Variants in *HAVCR1* Gene Region Contribute to Hepatitis C Persistence in African Americans

Genevieve Wojcik,[1] Rachel Latanich,[2] Tim Mosbruger,[2,a] Jacquie Astemborski,[2] Gregory D. Kirk,[1,2] Shruti H. Mehta,[1] James J. Goedert,[3] Arthur Y. Kim,[4] Eric C. Seaberg,[1] Michael Busch,[5,6] David L. Thomas,[1,2] Priya Duggal,[1] and Chloe L. Thio[2]

[1]Bloomberg School of Public Health, Johns Hopkins University, Baltimore, [2]Johns Hopkins University School of Medicine, Baltimore, and [3]Infections and Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland; [4]School of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston; [5]University of California, San Francisco; and [6]Viral Reference Laboratory and Repository Core at the Blood Systems Research Institute, San Francisco, California

To confirm previously identified polymorphisms in *HAVCR1* that were associated with persistent hepatitis C virus (HCV) infection in individuals of African and of European descent, we studied 165 subjects of African descent and 635 subjects of European descent. Because the association was only confirmed in subjects of African descent (rs6880859; odds ratio, 2.42; $P = .01$), we then used 379 subjects of African descent (142 with spontaneous HCV clearance) to fine-map *HAVCR1*. rs111511318 was strongly associated with HCV persistence after adjusting for *IL28B* and *HLA* (adjusted $P = 8.8 \times 10^{-4}$), as was one 81-kb haplotype (adjusted $P = .0006$). The *HAVCR1* genomic region is an independent genetic determinant of HCV persistence in individuals of African descent.

Keywords. human genetics; *Tim 1*; hepatitis C virus.

After an acute hepatitis C virus (HCV) infection, some individuals spontaneously clear the infection; however, the majority remain persistently infected. These dichotomous outcomes are not solely attributable to differences in the infecting inoculum as accidental injection with the same virus results in both outcomes [1]. Thus, host differences are important for determining the outcome of an acute HCV infection. One established host difference is ethnicity, with clearance occurring about 5 times more often in subjects of European descent compared to those of African descent [2]. This observation is partially explained by genetic differences in immune response genes, which has been shown with *IL28B* and *HLA* alleles [3, 4].

Another candidate immune response gene is *HAVCR1,* which was implicated in a previous large-scale candidate gene study of HCV clearance and persistence [5]. *HAVCR1* is a member of the T-cell immunoglobulin and mucin (*TIM*) gene family, which play a role in recognition and phagocytosis of apoptotic cells. In addition, *HAVCR1* acts as a costimulatory molecule leading to enhancement of T-cell proliferation [6]. *HAVCR1* polymorphisms have been associated with various diseases.

In our prior large-scale candidate gene analysis, the single-nucleotide polymorphism (SNP) rs6880859 in *HAVCR1* was associated with HCV persistence in individuals of African descent ($P < .001$) whereas another SNP, rs953569, was associated with persistence in those of European descent ($P = .007$) [5]. Thus, we sought to replicate these associations in an independent population and then more deeply investigate SNPs and haplotypes in the *HAVCR1* gene region to refine the chromosomal location responsible for the association in individuals of African descent.

METHODS

Study Population

To replicate the association of rs6880859 in subjects of African descent, we genotyped this SNP in 165 African American subjects who were not included in the initial study [5]. These subjects came from one of the following cohorts: the Women's Interagency Health Study (WIHS [n = 135]), Correlates of Resolved Versus Low-Level Viremic Hepatitis C Infection in Blood Donors (REVELL [n = 11]), and the Multicenter Hemophilia Cohort Study II (MHCS II [n = 19]), which have been described previously [7]. To replicate the rs953569 association, we used subjects of European descent from the same cohorts (WIHS [n = 47], REVELL [n = 198], and MHCS II [n = 390]).

To fine map the genetic region driving the single SNP association in subjects of African descent, we studied subjects of African descent from a prior HCV genome-wide association study (GWAS) from one of the following cohorts: ALIVE, WIHS, Boston Area HCV Study Transmission, Immunity, Outcomes Network, and REVELL [7]. These subjects came from

Received 19 June 2013; accepted 1 August 2013; electronically published 20 August 2013.
[a]Present affiliation: University of Utah, Department of Oncological Sciences, Salt Lake City.
Correspondence: Chloe L. Thio, MD, Johns Hopkins University School of Medicine, 855 N Wolfe St, Baltimore, MD 21205 (cthio@jhmi.edu).

