

## Research Article

# Genetic Variation in the *IL-6* and *HLA-DQB1* Genes Is Associated with Spontaneous Clearance of Hepatitis C Virus Infection

Paul Ravi Waldron,<sup>1,2</sup> Ilana Belitskaya-Lévy,<sup>3</sup> Aarthi Chary,<sup>1,2</sup>  
Johann Won,<sup>3</sup> Mark Winters,<sup>1,2</sup> Alexander Monto,<sup>4,5</sup> James Ryan,<sup>4,5</sup>  
Laura C. Lazzeroni,<sup>3,6</sup> and Mark Holodniy<sup>1,2</sup>

<sup>1</sup>VA Palo Alto Health Care System, Palo Alto, CA 94304, USA

<sup>2</sup>Division of Infectious Diseases & Geographic Medicine, Stanford University, Stanford, CA 94305, USA

<sup>3</sup>Cooperative Studies Program Coordinating Center, VA Palo Alto Health Care System, Palo Alto, CA 94304, USA

<sup>4</sup>San Francisco VA Medical Center, San Francisco, CA 94121, USA

<sup>5</sup>University of California, San Francisco, San Francisco, CA 94143, USA

<sup>6</sup>Department of Psychiatry, Stanford University, Stanford, CA 94305, USA

Correspondence should be addressed to Paul Ravi Waldron; [pwaldron@stanford.edu](mailto:pwaldron@stanford.edu)

Received 20 November 2015; Revised 9 March 2016; Accepted 23 March 2016

Academic Editor: Ghislain Opdenakker

Copyright © 2016 Paul Ravi Waldron et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Background.** Millions of people are infected with hepatitis C virus (HCV) worldwide and 30% spontaneously clear the infection. Reasons for HCV clearance without antiviral treatment are not well understood. **Methods.** Blood was collected for DNA analysis from patients with chronic HCV infection or evidence of spontaneous clearance. To overcome anticipated limitations of small sample size, primary analyses consisted of a candidate gene analysis of 12 preselected genes based on known association with host immunologic response to HCV infection. To further reduce the impact of multiple testing on power, a single likelihood ratio test was conducted for each gene using all associated SNPs assayed on the Illumina Quad 610/660W chip. Step-down permutation methods were used to adjust for multiple testing in all analyses. **Results.** Ninety-five and 62 patients with HCV chronic infection or spontaneous clearance, respectively, were included for analysis. *HLA-DQB1* ( $p = 1.76 \times 10^{-5}$ ) and *IL-6* ( $p = 0.0007$ ) genes were significantly associated with spontaneous HCV clearance. *IL-28B* was not significantly associated with spontaneous clearance ( $p = 0.17$ ). **Conclusion.** Our whole-gene analytic strategy identified a previously unreported association of *IL-6* with spontaneous clearance of HCV infection. We also confirmed the finding that *HLA-DQB1* is associated with spontaneous resolution of HCV infection.

## 1. Introduction

The hepatitis C virus (HCV) is a flavivirus which primarily affects the liver, with a seroprevalence of over 185 million infections worldwide [1]. Approximately 70% of people infected with HCV develop chronic infection and 30% are known to spontaneously clear the infection; however these rates vary by ancestry, gender, and other factors [2–4]. Chronic infection with HCV results in hepatic inflammation,

necrofibrosis, and may ultimately cause cirrhosis, hepatocellular carcinoma, and death due to liver failure [3]. Factors that lead to spontaneous clearance may provide important insight into disease pathogenesis and may be important for future assessments of prognosis as well as guiding discovery of novel therapies.

It has been shown that mutations that alter the expression of certain antiviral cytokines are predictive of HCV treatment response and disease prognosis and therefore may also be

instrumental in the biology of spontaneous clearance. Single-nucleotide polymorphisms (SNPs) rs12979860 and rs8099917, which are upstream of *interleukin 28B* (*IL-28B*, also known as interferon- $\lambda$ 3) and in strong linkage disequilibrium (LD) with each other, have been implicated as predictors of treatment response in separate studies [5–7]. These findings were confirmed with candidate gene studies and genotyping currently can be used clinically to advise patients on the likelihood of success prior to interferon-based HCV treatment [5].

Polymorphisms may also affect cell-mediated immune responses in part due to their effect on viral antigen presentation via HLA class II molecules. A specific *HLA* class II allele *DQB\*0301* was found by de Rueda et al. in a candidate gene analysis of samples from 428 treated HCV infected patients to correlate significantly with favorable treatment response via candidate gene analysis [8]. This suggests that *HLA-DQBI* is important for presentation of HCV antigens and resulting clearance of virus.

A recent genome-wide association study (GWAS) on a multicohort sample of 919 HCV spontaneous clearers and 1,482 patients with persistent infection identified genome-wide significant associations with rs12979860 in the promoter region of *IL-28B* and rs4272729 adjacent to *DQB\*0301* [9]. Together these SNPs explained 15% of the variation in HCV resolution in a sample comprising patients with European and African ancestry [9]. While GWAS provides an exhaustive analysis of the association of individual SNPs with a clinical outcome, it requires a large number of patients to empower these studies and minimize type I error [8, 9]. In addition, GWAS may not identify SNPs that may not reach significance individually but may be found to be important in aggregate. Whole-gene analysis is a technique for detecting the aggregate association of SNPs on a given gene with a clinical outcome that addresses these limitations of GWAS. Herein, we describe our whole-gene analysis for selected genes that are likely to be associated with HCV clearance.

## 2. Subjects and Methods

**2.1. Ethics Statement.** The study was approved by both the Stanford University and the University of California at San Francisco Institutional Review Boards and was conducted in accordance with the guidelines established by the Declaration of Helsinki. Written informed consent was obtained from all patients.

