

Genetic variation in *IL28B* predicts hepatitis C treatment-induced viral clearance

Dongliang Ge¹, Jacques Fellay¹, Alexander J. Thompson², Jason S. Simon³, Kevin V. Shianna¹, Thomas J. Urban¹, Erin L. Heinzen¹, Ping Qiu³, Arthur H. Bertelsen³, Andrew J. Muir², Mark Sulkowski⁴, John G. McHutchison² & David B. Goldstein¹

Chronic infection with hepatitis C virus (HCV) affects 170 million people worldwide and is the leading cause of cirrhosis in North America¹. Although the recommended treatment for chronic infection involves a 48-week course of peginterferon- α -2b (PegIFN- α -2b) or - α -2a (PegIFN- α -2a) combined with ribavirin (RBV), it is well known that many patients will not be cured by treatment, and that patients of European ancestry have a significantly higher probability of being cured than patients of African ancestry. In addition to limited efficacy, treatment is often poorly tolerated because of side effects that prevent some patients from completing therapy. For these reasons, identification of the determinants of response to treatment is a high priority. Here we report that a genetic polymorphism near the *IL28B* gene, encoding interferon- λ -3 (IFN- λ -3), is associated with an approximately twofold change in response to treatment, both among patients of European ancestry ($P = 1.06 \times 10^{-25}$) and African-Americans ($P = 2.06 \times 10^{-3}$). Because the genotype leading to better response is in substantially greater frequency in European than African populations, this genetic polymorphism also explains approximately half of the difference in response rates between African-Americans and patients of European ancestry.

To identify human genetic contributions to anti-HCV treatment response, we have performed a genome-wide association study of more than 1,600 individuals who were part of the IDEAL study², and we included a further 67 patients from another prospective treatment study³. Briefly, the IDEAL study compared the effectiveness of three treatment regimens involving PegIFN- α -2b or PegIFN- α -2a combined with RBV. It demonstrated similar efficacy of the two IFN preparations and a significantly lower efficacy in self-reported African-Americans compared with Americans of European ancestry (European-Americans). All patients included were treatment-naïve Americans who were chronically infected with genotype 1 HCV. Patients received 48 weeks of treatment and 24 weeks of follow-up. A total of 1,671 individuals were genotyped using the Illumina Human610-quad BeadChip, and we then searched for determinants of treatment response as a primary endpoint. We defined successful treatment response and non-response according to standard definitions⁴, concentrating on sustained virological response (SVR), which is the absence of detectable virus at the end of follow-up evaluation (Supplementary Information I). We included 1,137 patients who satisfied stringent compliance criteria (Supplementary Information I) in the analyses of treatment response, and 1,475 patients in a separate analysis of baseline viral load.

We found that a polymorphism on chromosome 19, rs12979860, is strongly associated with SVR in all patient groups (Fig. 1), with the

European-American population sample showing overwhelming genome-wide significance ($P = 1.06 \times 10^{-25}$). Combining the P values across the population groups, the variant shows association at 1.37×10^{-28} . The polymorphism resides 3 kilobases (kb) upstream of the *IL28B* gene (Fig. 2), encoding IFN- λ -3.

In patients of European ancestry, the CC genotype is associated with a twofold (95% confidence interval 1.8–2.3) greater rate of SVR than the TT genotype (Fig. 1), with similar ratios in both the African-American (threefold, 95% confidence interval 1.9–4.7) and the Hispanic (twofold, 95% confidence interval 1.4–3.2) population groups. The magnitude of this association is compared in Table 1 with other host or viral factors known to influence SVR in patients infected with genotype 1 HCV, including baseline viral load, fibrosis and ethnicity^{4,5}. Not only does the *IL28B* polymorphism strongly influence response within each of the major population groups, it also appears to explain much of the difference in response between different population groups (European-Americans compared with African-Americans). We estimate that approximately half of the difference in SVR between populations can be accounted for by the difference in frequency of the C allele between African-Americans and individuals of European ancestry (Supplementary Information XI). Interestingly, it has also been well documented that East Asians have