The Journal of Infectious Diseases 2014;209:355–9

DOI: 10.1093/infdis/jit444

both the candidate gene study [6] and the replication cohort described above. In total, this part of the study examined 379 individuals of African descent, as determined by principal components analysis (PCA) [7].

For all subjects in both the replication and fine mapping parts of this study, individuals with HCV recovery had HCV antibody (anti-HCV) and undetectable HCV RNA without any HCV therapy at 2 time points separated by 6 months. Persistently infected individuals had anti-HCV and HCV RNA for 6 months prior to any HCV therapy.

Informed consent was obtained from all patients, and the institutional review boards at all participating institutions approved the study.

Serologic Testing

All serum or plasma specimens were stored at −70°C. Human immunodeficiency virus type (HIV-1) antibody testing was done by enzyme immunoassay (EIA) with reactive results confirmed by Western blot analysis. Anti-HCV testing was done by Ortho HCV 2.0 or 3.0 EIA (Ortho Diagnostic Systems) or Abbott EIA 2.0 or 3.0 (Abbott Laboratories). HCV RNA was assessed by a branched DNA assay (Quantiplex HCV RNA 2.0 assay, Chiron Corporation), qualitative COBAS AMPLICOR HCV system (Roche Diagnostics), or transcription-mediated amplification (Novartis and Gen-Probe) [7]. All assays were performed according to the manufacturer's specifications.

Genotyping and Imputation

Genotyping of rs6880589 and rs953569 in the replication cohort was performed using a TaqMan SNP genotyping assay, performed according to the manufacturer's instructions (Applied Biosystems). All the subjects in the replication cohort had ancestry informative markers available for PCA.

To finely map the *HAVCR1* gene region in subjects of African descent, genotype data from 156.2–156.6 Mb on chromosome 5 were obtained from the prior GWAS performed with Illumina Human Omni-Quad array, as previously described [7]. To determine probable genotypes for SNPs that were not on the array, imputation was conducted using IMPUTE2 [8]. Imputation allows for in silico fine mapping by increasing the density of genetic markers to find the strongest signal in the region [9]. Probable genotypes are inferred by comparing the underlying linkage disequilibrium (LD) in the reference population (1092 individuals representing 4 continental populations from the 1000 Genomes Project [http://www.1000genomes.org/home]) with known genotypes to the genotypes in the study population.

The imputation protocol was verified by genotyping. The top imputed SNP did not have a predesigned TaqMan assay available; thus, we genotyped a SNP in perfect LD ($D' = 1$ and $r^2 = 1$) (rs17054099) using a predesigned TaqMan genotyping assay (Life Technologies). The genotyped alleles of rs17054099 correlated perfectly with the imputed alleles of the top imputed SNP, thus verifying the imputation protocol.

Statistical Analysis

For the replication study, the associations of rs6880589 and rs953569 were compared between those who cleared HCV infection and those who were persistently infected using logistic regression. Odds ratios (ORs) were adjusted for principal components and HIV status, and were modeled so that an OR >1 reflected an increased likelihood of being persistently HCV-infected.

For the fine-mapping phase of the study, the association of genotyped and imputed SNPs with spontaneous clearance was estimated using SNPTEST [10] with logistic regression using an additive genetic model adjusting for the first 2 principal components and HIV status. The SNPs with $P < .05$ were screened for predicted functional effects using ANNOVAR with SIFT [11] and Polyphen2 [12]. They were also used to estimate haplotype blocks, which are genomic regions determined by recombination patterns, using a solid spline of LD ($r^2 = 0.80$). A χ^2 test of association was performed for each block. For the significant blocks, we permuted the association 100 000 times to determine an empirical P value. For the blocks that remained significant after permutation, PHASE v.2.1.1 [13] was used to estimate individual haplotypes, the phased genotypes within these haplotype blocks. Logistic regression was used to determine the odds of spontaneous HCV clearance for each haplotype vs any other haplotype, adjusting as above.