**2.2. Subject Recruitment and Enrollment.** One hundred and sixty patients were recruited from 3 sites in the San Francisco Bay Area. Of these, 122 patients were enrolled at both the San Francisco Veterans Affairs Medical Center and the University of California at San Francisco Moffitt-Long Hospital from July 2007 to December 2009 as part of a study of immune responses related to hepatitis C. Patients included in the study were those with confirmed positive HCV antibody and sufficient DNA in extracted samples over six months, although all patients had infection status followed over many years. Another 38 patients were enrolled at the Veteran's Administration Palo Alto Health Care System between July

2007 and September 2011 as part of a cohort enrolled for the purpose of studying HCV infected host transcriptional responses to interferon as well as the host genomics of viral clearance. Blood was drawn at baseline and at 12, 24, 36, and 48 weeks after enrollment; patients from whom sufficient DNA was not obtained were contacted if possible and blood was redrawn.

**2.3. Study Assays and Sample Preparation.** Spontaneous clearance was defined as patients who were HCV antibody positive by screening EIA assay and confirmed by CHIRON® RIBA® HCV 3.0 SIA (RIBA) assay, who had a negative HCV viral load (Abbott Real time HCV assay, Abbott Diagnostics, North Chicago, IL) in plasma without a history of HCV-specific treatment or who had documented acute HCV infection by antibody and/or HCV viral load testing, and whose HCV viral load became undetectable without need for HCV-specific treatment. Chronic HCV infection was defined as those patients with a positive HCV antibody and positive HCV viral load result. HCV genotype was determined using the Versant INNO-LiPA HCV II line probe assay (Siemens Diagnostics, Tarrytown, NY) according to the manufacturer's recommendations. HIV-1 antibody testing was performed using a screening HIV1/HIV2 EIA and confirmatory western blot assay. This study was approved by the Stanford University and the University of California at San Francisco Institutional Review Boards.

Blood samples were collected in cell preparation tubes (CPT) (Becton Dickinson, Franklin Lakes, NJ) and peripheral blood mononuclear cells (PBMC) were separated from whole blood according to the manufacturer's instructions, washed with PBS twice, and stored in freezing medium (10% dimethyl sulfoxide (DMSO)/90% fetal bovine serum (FBS)) or as dry pellets at  $-80^{\circ}\text{C}$ . Cell DNA was purified from PBMC using the Qiagen kit systems (Qiagen, Valencia, CA). DNA purity was checked and any DNA samples with a 260/280 ratio of  $<1.7$  were repurified. DNA amounts were standardized with PicoGreen.

Genotyping was performed using the Illumina Human610-Quad array (samples analyzed prior to 2011) or the Illumina Human660W-Quad array (Illumina, San Diego, CA) on an Illumina BeadStation system, with GoldenGate Genotyping and Infinium Whole-Genome Amplification procedures. Only SNPs evaluated on both chips were included in the final analysis. Each array had several sample-dependent and sample-independent internal controls which were monitored with the BeadStudio Genotyping Module Integrated Controls Dashboard. Initial analyses of data and data quality were accomplished with the BeadStudio software.

**2.4. Candidate Gene Selection.** Due to sample size limitations, we carried out a primary analysis of candidate genes. We performed a detailed literature analysis and identified 67 host genes known or hypothesized to be integral both to the host immune response and to the life cycle of HCV. This included genes involved in cell-to-cell signaling, both innate and adaptive immune responses including interferon effector genes involved with the Janus Kinase-Signal Transducer

TABLE 1: Candidate genes selection.

Gene symbol	Chromosome location	Gene name	Number of SNPs in analysis
IL28B	19q13.13	Interleukin 28B (encoding IFN-gamma-3)	4
SOCS3	17q25.3	Suppressor of cytokine signaling 3	16
SOCS1	16p13.13	Suppressor of cytokine signaling 1	2
IL-10	1q31-q32	Interleukin 10	15
IL-18	11q22.2-q22.3	Interleukin 18	6
TNF-alpha	6p21.3	Tumor necrosis factor alpha	1
TGF-beta-1	19q13.1	Transforming growth factor, beta 1	4
IL-6	7p21	Interleukin 6	42
CTLA-4	2q33	Cytotoxic T-lymphocyte-associated protein 4	7
IFNG	12q14	Interferon, gamma	8
HLA genes:			
HLA-DRB1	6p21.3	Major histocompatibility complex, class II, DR beta 1	5
HLA-DQB1	6p21.3	Major histocompatibility complex, class II, DQ beta 1	40

and Activator of Transcription (JAK-STAT) pathway, major histocompatibility complex mediated antigen presentation, and facilitation of HCV entry into cells. The initial list was further reduced to twelve candidates that were determined to have the most evidence supporting their role in HCV clearance, either spontaneous or treatment-related (Table 1). A single overall test of each candidate gene was based on all available SNPs either within the coding region of the gene of interest or within the region of the next single 3' or 5' flanking/adjacent gene included on the array as defined by the array manufacturer (Illumina) based on the available literature at the time of production.

Rs12979860 was identified in the literature as a SNP of special interest. However, it was only assayed on the Illumina Human610-Quad array. Therefore, for the 32 samples assayed with the Illumina Human660W-Quad array, a separate allelic discrimination RT-PCR assay was performed for rs129798690 as described by Tillmann et al., with results included in our analysis [10].

