We selected the most strongly associated SNP, rs12979860, located upstream of this gene for genotyping in our cohort using the 5' nuclease assay with allele specific TaqMan probes [26].

**Sample preparation and LC-MS/MS analysis**

Sample preparation and LC-MS/MS analyses on these samples have been described previously [11]. Briefly, serum samples were immunodepleted using MARS14 columns and then subjected to in-solution digestion with trypsin. LC-MS/MS analysis was carried out on a nanoAcquity liquid chromatograph coupled to a QToF Premier mass spectrometer (Waters Corporation, Milford, MA). The Rosetta Elucidator® v3.3 software package (Rosetta Biosoftware, Rosetta Inpharmatics LLC, Seattle, WA) was used to import and align all LC-MS raw data files and perform feature quantitation. Mascot v2.2 (Matrix Sciences, Inc, Boston, MA) and Proteinlynx Global Server v2.4 (Waters Corporation, Milford, MA) database search engines were used to make peptide identifications.

**Metaprotein construction**

Several challenges exist in the analysis of proteomic data generated by LC-MS/MS. Incorrect peptide identifications, sharing of peptides among homologs, and post-translationally modified or processed peptides that may show a biologically relevant and different expression pattern than the proteotypic peptides, can all lead to errors in the quantitative analysis of proteins. In order to correct for these multiple sources of noise, we utilized our metaprotein factor model which groups peptides based on co-expression. We restricted our analysis to only those peptides with protein identifications. Because co-expression across the biological cohort is a central tenant of this modeling approach, peptides are not exclusively tied to their parent protein and therefore individual peptides arising from multiple proteins may potentially comprise a single metaprotein. Metaprotein nomenclature was based on predominance of peptides associated with a known protein. Complete details of how metaprotein expression levels are generated as well as mathematical details regarding the metaprotein factor model are described in Document S1. Protein identification frequency in samples from the 41 CHC patients is listed in Table S1.

**Association analysis**

Each metaprotein was analyzed for association with *IL28B* genotype using one-way analysis of variance (ANOVA). *IL28B* genotypes (TT, CT, CC) were analyzed assuming an additive model, controlling for sex and race. Subsequent analysis was carried out assuming a recessive model for the response allele (C), comparing CC to CT/TT genotypes, consistent with the effect of this locus on treatment response. All statistical analyses have been performed using MATLAB; a commented script (Materials S1) and workspace (Materials S2) have been included as supporting information.

**Acknowledgments**

We gratefully acknowledge Martha Stapels and Scott Geromanos for assistance in LC-MS method development and data collection, Cindy Chepanoske and Andrey Bondarenko for assistance in raw data processing, Diane Uzarski, Crystal Gates and Melissa Spain for assistance with database and biorepository serum sample processing.
Supporting Information

Document S1  Details of metaprotein expression level generation and metaprotein factor model construction. (PDF)

Table S1  Protein identification frequency in samples from 41 CHC patients. This table lists each of the proteins identified in the data set along with the proteins’ full names and the number of different isoforms from that protein that were identified. (DOC)

References


Materials S1  Matlab script. (M)

Materials S2  Matlab workspace. (MAT)

Author Contributions

Conceived and designed the experiments: JJM JGM. Performed the experiments: JWT MAM. Analyzed the data: DDC JEL. Contributed reagents/materials/analysis tools: JGM KP. Wrote the paper: JJM KP PJC AT HLT.