

## Variant in *PNPLA3* is associated with alcoholic liver disease

Chao Tian<sup>1</sup>, Renee P Stokowski<sup>1</sup>, David Kershenobich<sup>2</sup>,  
Dennis G Ballinger<sup>1,3</sup> & David A Hinds<sup>1</sup>

**Two genome-wide association studies (GWAS) have described associations of variants in *PNPLA3* with nonalcoholic fatty liver and plasma liver enzyme levels. We investigated the contributions of these variants to liver disease in Mestizo subjects with a history of alcohol dependence. We found that rs738409 in *PNPLA3* is strongly associated with alcoholic liver disease and clinically evident alcoholic cirrhosis (unadjusted OR = 2.25,  $P = 1.7 \times 10^{-10}$ ; ancestry-adjusted OR = 1.79,  $P = 1.9 \times 10^{-5}$ ).**

Liver cirrhosis is the 12th leading cause of death in the United States, with a total of 28,175 deaths reported in 2005<sup>1</sup>. Nearly half of these were classified as alcohol related. Individuals of Hispanic origin are disproportionately affected by chronic liver disease<sup>2</sup>. In 2007, cirrhosis was the fourth leading cause of death in Mexico and the second leading cause in adults age 15–64 (<http://sinais.salud.gob.mx/mortalidad/>). Only 10–15% of alcoholics develop cirrhosis, and although patterns of alcohol consumption are clearly important, they do not appear to fully account for the ethnic differences in cirrhosis incidence rates<sup>3</sup>.

Recently, Romeo *et al.*<sup>4</sup> carried out a GWAS of nonsynonymous sequence variations in a population comprising Hispanics, African Americans and European Americans to identify genetic variants contributing to nonalcoholic fatty liver disease. They found strong evidence of association of an allele in *PNPLA3* (rs738409: C>G, NP\_079501.2:p.I148M) with increased hepatic fat levels ( $P = 5.9 \times 10^{-10}$ ), and the association remained highly significant ( $P = 7.0 \times 10^{-14}$ ) after adjustment for body mass index, diabetes status, ethanol use and genetic ancestry. The study also revealed a significant elevation in serum concentrations of alanine aminotransferase (ALT) in association with the rs738409[G] allele in Hispanics ( $P = 1.3 \times 10^{-5}$ ). Resequencing the region revealed another allele (rs6006460: G>T, NP\_079501.2:p.S453I) that was independently associated with lower hepatic fat content

in African Americans, the group at lowest risk of nonalcoholic fatty liver disease. An independent GWAS<sup>5</sup> aiming to find genes influencing plasma levels of liver enzymes also showed evidence that variants in *PNPLA3* were associated with plasma levels of ALT. This study gave additional support to the hypothesis that rs738409[G] may confer increased susceptibility to hepatic injury. These associations have recently been confirmed in additional cohorts<sup>6,7</sup>.

We set out to assess whether variation in *PNPLA3* was also associated with clinically evident liver disease in Mestizo (mixed European and Native American ancestry) individuals from Mexico City with a history of substantial alcohol abuse. In addition to the reported nonsynonymous variants rs738409 and rs6006460, we assayed 15 common tagging SNPs from the *PNPLA3* region, 291 SNPs for assessing global ancestry<sup>8</sup>, 16 ancestry-informative markers flanking the *PNPLA3* region for assessing local ancestry<sup>9</sup> (Supplementary Table 1) and 7 SNPs previously reported to be associated with cirrhosis in individuals with hepatitis C<sup>10</sup>. We successfully genotyped 305 individuals with a history of alcohol dependence but apparently normal liver function (controls), 434 with intermediate alcoholic liver disease (ALD) and 482 with clinically evident alcoholic cirrhosis. Diagnoses were based on biochemical and clinical assessments supported by imaging, without histological confirmation (Supplementary Fig. 1 and Supplementary Methods).