**2.5. Statistical Analysis.** Primary analyses focused on whole-gene tests of 12 preselected genes of interest as described above to reduce the impact of required multiple testing corrections on power, given the small sample. A genome-wide SNP-based analysis was also performed in the event that single SNPs in genes or regions not previously associated with HCV spontaneous clearance might be detected and serve to inform other candidate whole-gene analyses.

**2.6. Data Preprocessing.** Approximately 610–660 K SNP markers were genotyped in a genome-wide scan. SNPs were excluded from the analysis if they had a call rate of <0.85, a minor allele frequency <0.01, or a Hardy-Weinberg equilibrium  $p$  value < 0.0001, leaving 150 SNPs in the 12 candidate genes and 540,031 additional genome-wide markers (excluding sex chromosomes), which were designated as potential null SNPs to be used to control for population stratification. We used the following “100, 5, 0.5”

criterion for LD pruning to derive null SNPs as follows: (1) consider a window of 100 SNPs; (2) calculate LD between pairs of SNPs in the window; (3) remove one of the pairs of SNPs if the LD is greater than 0.5; and (4) shift the window 5 SNPs forward and repeat the procedure. After linkage disequilibrium pruning, 309,470 null SNPs were identified. These data preprocessing steps were performed using SNP & Variation Suite version 8 (Golden Helix, Bozeman, MT) and PLINK [11].

**2.7. Population Stratification.** To control for population stratification, principal components analysis was used to derive “ancestry eigenvectors” from the set of 309,470 null SNPs. As self-identified race is not a perfect predictor of ancestry group, we used genotype-based principal components to adjust for population stratification. The genomic inflation factor, calculated before and after correction for population structure, was used to evaluate how successful the principal components were in correcting for population structure effects [12].

The eigenvectors were calculated using the Bioconductor R package *snpStats* and the R package *eigen* [13–15]. Significant ancestry eigenvectors were added as covariates in the logistic regression model of HCV clearance.

**2.8. Gene-Based Analysis.** An additive genetic model was used in all analyses. For each of the 12 selected genes, a single likelihood ratio test (LRT) compared two logistic regression models of HCV clearance: the “full” model with all SNPs for that gene as well as the ancestry eigenvectors included as predictor variables and the “null” model which included the ancestry eigenvectors as the only predictors. The whole-gene LRT produced a single  $p$  value for each gene, representing the total collective association of the SNPs in that gene with HCV clearance. Whole-gene tests have the potential to increase power by combining statistical signal across SNPs, while reducing the number of tests for which a multiple testing correction is needed. Many alternative gene-based methods

of association have been proposed, especially for rare variant data [16]. The LRT method was selected to avoid complex or adaptive models that might require a larger sample for robust inference.

**2.9. Step-Down Permutation Procedure to Adjust for Multiple Comparisons.** To account for multiple testing of 12 separate candidate genes, a step-down permutation procedure was implemented as follows [17]. The clinical variables of the 157 patients (i.e., the outcome variable, covariates, such as HIV status, and the ancestry eigenvector) were permuted randomly with respect to their genotype data. The LRT tests above were repeated using the permuted data yielding 12 permutation LRT  $p$  values. These  $p$  values were ordered from smallest to largest and the permutation procedure was repeated 10,000 times to produce a permutation distribution of the smallest  $p$  values, the second smallest  $p$  values, and so on. The fifth percentiles of these distributions provided significance thresholds that were used to assess the significance of the 12 genes based on their observed LRT  $p$  values. The step-down procedure was applied as follows: if a gene with the smallest  $p$  value passed its corresponding significance threshold, it was declared significant and the next gene (with the second smallest  $p$  value) was evaluated. If the second gene passed its significance threshold, then it was declared significant and the next gene was evaluated, and so on. Once any gene did not pass its threshold, it and all the remaining genes were declared to be nonsignificant. The permutation step-down process provides strong control of familywise error rate (i.e., it ensures that the probability of more than zero false positives is 5% or less). Both stratified and random permutations were used and the results were compared. Stratified permutations were performed within the ethnic strata defined by the genotype-based principal components.

**2.10. Univariate Analysis of Candidate SNPs.** Each of the 150 SNPs in the candidate genes was also analyzed in a single-SNP logistic regression model adjusted for population stratification. To control for multiple comparisons, the step-down permutation procedure as the one described above was used.

**2.11. Genome-Wide Scan.** To perform a genome-wide association analysis of SNPs with HCV clearance, a single-SNP logistic regression was performed with HCV clearance as the outcome variable, with and without adjustment for the population genetic structure. Benjamini and Hochberg procedure was used to adjust the raw  $p$  values for multiple comparisons [18].

Most analyses were performed in R [13]. *SNP\_tools* [19] and *Haploview* [20] were used to create LD maps.  $D'$  was used as the LD measure [21].

### 3. Results

Of the 163 patients initially recruited, three patients (1 spontaneous resolution and 2 HCV infected) were excluded from analysis due to insufficient DNA isolated. Three patients

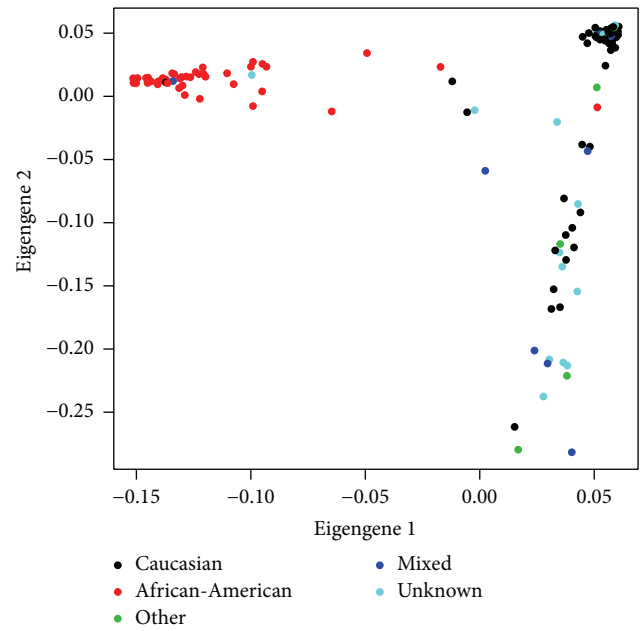


FIGURE 1: Population stratification via null genes.