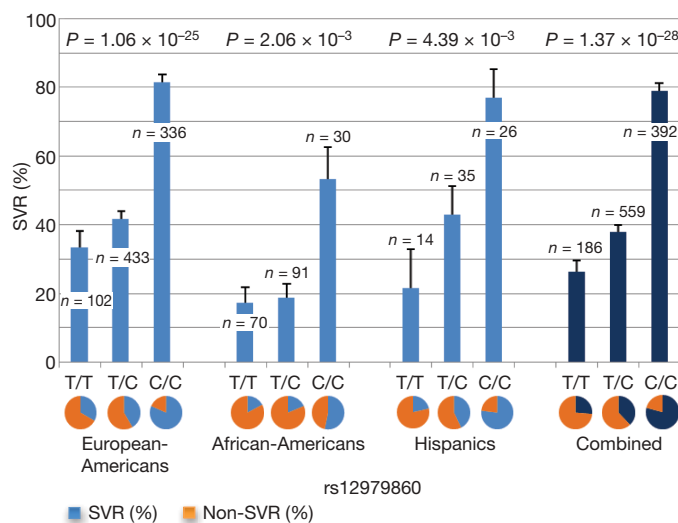


Figure 1 | Percentage of SVR by genotypes of rs12979860. Data are percentages + s.e.m.

¹Institute for Genome Sciences & Policy, Center for Human Genome Variation, Duke University, Durham, North Carolina 27708, USA. ²Duke Clinical Research Institute and Division of Gastroenterology, School of Medicine, Duke University, Durham, North Carolina 27705, USA. ³Schering-Plough Research Institute, Kenilworth, New Jersey 07033, USA. ⁴Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA.

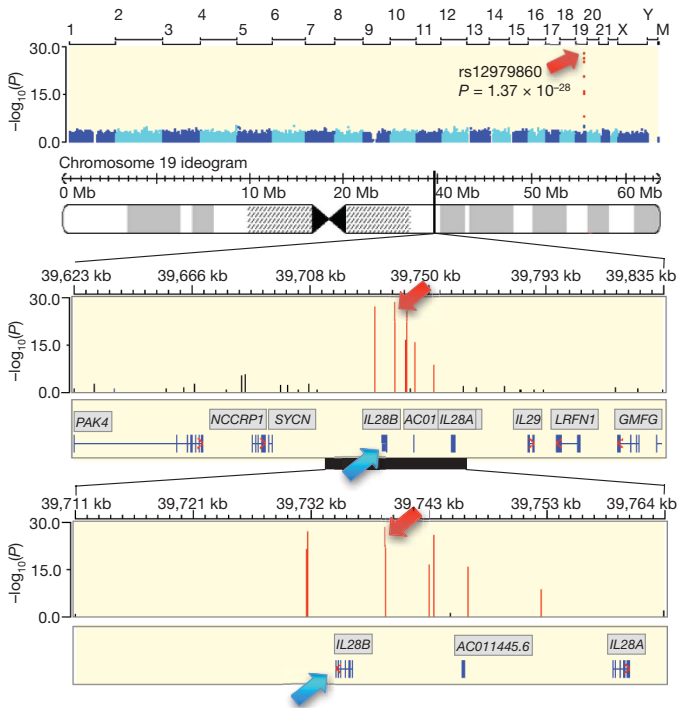


Figure 2 | Genomic overview of the region of 19q13.13 surrounding the genome-wide significant determinant of response to treatment and including the *IL28B* gene. The top panel shows a genome-wide view of the P values [$-\log_{10}(P)$]. Panels below show all genotyped SNPs in the region of significance and the structures of the surrounding genes. The SNPs that show genome-wide significant association with SVR are marked in red. The polymorphism rs12979860 (red arrow) is 3 kb upstream to the gene encoding IFN- λ -3 (*IL28B*, blue arrow). Other SNPs in the same region showing genome-wide significant P values largely reflect the same signal (Supplementary Information IX). The results were annotated using the WGAViewer software¹⁹.

higher SVR rates than patients of European ancestry^{6,7}. By looking at a random multi-ethnic population sample with unknown hepatitis C status (Supplementary Information II), we observed a substantially higher frequency of the C allele in East Asians (Fig. 3). Collectively, the SVR rates across different population groups displayed a striking concordance with C-allele frequency (Fig. 3). Finally, it is also noteworthy that African-Americans with the CC genotype have a significantly higher rate of response (53.3%) than individuals of European ancestry who have the TT genotype (33.3%, $P < 0.05$), which emphasizes the greater importance of individual genotype compared with ethnicity in predicting treatment response⁸.