The clinical characteristics of our samples are shown in Table 1. The control, ALD and cirrhosis diagnosis groups had significant differences (all  $P < 10^{-10}$ ) in mean age, mean current alcohol intake and duration of alcohol consumption, and in mean global and local loadings from a principal component analysis (PCA) of the genotyping data

**Table 1 Clinical characteristics of the genotyped subjects**

| Characteristic                               | Control         | ALD              | Cirrhosis      | <i>P</i> value         |
|--|-----------------|------------------|----------------|------------------------|
| Gender: <i>n</i> (%)                         |                 |                  |                | $9.2 \times 10^{-1}$   |
| F  | 48 (0.16)       | 67 (0.15)        | 71 (0.15)      |                        |
| M  | 257 (0.84)      | 367 (0.85)       | 411 (0.85)     |                        |
| Current alcohol intake, g/week: <i>n</i> (%) |                 |                  |                | $9.4 \times 10^{-11}$  |
| 200–500                                      | 186 (0.61)      | 162 (0.37)       | 179 (0.37)     |                        |
| 500–750                                      | 20 (0.07)       | 46 (0.11)        | 48 (0.10)      |                        |
| >750   | 99 (0.32)       | 226 (0.52)       | 255 (0.53)     |                        |
| Duration, years: mean (s.d.)                 | 17.5 (9.62)     | 19.5 (9.89)      | 26.1 (11.42)   | $<2.2 \times 10^{-16}$ |
| Age, years: mean (s.d.)                      | 39 (12.7)       | 41 (12.4)        | 52 (11.6)      | $<2.2 \times 10^{-16}$ |
| Global PC1 loading: mean (s.d.)              | –0.0098 (0.027) | –0.00004 (0.024) | 0.0058 (0.026) | $<2.2 \times 10^{-16}$ |
| Local PC1 loading: mean (s.d.)               | –0.0094 (0.029) | 0.00106 (0.026)  | 0.0049 (0.027) | $5.3 \times 10^{-12}$  |
| Child-Pugh class: <i>n</i> (%)               |                 |                  |                | $<2.2 \times 10^{-16}$ |
| A  | 294 (1.00)      | 312 (1.00)       | 209 (0.44)     |                        |
| B  | 0 (0.00)        | 0 (0.00)         | 215 (0.45)     |                        |
| C  | 0 (0.00)        | 0 (0.00)         | 55 (0.11)      |                        |

*P* values are from *F* tests for continuous outcomes and Pearson  $\chi^2$  tests for categorical outcomes.

<sup>1</sup>Perlegen Sciences, Mountain View, California, USA. <sup>2</sup>Departamento de Medicina Experimental, Facultad de Medicina de la Universidad Nacional Autónoma de México, Mexico City, Mexico. <sup>3</sup>Complete Genomics, Mountain View, California, USA. Correspondence should be addressed to D.A.H. (david\_hinds@perlegen.com).

Received 27 July; accepted 21 October; published online 29 November 2009; doi:10.1038/ng.488

**Table 2 Association test results for rs738409**

|                       | Genotype count |     |     |      |     |     | Logistic regression likelihood ratio tests |                       |                   |                      |                             |                      |
|-----------------------|----------------|-----|-----|------|-----|-----|--|-----------------------|-------------------|----------------------|-----------------------------|----------------------|
|                       | Control        |     |     | Case |     |     | No ancestry correction                     |                       | Global correction |                      | Global and local correction |                      |
|                       | CC             | CG  | GG  | CC   | CG  | GG  | OR (95% CI)                                | <i>P</i>              | OR (95% CI)       | <i>P</i>             | OR (95% CI)                 | <i>P</i>             |
| Cirrhosis vs. control | 83             | 198 | 111 | 59   | 264 | 371 | 2.25 (1.74–2.90)                           | $1.7 \times 10^{-10}$ | 1.79 (1.37–2.35)  | $1.9 \times 10^{-5}$ | 1.81 (1.36–2.41)            | $4.7 \times 10^{-5}$ |
| Cirrhosis vs. ALD     | 91             | 266 | 305 | 59   | 264 | 371 | 1.43 (1.15–1.78)                           | $1.0 \times 10^{-3}$  | 1.33 (1.06–1.66)  | $1.4 \times 10^{-2}$ | 1.45 (1.13–1.84)            | $2.8 \times 10^{-3}$ |
| ALD vs. control       | 83             | 198 | 111 | 91   | 266 | 305 | 1.45 (1.16–1.80)                           | $8.4 \times 10^{-4}$  | 1.26 (1.00–1.58)  | $5.1 \times 10^{-2}$ | 1.18 (0.92–1.51)            | $1.9 \times 10^{-1}$ |

Genotypes describe the forward strand on NCBI build 36, and odds ratios are per G allele.