(1 spontaneous resolution and 2 HCV infected) were excluded from analysis due to gender mismatch. Demographic and clinical characteristics of the 157 study participants for which data could be obtained are presented in Table 2. Ninety-five patients (61%) had chronic infection and 62 (39%) had spontaneous resolution. Between the groups, only self-identified race was significantly different with African-American patients representing 13% of the spontaneous resolution group and 37% of the chronic infection group ( $p = 0.01$ ).

**3.1. Population Stratification.** Shown in Figure 1 is a two-dimensional representation of data on the first two principal components (“ancestry eigenvectors”). The first eigenvector provides a near perfect separation of Caucasian and African-American populations, the two largest groups in the study. When the first four principal components were entered in the logistic regression model of chronic infection versus spontaneous resolution, the first eigenvector was found to be significant ( $p = 0.0026$ ). The genomic inflation factor was estimated to be 1.64 before correction for population genetic structure and 1.05 after correction using the first ancestry eigenvector. The quantile-quantile plots of the null SNP test statistics before and after correction for population stratification are provided in Supplementary Figure 1 in Supplementary Material available online at <http://dx.doi.org/10.1155/2016/6530436>. The first eigenvector was used in all subsequent analyses to adjust for population stratification. Using additional eigenvectors did not alter the results.

**3.2. Gene-Based Analysis.** Twelve candidate genes, including 150 SNP, were selected for primary analysis (Table 1).

The number of SNPs per gene ranged from 1 to 42. The left panel of Table 3 presents the observed LRT  $p$  value for each



TABLE 2: Demographic and clinical characteristics of subjects.

	Total N = 157	Spontaneous resolution N = 62 (39%)	Chronic infection N = 95 (61%)	<i>p</i> value (Chi-squared test unless otherwise noted)
Age, mean (SD) [range], years	55.9 (8.2) [25, 81]	55.7 (8.7) [25, 77]	55.9 (7.8) [35, 81]	0.87 (two-sample <i>t</i> -test)
Gender, number (%)				0.44
Male	132 (84%)	54 (87%)	78 (82%)	
Female	23 (15%)	8 (13%)	15 (16%)	
Trans	2 (1%)	0	2 (2%)	
Ethnicity, number (%)				0.53
Not Hispanic	126 (80%)	48 (77%)	78 (82%)	
Hispanic	20 (13%)	10 (16%)	10 (11%)	
Mixed	6 (4%)	3 (5%)	3 (3%)	
Missing	5 (3%)	1 (2%)	4 (4%)	
Self-identified race, number (%)				0.01
Caucasian	84 (54%)	36 (58%)	48 (51%)	
Black	43 (27%)	8 (13%)	35 (37%)	
Other or mixed	14 (9%)	7 (11%)	7 (7%)	
Missing	16 (10%)	11 (18%)	5 (5%)	
HIV status, number (%)				0.26
Infected	42 (27%)	13 (21%)	29 (31%)	
Not infected	115 (73%)	49 (79%)	66 (69%)	
HCG genotype, number (%)				NA
1	NA	NA	67 (71%)	
2	NA	NA	9 (10%)	
3	NA	NA	11 (12%)	
4	NA	NA	4 (4%)	
Unknown	NA	NA	4 (4%)	
HCV treatment, number (%)				NA
SVR	NA	NA	23 (24%)	
NR	NA	NA	7 (7%)	
Not treated	NA	NA	65 (68%)	

NA: not available.

gene and the corresponding significance threshold used to adjust for multiple testing. The right panel of Table 3 presents results for a parallel analysis with HIV status included as a covariate. The same two genes, *HLA-DQB1* and *IL-6*, reached statistical significance in both analyses. The same results were obtained when stratified permutations were performed where strata, defined by the first ancestry eigenvector, roughly corresponded to Caucasian and African-American populations.

Genotype counts and SNP frequencies for *HLA-DQB1* and *IL-6* are provided in Supplementary Table 1. LD maps for the two significant genes are presented in Figure 2 and for the remaining genes in Supplementary Figure 2. The LD map for *HLA-DQB1* demonstrates a high degree of disequilibrium, which may indicate that a few SNPs in LD with the others may be driving the observed association with the clinical outcome. Less LD is present in *IL-6* which may indicate that several SNPs independent of each other may be driving the association of the gene with the clinical outcome of interest [Figure 2].

**3.3. Single-SNP Analyses.** We also conducted separate single-SNP analyses similar to those above, putting only one SNP at a time in the full model. For the 150 candidate gene SNPs, we adjusted for population stratification and multiple testing of 150 SNPs (Supplementary Table 2). We conducted genome-wide analyses corrected for multiple testing, both with and without adjustment for population stratification (Figure 3). As expected given the small sample size, no SNP reached statistical significance in any of these analyses, given the less powerful SNP-based testing strategy, with its stricter multiple testing penalty.

The smallest raw *p* value of  $1.62 \times 10^{-6}$  (adjusted *p* = 0.328) corresponded to SNP rs2303135 on chromosome 5.