RESULTS

Subjects of African descent in the replication study included 78 with spontaneous HCV clearance and 87 with persistent infection. The C allele at rs6880589 was present in 13.6% overall and was associated with HCV persistence (OR, 2.42; 95% confidence interval [CI], 1.24–4.73; $P = .01$), similar to the prior candidate gene study, which had an OR of 2.44. In the 287 subjects of European descent with spontaneous HCV clearance and 348 with persistent HCV infection, our prior association of rs953569 with HCV outcomes was not replicated (OR, 0.73; $P = .24$).

Because the findings were only replicated in subjects of African descent, we fine-mapped the region by performing in silico genotyping using data from the prior GWAS from 237 subjects of African descent with persistent HCV and 142 with spontaneous clearance (Supplementary Table 1). A total of 88 SNPs were evaluated in the specified genomic region with the strongest association with SNP rs7706174 (adjusted $P = 2.1 \times 10^{-3}$), which lies within *HAVCR1* (Figure 1). The minor allele frequency (MAF) of this allele was 35% overall and was 28% and 39% when stratified by HCV clearance and persistence, respectively. An additional 1807 SNPs were imputed in this region, and the most strongly associated imputed SNP after

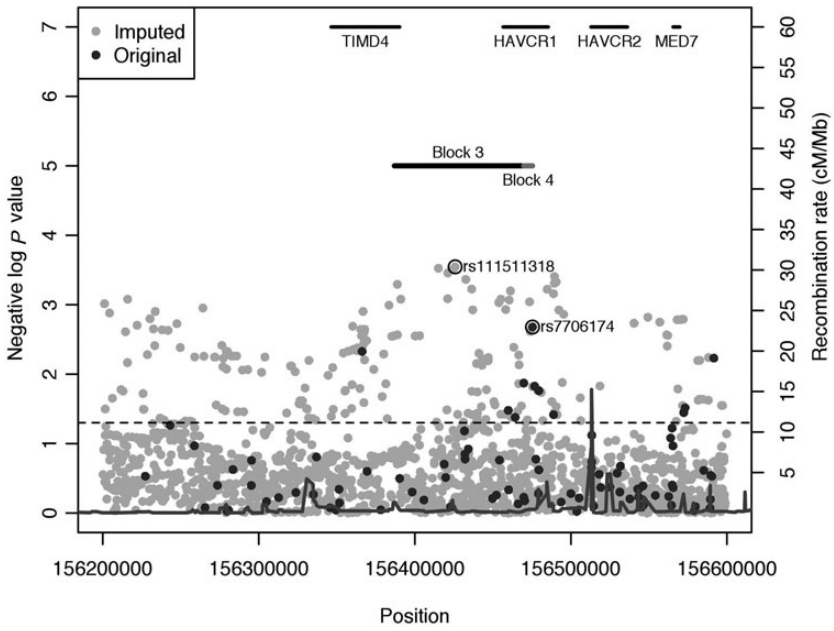

Figure 1. Each point corresponds to a *P* value from a test of association for a genotyped (dark gray) or imputed (light gray) single-nucleotide polymorphism (SNP). The negative $\log_{10}$ *P* values (left y-axis) are plotted by genomic location across the *HAVCR1* gene region (x-axis). The dashed line represents $P = .05$. The imputed and genotyped SNPs with the strongest association are labeled, as are the haplotype blocks. The recombination rate across this area of the genome is demonstrated by the black line with peaks representing a higher recombination rate (right y-axis). The coding regions for the genes in the region are identified at the top of the figure.

adjustment was between *TIMD4* and *HAVCR1* (rs111511318, $P = 2.8 \times 10^{-4}$). The overall MAF of this SNP was 11% and was 6% and 14% when stratified by HCV clearance and persistence, respectively. This association remained ($P = 8.8 \times 10^{-4}$) after adjusting for the top SNPs in the overall GWAS representing the known genetic associations in *IL28B* (rs12979860) and *HLA* (rs4273729).