We next tested whether this variant influences baseline (pre-treatment) viral load and found a significant association in all groups (Supplementary Information XIII). Interestingly, the C allele, associated with better treatment response, is also associated with higher baseline viral load (CC 6.35, $n = 485$; TC 6.33, $n = 744$; TT 6.16, $n = 246$; $P = 1.21 \times 10^{-10}$; viral loads given as \log_{10} international

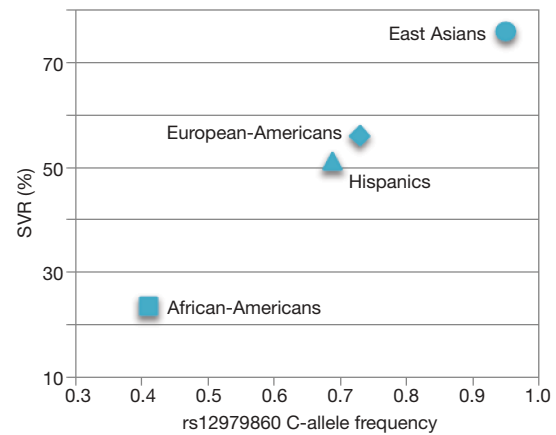


Figure 3 | Rate of SVR and rs12979860 C-allele frequency in diverse ethnic groups. The SVR rate in East Asians is adopted from Liu *et al.*⁷. Sample sizes for C-allele frequency: $n = 61$ (African-Americans); $n = 271$ (European-Americans); $n = 16$ (Hispanics); $n = 107$ (East Asians); sample sizes for SVR rate: $n = 191$ (African-Americans); $n = 871$ (European-Americans); $n = 75$ (Hispanics); $n = 154$ (East Asians).

units (IU ml^{-1}). Although this finding is counter-intuitive in that lower baseline viral loads predict a better response to treatment, it could relate to recent speculation about the role of IFN-stimulated genes in modulating response to PegIFN⁹, and it seems plausible that the *IL28B* polymorphism has a role in the regulation of intra-hepatic IFN-stimulated gene expression with consequences both for viral load and treatment response (Supplementary Information XIII). We also note that the polymorphism has no association with whether individuals are classified as having baseline viral loads above or below a commonly used threshold that predicts respectively worse or better treatment response (Supplementary Information XIII), indicating that the association of the polymorphism with clearance and viral load may be independent. In addition, we note that the C-allele frequency was significantly reduced in the chronically infected cohort compared with ethnically matched controls (0.63 versus 0.73, controlled for population structure, $P < 2.5 \times 10^{-6}$, Supplementary Information II and XVI), which suggests an association between the C allele and a higher rate of natural clearance of hepatitis C. We note, however, that determination of the precise effect of the C allele on clearance will require comparison between matched cohorts known to have and have not naturally cleared this viral infection.

We sequenced the *IL28B* gene in 96 individuals, and found two variants highly associated with rs12979860 ($r^2 > 0.85$ for all comparisons in all populations): a G > C transition 37 base pairs (bp) upstream of the translation initiation codon (rs28416813), and a non-synonymous coding single nucleotide polymorphism (SNP) (rs8103142) encoding the amino-acid substitution Lys70Arg. These new variants were then genotyped in the full cohort. Owing to the high degree of correlation among the three SNPs, tests for independence among these variants, using all available patients, were not able

Table 1 | Comparison between the genetic and conventional clinical factors associating with SVR

| | Odds ratio (95% confidence interval) | | |
|--|--------------------------------------|-------------------|----------------|
| | European-Americans | African-Americans | Hispanics |
| <i>IL28B</i> rs12979860 genotype CC (versus CT and TT)* | 7.3 (5.1–10.4) | 6.1 (2.3–15.9) | 5.6 (1.4–22.1) |
| Baseline viral load (<600,000 IU ml^{-1} versus $\geq 600,000$ IU ml^{-1})† | 4.2 (2.6–6.6) | 5.1 (1.9–13.9) | 2.4 (0.7–8.8) |
| Baseline fibrosis (METAVIR F0–2 versus F3–4)‡ | 3.0 (1.8–5.1) | 1.1 (0.3–5.2) | 4.1 (0.7–25.5) |
| Ethnicity (European-Americans/African-Americans) | 3.1 (2.1–4.7) | | |

Odds ratios and 95% confidence intervals are generated from the logistic regression model.