## 4. Discussion

We used a statistical methodology designed for small samples to confirm the finding of Duggal et al. demonstrating a significant association between spontaneous resolution of

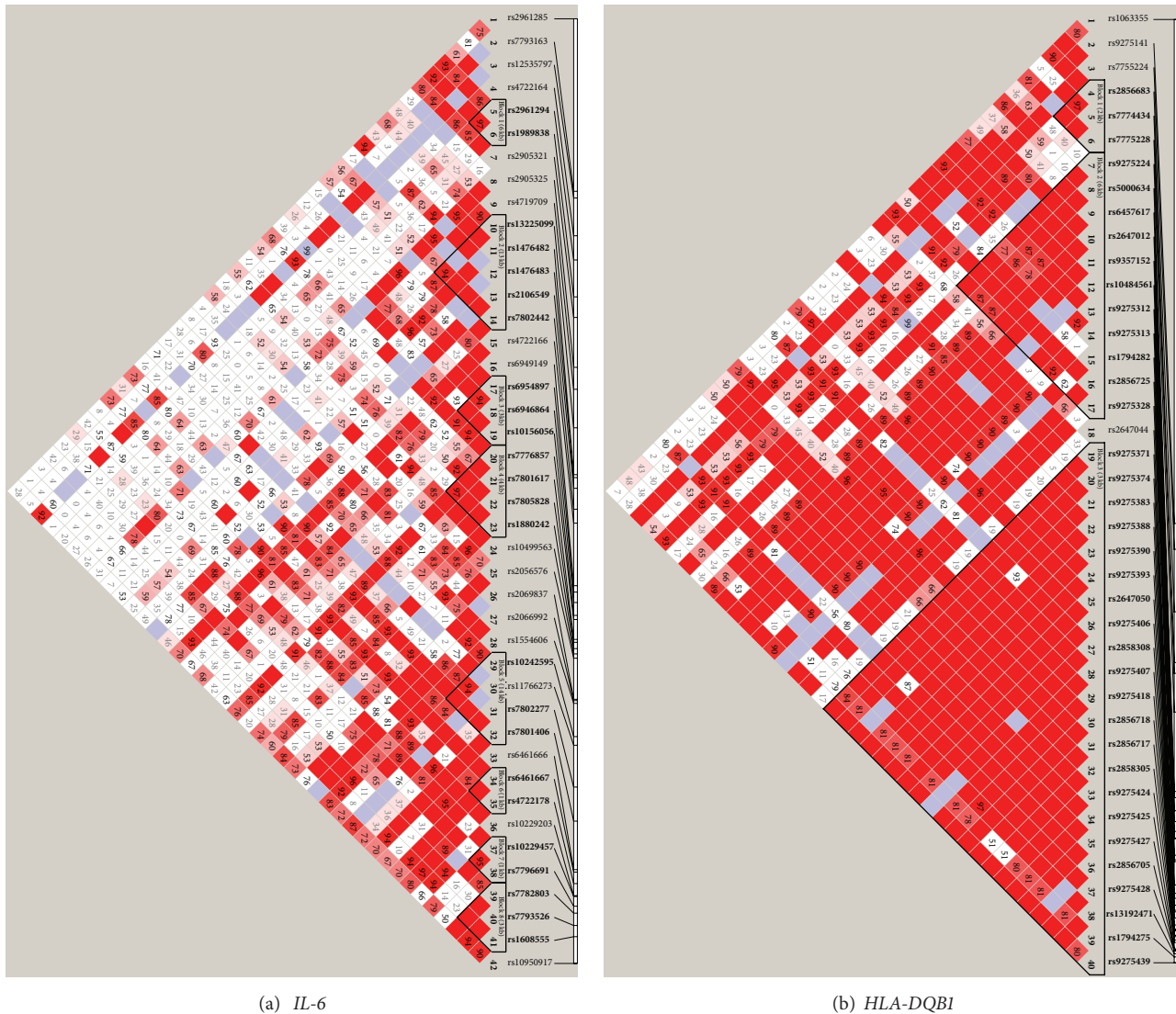


FIGURE 2: Linkage disequilibrium maps of the two significant genes, measured as  $D'$ .

HCV infection and the *HLA-DQB1* region [9]. Their results identified an association between HCV clearance and single SNPs in the gene; our results demonstrate a genetic association that reflects a contribution from one or more SNPs in the gene region. In addition, we identified an association with another gene, encoding the cytokine interleukin 6 (*IL-6*), which is novel [22].

*IL-6* is known as a marker of host inflammatory response and has been thought of as a mediator of innate immunity via its receptor interactions with gp130 and resulting proinflammatory transcriptional activity via both the JAK/STAT pathway and extracellular-signal-regulated kinase/mitogen-activated protein kinase (ERK/MAPK) pathways [23]. More recently, *IL-6* has been shown to be active in cell-mediated immunity as an effector molecule that promotes differentiation of T-helper 17 (Th-17) cells which use interleukin-17 to increase transcription of proinflammatory genes [24]. Several studies have demonstrated that serum *IL-6* levels are

significantly elevated in HCV infected patients as compared to healthy controls [25–28]. Although one study showed significantly lower levels of *IL-6* gene expression, the vast majority of studies demonstrate significantly higher cytokine levels in HCV patients [29]. Relatively higher *IL-6* elevations have been associated with patients with treatment failure versus successful treatment [26, 30].

The *IL-6* –174 G/C SNP (rs1800795) in the *IL-6* promoter region, although not assayed in this study, has been associated with HCV treatment response and differential allele expression among races [31–33]. However, controversy exists regarding whether differential allele expression is associated with favorable treatment response. A possible confounder may be varying prevalence of other neighboring *IL-6* promoter mutations (including rs1800796 and rs1800797), which may explain differences noted between studies to date [34–36]. Other studies suggest that the rs1800795 polymorphism is directly correlated with HCV spontaneous clearance [22].