Haplotype blocks were determined from the 97 genotyped and imputed SNPs with $P < .05$. Of the 7 haplotype blocks, blocks 3 and 4 were the most strongly associated with HCV outcome ($P = 3.0 \times 10^{-4}$ and 6.0×10^{-4}, respectively; Supplementary Table 2). After permuting the case-control status 100 000 times, block 3 had the strongest association (empirical $P = 6.4 \times 10^{-3}$; Figure 1).

From the 33 SNPs within block 3, we estimated 23 distinct haplotypes, of which 11 had a frequency >1% in both the clearance and persistence groups. Of these 11 only haplotype 22, which included the top imputed SNP as well as the original

Table 1. Eleven Haplotypes With Frequency >1% in Block 3 of Subjects of African Descent

Haplotype	No. With Haplotype	No. Without Haplotype	*P* Value	Frequency in Persistent HCV, %	Frequency in HCV Clearance, %	Odds Ratio[a]
22	82	676	.0006	13.9	5.6	2.72
5	48	710	.046	5.1	8.5	0.55
15	22	736	.049	1.9	4.6	0.41
6	32	726	.112	3.2	6.0	0.56
19	11	747	0.198	1.1	2.1	0.45
7	63	695	.247	9.1	7.0	1.40
12	72	686	.291	10.5	7.7	1.34
23	18	740	.547	2.1	2.8	0.74
3	32	726	.672	4.0	4.6	0.85
13	10	748	0.687	1.5	1.1	1.33
9	333	425	.995	43.9	44	1.00

Abbreviation: HCV, hepatitis C virus.

[a] An odds ratio of >1 is associated with persistent HCV infection.

SNP from the candidate gene study, was significantly associated with HCV persistence in the adjusted analysis (OR, 2.72; $P = .0006$; Table 1) even after adjusting for cohort (data not shown). As expected, the haplotype with the nearly complementary SNPs to haplotype 22 (haplotype 5) was associated with increased likelihood for spontaneous HCV clearance (OR, 0.55; $P = .05$; Table 1). We combined the 38 SNPs in blocks 3 and 4, and estimated 29 haplotypes of which 12 had a frequency >1%. Analysis of these haplotypes was identical to that of block 3 by itself. Adjustment for cohort and mode of transmission did not alter the results.

The 97 genotyped and imputed SNPs with a $P < .05$ in the *HAVCR1* gene region were screened for functional prediction. Only 1 SNP was exonic, but this SNP was categorized as benign (rs12522248, $P = .02$), and thus is likely not functional.

DISCUSSION

This study confirms a previous association of the *HAVCR1* gene region with HCV persistence in individuals of African descent but not in those of European descent. Fine mapping in subjects of African descent identified a SNP (rs111511318) with the strongest association independent of signals from *IL28B* and *HLA*. We also identified 1 haplotype with the strongest association with HCV persistence in individuals of African descent. Although these data confirm that the *HAVCR1* region is a determinant of HCV persistence in individuals of African descent, further work is needed to determine the functional variant within our identified haplotype.

The most strongly associated SNP in the *HAVCR1* gene region, rs111511318, was present in 11% of the individuals of African descent in this study. Interestingly, this SNP is not polymorphic in individuals of European ancestry (www.1000genomes.org), which may explain why the *HAVCR1* association could not be replicated in subjects of European descent. Due to the long-range LD in this region, we were unable to isolate a single causal SNP. However, we did narrow the region of interest to an 81-kb haplotype extending from *TIMD4* to *HAVCR1* that likely contains the causal genetic variant. Data from the ENCODE project demonstrate that this region of interest (block 3) contains several potential areas of important regulatory elements, which may regulate HCV clearance (Supplementary Figure 1).

A Spanish group genotyped the single SNP (rs953569) in *HAVCR1* that gave a potential signal in European-Americans in our prior candidate gene study, but they did not confirm a role for this SNP with HCV persistence, [14] which is consistent with our findings.

It is intriguing to consider why this gene region is only important in individuals of African descent. The *TIM* gene family is a regulator of the immune response and the balance between the Th1 and Th2 response, so it is likely to be involved in the response to acute HCV in all ethnic groups. However, the associated haplotype only exists in those of African descent. Furthermore, there is greater heterogeneity of this region in African Americans than in European Americans, suggesting that there was more selection pressure from another disease leading to faster evolution. Thus, it is possible that the variants that are unique or more common in persons of African descent may have offered protection from this other disease but are detrimental to spontaneous HCV clearance.