(**Supplementary Figs. 2–4**). Both the local and global mean individual proportions of Native American ancestry were significantly higher among the individuals with cirrhosis than the control individuals ( $P < 2.2 \times 10^{-16}$ ), consistent with the higher prevalence of cirrhosis in Hispanics compared to individuals of European or African ancestry<sup>2</sup>.

We tested each SNP for association with clinically evident cirrhosis or ALD using likelihood ratio tests from a logistic regression analysis, adjusted for age, alcohol intake and duration, an interaction between age and duration, and global genetic ancestry estimated using PCA (**Supplementary Table 2**). The association analysis was carried out for three pairwise combinations of diagnosis status: cirrhosis versus control, cirrhosis versus ALD, and ALD versus control. We also performed tests controlling for local ancestry along the 9.4-Mb region flanking rs738409 estimated from PCA using the 16 ancestry-informative markers. SNP rs738409 was strongly associated with alcoholic cirrhosis (**Table 2**).

In tests of cirrhosis versus control status, rs738409 showed strong association before and after controlling for global ancestry ( $P = 1.7 \times 10^{-10}$ ,  $P = 1.9 \times 10^{-5}$ ) as well as after controlling for local ancestry ( $P = 4.7 \times 10^{-5}$ ). Test results for cirrhosis versus ALD, and for ALD versus control status, suggest that rs738409 has an intermediate effect on the ALD phenotype. We used the Akaike information criterion to compare four genetic models (2-d.f. general model and 1-d.f. additive, dominant and recessive risk-allele models) using logistic regression adjusting for covariates and the global individual ancestry. The most parsimonious model was an additive model for rs738409 [G] (**Supplementary Table 3**). Logistic regression analysis suggested that the rs738409 sequence variation accounts for 49% of the observed ancestry-related difference in cirrhosis susceptibility. Further tests showed no association or interaction of rs738409 with other covariates, including age, alcohol intake and duration of intake. Matched analyses gave very similar estimates of the genotype effect of rs738409 (**Supplementary Methods**).

Association test results for the tagging SNPs were generally consistent with their extent of linkage disequilibrium (LD) with rs738409 (**Supplementary Fig. 5** and **Table 2**). SNP rs738408 is 3 base pairs away from rs738409 and these two SNPs are in nearly complete LD ( $r^2 = 0.99$ ). Association tests for rs738408 gave results nearly identical to those for rs738409. When rs738409 was treated as causal by including its genotypes as a covariate in the regression model, the additional associations in the *PNPLA3* region were eliminated. All common haplotypes containing the rs738409[G] allele were more common among individuals with cirrhosis than in the control group (**Supplementary Table 4**). The rare variant rs6006460[T] reported by Romeo *et al.*<sup>4</sup> is also rare in our Mestizo population (minor allele frequency = 0.002), so we had no power to detect an effect of this variant. The minor T allele of rs6006460 was observed in both the cirrhosis and control groups. We did not observe significant associations with any of the markers previously reported to be associated with cirrhosis in individuals with hepatitis C<sup>10</sup> (**Supplementary Table 2**).

We further tested for an association of rs738409 with a prognostic measure of disease severity in individuals with cirrhosis, the Child-Pugh score<sup>11</sup>. We coded Child-Pugh classes as numeric scores from 1 to 3 in order of severity and fit the data using linear regression with the same covariates used in the binary outcome models and an additive genotype term. The high-risk G allele showed a suggestive association with increasing disease severity (Wald test, one-sided  $P = 0.05$ ). The frequencies of rs738409[G] in individuals scored as Child-Pugh class A, B and C (with class C being the most severe), were 0.70, 0.75 and 0.77, respectively.