* Corresponding relative risks for rs12979860: 2.0 (95% confidence interval 1.8–2.3) in European-Americans; 3.0 (95% confidence interval 1.9–4.7) in African-Americans; 2.1 (95% confidence interval 1.4–3.2) in Hispanics.

† In clinical practice it is customary to divide patients into high and low viral-load groups, reflecting a well-described threshold effect. The IDEAL trial used a threshold of 600,000 (ref. 2).

‡ Fibrosis was scored by METAVIR stage on a baseline centrally evaluated liver biopsy^{2,18}.

to resolve which, if any, of these sites is uniquely responsible for the association with SVR. Additional HCV-infected cohorts may help to determine whether one of these SNPs, or any other SNP in the region, is causal for the association, as the pattern of association suggests the possibility of more than one functional variant in the region (Supplementary Information IX). Ultimately, identification and elucidation of the functional SNPs will depend on in-depth functional studies.

Given the significant effect of the *IL28B* polymorphism on treatment response, and its likely clinical relevance, it was considered important to compare the magnitude of different predictors of response for the patients studied here. We developed a logistic regression model that related clinical predictors to response rates (Supplementary Information XI). We noted that the regression model showed that the CC genotype is associated with a more substantial difference in rate of response than other known baseline predictors included in the model.

It seems likely therefore that advance knowledge of host genotype of patients infected with HCV could in the future become an important component of the clinical decision to initiate treatment with PegIFN and RBV. Many important clinical questions remain. The current data are specific to patients with genotype 1 infection. It will therefore be necessary to evaluate the role of host *IL28B* genotype and treatment response in other less common HCV genotypes. Novel small molecules, including HCV protease inhibitors, are currently being developed and may soon be used in combination with PegIFN and RBV for the treatment of genotype 1 HCV¹⁰; the role of the *IL28B* genotype in these novel treatment regimens should therefore be investigated.

In conclusion, we have identified a polymorphism 3 kb upstream of *IL28B* that is significantly associated with response to PegIFN and RBV for patients with chronic genotype 1 HCV infection. The polymorphism explains much of the difference in response between European-American and African-American patients. Given that the polymorphism appears to associate with natural clearance as well as treatment response, it seems likely that the gene product is involved in the innate control of HCV. Indeed, IFN- λ s have demonstrated antiviral activity against genotype 1 HCV *in vitro*¹¹ and *in vivo*¹². The IFN- λ proteins, encoded by the *IL28A/B* and *IL29* genes, were first described in 2003 (refs 13, 14). These IFNs signal through a unique receptor but appear to share a common downstream signalling system with the type 1 IFNs, including IFN- α . These findings, and further study of the functional mechanism underlying the *IL28B*-response association, may help identify patients for whom therapy is likely to be successful, and highlight the IFN- λ signalling axis as a potential target for novel antiviral drug development.

METHODS SUMMARY

Our primary association tests involved single-marker genotype trend tests performed in three independent groups (European-Americans, $n = 871$; African-Americans, $n = 191$; Hispanics, $n = 75$; Supplementary Information I), using logistic regression models for treatment response and linear regression for baseline viral load (Supplementary Information VI). Association tests were implemented in the PLINK software¹⁵, correcting for several clinical covariates, including baseline (pre-treatment) HCV viral load and severity of fibrosis. Then the association signals (P values) were combined using Stouffer's weighted Z -method¹⁶, correctly taking into account sample sizes, effect sizes and effect directions in each population. This combined P value was then reported as the main result, along with the P values in each ethnic group. A series of quality-control steps resulted in 565,759 polymorphisms for the association tests. We applied methods to assess copy number variants and tested the relation between copy number variants and SVR. To control for the possibility of spurious

associations resulting from population stratification, we used a modified EIGENSTRAT¹⁷ method and corrected for population ancestry within each group. We assessed significance with a Bonferroni correction ($P_{\text{cutoff}} = 4.4 \times 10^{-8}$; see Supplementary Information VIII for details).

Received 27 May; accepted 17 July 2009.

Published online 16 August; corrected 17 September 2009 (see full-text HTML version for details).