The major strengths of this study are the replication of the prior association of the *HAVCR1* genomic region in an independent cohort, the large number of subjects of African descent, and the comprehensive study of SNPs in the region. The major limitation of this study is the large LD block in this region, which is a barrier to identifying a single, causal SNP.

In summary, this study demonstrates that the *HAVCR1* gene region is a determinant of HCV clearance in persons of African descent. Further work is needed to determine the precise variant responsible for this association, which will further our understanding of differences in HCV outcomes between individuals of African and European descent.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online (http://jid.oxfordjournals.org/). Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

Notes

Acknowledgments. Some of the data in this manuscript were collected by the WIHS Collaborative Study Group with centers at the New York City/Bronx Consortium (principal investigator [PI], Kathryn Anastos); Brooklyn, New York (PI, Howard Minkoff); Washington, DC, Metropolitan Consortium (PI, Mary Young); The Connie Wofsy Study Consortium of Northern California (PI, Ruth Greenblatt); Los Angeles County/Southern California Consortium (PI, Alexandra Levine); Chicago Consortium (PI, Mardge Cohen); Data Coordinating Center (PI, Stephen Gange). The authors thank Leslie Tobler for her contributions to recruiting subjects for the REVELL cohort.

Financial support. This project was supported by the National Institutes of Health (NIH; grant number R01 DA013324 to D. L. T.). REVELL is funded by the NIH (grant number RO1-HL-076902). ALIVE is funded by the NIH (grant numbers DA04334 to G. D. K., DA12568 to S. H. M.). The WIHS is funded by the NIH (grant numbers UO1-AI-35004, UO1-AI-31834, UO1-AI-34994, UO1-AI-34989, UO1-AI-34993, and UO1-AI-42590) and by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (grant number UO1-HD-32632). WIHS is cofunded by the National Cancer Institute, the National Institute on Drug Abuse, the National Institute on Deafness and Other Communication Disorders, and the National Center for Research Resources (UCSF-CTSI grant number UL1 RR024131). MHCS II is supported in part by the Intramural Research Program of the National Cancer Institute. A. Y. K. is funded by the NIH (grant numbers U19-AI-066345, R01-DA031056, and R01-DA-033541).

Disclaimer. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Kenny-Walsh E. Clinical outcomes after hepatitis C infection from contaminated anti-D immune globulin. Irish Hepatology Research Group. N Engl J Med **1999**; 340:1228–33.
2. Thomas DL, Astemborski J, Rai RM, et al. The natural history of hepatitis C virus infection: host, viral, and environmental factors. JAMA **2000**; 284:450–6.
3. Thomas DL, Thio CL, Martin MP, et al. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. Nature **2009**; 461:798–801.
4. Thio CL, Thomas DL, Goedert JJ, et al. Racial differences in HLA class II associations with hepatitis C virus outcomes. J Infect Dis **2001**; 184: 16–21.
5. Mosbruger TL, Duggal P, Goedert JJ, et al. Large-scale candidate gene analysis of spontaneous clearance of hepatitis C virus. J Infect Dis **2010**; 201:1371–80.
6. Umetsu SE, Lee WL, McIntire JJ, et al. TIM-1 induces T cell activation and inhibits the development of peripheral tolerance. Nat Immunol **2005**; 6:447–54.
7. Duggal P, Thio CL, Wojcik GL, et al. Genome-wide association study of spontaneous resolution of hepatitis C virus infection: data from multiple cohorts. Ann Intern Med **2013**; 158:235–45.
8. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet **2009**; 5:e1000529.
9. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet **2010**; 11:499–511.
10. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet **2007**; 39:906–13.
11. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res **2010**; 38:e164.
12. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods **2010**; 7:248–9.
13. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet **2001**; 68:978–89.
14. Abad-Molina C, Garcia-Lozano JR, Montes-Cano MA, et al. HAVCR1 gene haplotypes and infection by different viral hepatitis C virus genotypes. Clin Vaccine Immunol **2012**; 19:223–7.