The biochemical function of the *PNPLA3* protein is unclear, though it appears to have lipogenic transacylase activity<sup>12</sup>. The gene is highly expressed in adipocytes and liver<sup>7</sup>, and the protein may have a role in energy homeostasis<sup>13</sup>. Obesity and diabetes are known risk factors for non-alcoholic as well as alcoholic liver disease, and expression of *PNPLA3* is correlated with obesity<sup>6</sup>. However, studies associating rs738409 with liver dysfunction have not detected associations with systemic metabolic phenotypes, including obesity, plasma lipid levels or insulin resistance<sup>4,6</sup>. These and other results<sup>7</sup> suggest that rs738409 is an independent risk factor for liver dysfunction. Our work supports a central role for altered lipid processing in liver pathogenesis<sup>14</sup>, but further investigation of the mechanism of *PNPLA3* activity is warranted. As our study was limited to analyses of categorical diagnoses without detailed supporting quantitative phenotype data, additional work will be required to more precisely characterize the disease states associated with this polymorphism (also see **Supplementary Note**).

Our study extends the previously reported associations of rs738409 with subclinical nonalcoholic liver disease to clinically relevant diagnoses of alcoholic liver disease. This single variant accounts for a substantial share of the increased risk of cirrhosis associated with Hispanic ancestry. Hispanics with hepatitis C are also at substantially elevated risk for hepatic injury compared to individuals of other ethnicities<sup>15</sup>, and it will be important to determine whether rs738409 is associated in that context as well. The effect size of rs738409 is large for an association with complex disease in humans and may be the largest known genetic modifier for a disease that is a major cause of preventable death. For these reasons, this variant may be an attractive target for genetic screening to identify individuals at high risk of liver disease for more aggressive interventions.

*Note: Supplementary information is available on the Nature Genetics website.*

#### ACKNOWLEDGMENTS

We thank the individual participants in the study who made this work possible; D. Cox, N. Patil, N. Shen and C. Hacker for their work to obtain and characterize the samples; E. Beasley and M. Mealiffe for guidance; and E. Beilharz, T. Genschoreck, M. Kelleher, M. Kenner, A. Ollmann, K. Pant and S. Pitts for technical assistance.

#### AUTHOR CONTRIBUTIONS

D.G.B. participated in the conception and design of the study. D.K. provided the samples used in the study. R.P.S. directed the genotyping experiments. D.A.H. and C.T. analyzed the data and drafted the manuscript. All authors contributed to the final manuscript.

## COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at <http://www.nature.com/naturegenetics/>.

Published online at <http://www.nature.com/naturegenetics/>.

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>.

1. Yoon, Y.H. & Yi, H.Y. Surveillance Report #83: Liver cirrhosis mortality in the United States, 1970–2005. (National Institute on Alcohol Abuse and Alcoholism, Arlington, Virginia, USA, 2008).
2. Flores, Y.N. *et al. Am. J. Gastroenterol.* **103**, 2231–2238 (2008).
3. Mann, R.E., Smart, R.G. & Govoni, R. *Alcohol Res. Health* **27**, 209–219 (2003).
4. Romeo, S. *et al. Nat. Genet.* **40**, 1461–1465 (2008).
5. Yuan, X. *et al. Am. J. Hum. Genet.* **83**, 520–528 (2008).
6. Kotronen, A. *et al. Diabetologia* **52**, 1056–1060 (2009).
7. Kollerits, B. *et al. J. Med. Genet.* published online doi:10.1136/jmg.2009.066597 (18 June 2009).
8. Hinds, D.A. *et al. Am. J. Hum. Genet.* **74**, 317–325 (2004).
9. Tian, C. *et al. Am. J. Hum. Genet.* **80**, 1014–1023 (2007).
10. Huang, H. *et al. Hepatology* **46**, 297–306 (2007).
11. Pugh, R.N., Murray-lyon, I. & Dawson, J. *Br. J. Surg.* **60**, 646–649 (1973).
12. Jenkins, C.M. *et al. J. Biol. Chem.* **279**, 48968–48975 (2004).
13. Baulande, S., Lasnier, F., Lucas, M. & Pairault, J. *J. Biol. Chem.* **276**, 33336–33344 (2001).
14. Sozio, M. & Crabb, D.W. *Am. J. Physiol. Endocrinol. Metab.* **295**, E10–E16 (2008).
15. Hoyumpa, A. *et al. Hepatology* **42** S1, 442A (2005).