TABLE 3: Likelihood ratio test (LRT) results comparing full models (consisting of all SNPs for the gene) and null models after controlling for HIV infection.

Not controlling for HIV			Controlling for HIV		
Gene name	LRT $p$ value	5% significance threshold*	Gene name	LRT $p$ value	5% significance threshold*
HLA-DQB1	$1.76 * 10^{-5}$	0.0006	HLA-DQB1	$1.84 * 10^{-5}$	0.0005
IL-6	0.0007	0.0117	IL-6	0.0008	0.0110
IL-28B	0.1700	0.0347	IL28B	0.1600	0.0324
IL-10	0.1987	0.0697	IL-10	0.2603	0.0645
CTLA-4	0.4514	0.1124	CTLA4	0.4305	0.1059
HLA-DRB1	0.4981	0.1661	HLA-DRB1	0.4759	0.1602
SOCS1	0.5313	0.2264	SOCS1	0.5071	0.2195
TNF	0.6957	0.2952	TNF	0.6508	0.2888
IFNG	0.7391	0.3755	SOCS3	0.7475	0.3675
TGF-beta-1	0.7590	0.4677	TGF-beta-1	0.8025	0.4548
SOCS3	0.7742	0.5738	IFNG	0.8027	0.5630
IL-18	0.9567	0.7027	IL-18	0.9649	0.7037

\*5% significance thresholds derived from permutation distributions (unstratified). A gene reaches statistical significance under the step-down procedure when the observed LRT  $p$  values for it and all more significant genes are less than their corresponding 5% thresholds.

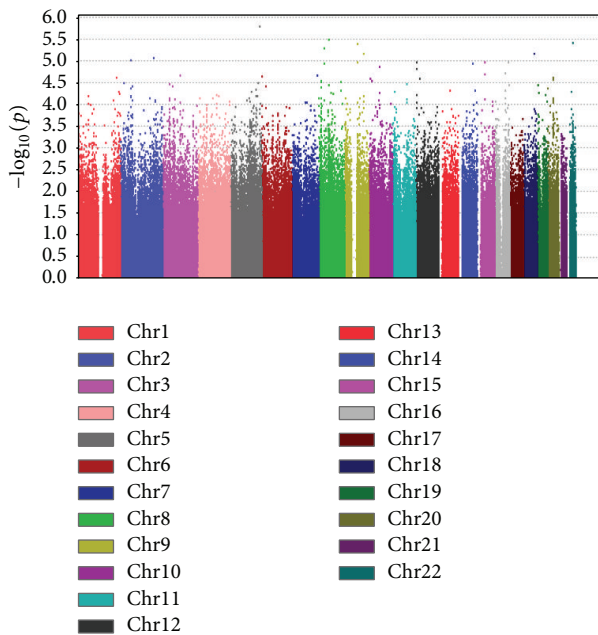


FIGURE 3: Manhattan plot summarizing the genome-wide association results. Smallest  $p = 1.62e - 6$  (adjusted  $p = 0.328$ ).

Another study has demonstrated that this SNP may be important in autoimmune disease pathogenesis [37].

Our whole-gene analysis may have identified a previously undetected association of *IL-6* with spontaneous clearance by testing the total combined effect of multiple SNPs which are not in linkage disequilibrium with each other, each with a relatively small effect, and are therefore undetectable in a single-SNP-based test, especially when subjected to a genome-wide multiple testing correction. Thus, it can be possible by examining the gene as a whole to identify novel genetic associations with a clinical outcome that might elude single-SNP analysis.

Our study is limited by the small sample, but this limitation can be partly overcome by restricting analysis to a preselected set of candidate genes with *a priori* evidence and by testing genes rather than SNPs in order to reduce the number of tests and increase power. Large samples are expensive and sometimes unavailable, if few patients meet appropriate inclusion/exclusion criteria. In fact, it is sometimes impossible to obtain a larger sample except by introducing greater heterogeneity into the study population, which might sometimes reduce power to the point of offsetting any advantage of sample size. The present strategy may also be more successful when a gene has many individual SNPs, each making a small contribution to the phenotype.

Race was significantly different between the chronic infection and spontaneous clearer groups and was missing for 16 patients. The most common reasons for missing data were patients listing race as Hispanic (which was coded as an ethnicity for the purposes of this study) and not defining any other racial category. This may have contributed to the difference as there was more missing data in the spontaneous clearance group. The other reason may be a true reflection of racial differences in spontaneous clearance as prior studies have also shown a higher rate of persistent HCV infection in African ancestry populations [9].

Some SNPs, possibly associated with spontaneous clearance in prior literature, were not included on the chip array. Rs4272729 (in *HLA-DQB1* gene region), rs1800795, rs1800796, and rs1800797 (in *IL-6* gene region) were not included in either the Illumina 610 or 660W Quad chip used to assay sample DNA. Although included in the analysis, *IL-28B* SNPs rs12979860 and rs8099917 were not found to be significantly associated with spontaneous viral clearance. In addition, the *IL-28B* gene as a whole was not found to be significantly associated with spontaneous clearance in this analysis. This may be due to an insufficient number of patients to demonstrate a difference, as well as the small number of SNPs assayed, as the association of SNPs in the *IL-28B* region

with viral clearance is strongly suggested in prior literature [4–7, 9].