1. World Health Organization. *Hepatitis C* (<http://www.who.int/mediacentre/factsheets/fs164/en/>) (2009).
2. McHutchison, J. G. *et al.* Peginterferon alfa-2b or alfa-2a with ribavirin for treatment of hepatitis C infection. *N. Engl. J. Med.* **361**, 580–593 (2009).
3. Muir, A. J., Bornstein, J. D. & Killenberg, P. G. Peginterferon alfa-2b and ribavirin for the treatment of chronic hepatitis C in blacks and non-Hispanic whites. *N. Engl. J. Med.* **350**, 2265–2271 (2004).
4. Ghany, M. G., Strader, D. B., Thomas, D. L. & Seeff, L. B. Diagnosis, management, and treatment of hepatitis C: an update. *Hepatology* **49**, 1335–1374 (2009).
5. Shiffman, M. L. *et al.* Peginterferon alfa-2a and ribavirin in patients with chronic hepatitis C who have failed prior treatment. *Gastroenterology* **126**, 1015–1023; discussion 947 (2004).
6. Yan, K. K. *et al.* Treatment responses in Asians and Caucasians with chronic hepatitis C infection. *World J. Gastroenterol.* **14**, 3416–3420 (2008).
7. Liu, C. H. *et al.* Pegylated interferon- α -2a plus ribavirin for treatment-naive Asian patients with hepatitis C virus genotype 1 infection: a multicenter, randomized controlled trial. *Clin. Infect. Dis.* **47**, 1260–1269 (2008).
8. Wilson, J. F. *et al.* Population genetic structure of variable drug response. *Nature Genet.* **29**, 265–269 (2001).
9. Sarasin-Filipowicz, M. *et al.* Interferon signaling and treatment outcome in chronic hepatitis C. *Proc. Natl Acad. Sci. USA* **105**, 7034–7039 (2008).
10. McHutchison, J. G. *et al.* Telaprevir with peginterferon and ribavirin for chronic HCV genotype 1 infection. *N. Engl. J. Med.* **360**, 1827–1838 (2009).
11. Robek, M. D., Boyd, B. S. & Chisari, F. V. Lambda interferon inhibits hepatitis B and C virus replication. *J. Virol.* **79**, 3851–3854 (2005).
12. Shiffman, M. L. *et al.* PEG-IFN- λ : antiviral activity and safety profile in a 4-week phase 1b study in relapsed genotype 1 hepatitis C infection. *J. Hepatol.* **50** (suppl. 1), abstr. A643 s237 (2009).
13. Kotenko, S. V. *et al.* IFN- λ s mediate antiviral protection through a distinct class II cytokine receptor complex. *Nature Immunol.* **4**, 69–77 (2003).
14. Sheppard, P. *et al.* IL-28, IL-29 and their class II cytokine receptor IL-28R. *Nature Immunol.* **4**, 63–68 (2003).
15. Purcell, S. *et al.* PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
16. Whitlock, M. C. Combining probability from independent tests: the weighted Z -method is superior to Fisher's approach. *J. Evol. Biol.* **18**, 1368–1373 (2005).
17. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genet.* **38**, 904–909 (2006).
18. The French METAVIR Cooperative Study Group. Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. *Hepatology* **20**, 15–20 (1994).
19. Ge, D. *et al.* WGAViewer: software for genomic annotation of whole genome association studies. *Genome Res.* **18**, 640–643 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are indebted to the IDEAL principal investigators, the study coordinators, nurses and patients involved in the study. We also recognize E. Gustafson, P. Savino, D. Devlin, S. Noviello, M. Geffner, J. Albrecht and A. C. Need for their contributions to the study. This study was funded by Schering-Plough Research Institute, Kenilworth, New Jersey. A.J.T. received funding support from the National Health and Medical Research Council of Australia and the Gastroenterological Society of Australia.

Author Contributions D.G., J.F., A.J.T. and J.S.S. contributed equally to this work. D.G., J.F. and A.J.T. performed the statistical and bioinformatical analyses. J.F. and A.J.T. defined the clinical phenotypes. K.V.S. performed the genotyping. D.B.G. and D.G. drafted the manuscript. J.S.S., J.G.M. and D.B.G. designed the study. All authors collected and analysed data and contributed to preparing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.B.G. (d.goldstein@duke.edu).