Understanding the immunologic host factors that predict clearance of HCV is important in estimating treatment prognosis for different patients. However, it may also be of critical importance in the future for selecting antivirals for treatment, for vaccine development, and for the development of more potent antivirals. Additionally, understanding the genes involved in the clearance of HCV may provide a clearer understanding of the human immune response to viral pathogens and disease. Our whole-gene analytic strategy identified a previously unreported association of IL-6 with spontaneous clearance of HCV infection. We also confirmed an earlier finding that HLA-DQB1 is associated with spontaneous clearance.

## Disclosure

Laura C. Lazzeroni and Mark Holodniy are co-senior authors. The views expressed herein are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the US Government. This study was presented in part at the Joint Statistical Meetings Conference, August 2012, and at the International AIDS Society Conference, July 2015.

## Competing Interests

The authors declare no competing financial interests.

## Authors' Contributions

Paul Ravi Waldron authored the first draft and finalized the paper and was primarily responsible for scientific analysis of the data. Ilana Belitskaya-Lévy contributed significantly to the writing of the paper and was primarily responsible for the statistical analysis. Aarthi Chary contributed to the paper writing, data collection, and scientific analysis of the data. Johann Won contributed to the statistical analysis. Mark Winters performed the sample preparation and the lab assays described. Alexander Monto and James Ryan contributed to data collection and data analysis. Laura C. Lazzeroni contributed to the conception of the study and the statistical analysis. Mark Holodniy contributed to the conception of the study and the scientific analysis. All authors reviewed and approved the first submission and revised the paper. Paul Ravi Waldron and Ilana Belitskaya-Lévy contributed equally to this work.

## Acknowledgments

This work was supported by research funding to Mark Holodniy and the Cooperative Studies Program Coordinating Center, Palo Alto, from the Clinical Science Research & Development Service of the Department of Veterans Affairs (A Cooperative Studies Program-Wide DNA Bank, CSP #478). Additional support was provided by NIH Training Grant T32 AI007502 (Upinder Singh PI).

## References

- [1] K. Mohd Hanafiah, J. Groeger, A. D. Flaxman, and S. T. Wiersma, "Global epidemiology of hepatitis C virus infection: new estimates of age-specific antibody to HCV seroprevalence," *Hepatology*, vol. 57, no. 4, pp. 1333–1342, 2013.
- [2] E. Kenny-Walsh and F. Shanahan, "Clinical outcomes after hepatitis C infection from contaminated anti-D immune globulin," *The New England Journal of Medicine*, vol. 340, no. 16, pp. 1228–1233, 1999.
- [3] D. L. Thomas, J. Astemborski, R. M. Rai et al., "The natural history of Hepatitis C virus infection: host, viral, and environmental factors," *The Journal of the American Medical Association*, vol. 284, no. 4, pp. 450–456, 2000.
- [4] D. L. Thomas, C. L. Thio, M. P. Martin et al., "Genetic variation in IL28B and spontaneous clearance of hepatitis C virus," *Nature*, vol. 461, no. 7265, pp. 798–801, 2009.
- [5] N. H. Afdhal, J. G. McHutchison, S. Zeuzem et al., "Hepatitis C pharmacogenetics: state of the art in 2010," *Hepatology*, vol. 53, no. 1, pp. 336–345, 2011.
- [6] D. Ge, J. Fellay, A. J. Thompson et al., "Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance," *Nature*, vol. 461, no. 7262, pp. 399–401, 2009.
- [7] V. Suppiah, M. Moldovan, G. Ahlenstiel et al., "IL28B is associated with response to chronic hepatitis C interferon- $\alpha$  and ribavirin therapy," *Nature Genetics*, vol. 41, no. 10, pp. 1100–1104, 2009.
- [8] P. M. de Rueda, M.-N. López-Nevot, P. Sáenz-López et al., "Importance of host genetic factors HLA and IL28B as predictors of response to pegylated interferon and ribavirin," *The American Journal of Gastroenterology*, vol. 106, no. 7, pp. 1246–1254, 2011.
- [9] P. Duggal, C. L. Thio, G. L. Wojcik et al., "Genome-wide association study of spontaneous resolution of hepatitis C virus infection: data from multiple cohorts," *Annals of Internal Medicine*, vol. 158, no. 4, pp. 235–245, 2013.
- [10] H. L. Tillmann, A. J. Thompson, K. Patel et al., "A polymorphism near IL28B is associated with spontaneous clearance of acute hepatitis C virus and jaundice," *Gastroenterology*, vol. 139, no. 5, pp. 1586–1592.e1, 2010.
- [11] S. N. Purcell, K. Todd-Brown, L. Thomas et al., "PLINK: a tool set for whole-genome association and population-based linkage analyses," *The American Journal of Human Genetics*, vol. 81, no. 3, pp. 559–575, 2007.
- [12] B. Devlin and K. Roeder, "Genomic control for association studies," *Biometrics*, vol. 55, no. 4, pp. 997–1004, 1999.
- [13] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2016, <https://www.R-project.org/>.
- [14] D. Clayton and H.-T. Leung, "An R package for analysis of whole-genome association studies," *Human Heredity*, vol. 64, no. 1, pp. 45–51, 2007.
- [15] D. Clayton, "snpStats: SnpMatrix and XSNPMatrix classes and methods," in *R Package Version 1.16.0*, 2014.
- [16] L. Moutsianas, V. Agarwala, C. Fuchsberger et al., "The power of gene-based rare variant methods to detect disease-associated variation and test hypotheses about complex disease," *PLoS Genetics*, vol. 11, no. 4, Article ID e1005165, 2015.
- [17] P. H. Westfall and S. S. Young, *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*, Wiley-Interscience, New York, NY, USA, 1993.



- [18] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society*, vol. 57, no. 1, pp. 289–300, 1995.
- [19] B. Chen, S. Wilkening, M. Drechsel, and K. Hemminki, "SNP-tools: a compact tool package for analysis and conversion of genotype data for MS-Excel," *BMC Research Notes*, vol. 2, article 214, 2009.
- [20] J. C. Barrett, B. Fry, J. Maller, and M. J. Daly, "Haploview: analysis and visualization of LD and haplotype maps," *Bioinformatics*, vol. 21, no. 2, pp. 263–265, 2005.
- [21] R. C. Lewontin, "The interaction of selection and linkage. I. General considerations; heterotic models," *Genetics*, vol. 49, no. 1, pp. 49–67, 1964.
- [22] S. Barrett, M. Collins, C. Kenny, E. Ryan, C. O. Keane, and J. Crowe, "Polymorphisms in tumour necrosis factor- $\alpha$ , transforming growth factor- $\beta$ , interleukin-10, interleukin-6, interferon- $\gamma$ , and outcome of hepatitis C virus infection," *Journal of Medical Virology*, vol. 71, no. 2, pp. 212–218, 2003.
- [23] P. C. Heinrich, I. Behrmann, S. Haan, H. M. Hermanns, G. Müller-Newen, and F. Schaper, "Principles of interleukin (IL)-6-type cytokine signalling and its regulation," *The Biochemical Journal*, vol. 374, no. 1, pp. 1–20, 2003.
- [24] E. Bettelli, Y. Carrier, W. Gao et al., "Reciprocal developmental pathways for the generation of pathogenic effector TH17 and regulatory T cells," *Nature*, vol. 441, no. 7090, pp. 235–238, 2006.
- [25] C. P. M. S. Oliveira, C. R. Kappel, E. R. Siqueira et al., "Effects of Hepatitis C virus on cardiovascular risk in infected patients: a comparative study," *International Journal of Cardiology*, vol. 164, no. 2, pp. 221–226, 2013.
- [26] A. Panasiuk, D. Prokopowicz, and B. Panasiuk, "Monocyte chemotactic protein-1 and soluble adhesion molecules as possible prognostic markers of the efficacy of antiviral treatment in chronic hepatitis C," *World Journal of Gastroenterology*, vol. 10, no. 24, pp. 3639–3642, 2004.
- [27] M. G. Neuman, J.-P. Benhamou, I. M. Malkiewicz et al., "Cytokines as predictors for sustained response and as markers for immunomodulation in patients with chronic hepatitis C," *Clinical Biochemistry*, vol. 34, no. 3, pp. 173–182, 2001.
- [28] Y. Zhang, D. Guo, Y. Zhao et al., "The effect of cytokine profiles on the viral response to re-treatment in antiviral-experienced patients with chronic hepatitis C virus infection," *Antiviral Research*, vol. 92, no. 2, pp. 247–254, 2011.
- [29] P. H. McGuinness, D. Painter, S. Davies, and G. W. McCaughan, "Increases in intrahepatic CD68 positive cells, MAC387 positive cells, and proinflammatory cytokines (particularly interleukin 18) in chronic hepatitis C infection," *Gut*, vol. 46, no. 2, pp. 260–269, 2000.
- [30] M. Guzmán-Fulgencio, J. L. Jiménez, J. Berenguer et al., "Plasma IL-6 and IL-9 predict the failure of interferon- $\alpha$  plus ribavirin therapy in HIV/HCV-coinfected patients," *Journal of Antimicrobial Chemotherapy*, vol. 67, no. 5, pp. 1238–1245, 2012.
- [31] A. L. Van Dyke, M. L. Cote, A. S. Wenzlaff, S. Land, and A. G. Schwartz, "Cytokine SNPs: comparison of allele frequencies by race and implications for future studies," *Cytokine*, vol. 46, no. 2, pp. 236–244, 2009.
- [32] A. Meenagh, F. Williams, O. A. Ross et al., "Frequency of cytokine polymorphisms in populations from Western Europe, Africa, Asia, the Middle East and South America," *Human Immunology*, vol. 63, no. 11, pp. 1055–1061, 2002.
- [33] E. D. Cox, S. C. Hoffmann, B. S. DiMercurio et al., "Cytokine polymorphic analyses indicate ethnic differences in the allelic distribution of interleukin-2 and interleukin-6," *Transplantation*, vol. 72, no. 4, pp. 720–726, 2001.
- [34] A. Cussigh, E. Falletti, C. Fabris et al., "Interleukin 6 promoter polymorphisms influence the outcome of chronic hepatitis C," *Immunogenetics*, vol. 63, no. 1, pp. 33–41, 2011.
- [35] L. J. Yee, K. Im, B. Borg, H. Yang, and T. J. Liang, "Interleukin-6 haplotypes and the response to therapy of chronic hepatitis C virus infection," *Genes and Immunity*, vol. 10, no. 4, pp. 365–372, 2009.
- [36] J. Nattermann, M. Vogel, T. Berg et al., "Effect of the interleukin-6 C174G gene polymorphism on treatment of acute and chronic hepatitis C in human immunodeficiency virus coinfecting patients," *Hepatology*, vol. 46, no. 4, pp. 1016–1025, 2007.
- [37] L. Ewald, L.-W. Beate, S. Stephanie, R. Wilfried, and E.-S. Yosuf, "Analysis of a functional IL-6 gene polymorphism in HLAB27 associated and intermediate uveitis gives new insight in disease pathogenesis and commonality with other autoimmune diseases," *Journal of Immunology Research*, vol. 2015, Article ID 174062, 5 pages, 2015.