ORIGINAL ARTICLE

Risk alleles for chronic hepatitis B are associated with decreased mRNA expression of HLA-DPA1 and HLA-DPB1 in normal human liver

TR O'Brien[1], I Kohaar[2], RM Pfeiffer[3], D Maeder[4], M Yeager[4,5], EE Schadt[6,7] and L Prokunina-Olsson[2]

[1]*Infections and Immunoepidemiology Branch, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA;* [2]*Laboratory of Translational Genomics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA;* [3]*Biostatistics Branch, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA;* [4]*Human Genetics Program, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA;* [5]*NCI Core Genotyping Facility, SAIC-Frederick, Inc., NCI-Frederick, Gaithersburg, MD, USA;* [6]*Sage Bionetworks, Seattle, WA, USA and* [7]*Pacific Biosciences, Menlo Park, CA, USA*

A genome-wide association study identified single nucleotide polymorphisms (SNPs) rs3077 and rs9277535 located in the 3′ untranslated regions of human leukocyte antigen (HLA) class II genes HLA-DPA1 and HLA-DPB1, respectively, as the independent variants most strongly associated with chronic hepatitis B. We examined whether these SNPs are associated with mRNA expression of HLA-DPA1 and HLA-DPB1. We identified gene expression-associated SNPs (eSNPs) in normal liver samples obtained from 651 individuals of European ancestry by integrating genotype ($\sim$650 000 SNPs) and gene expression (>39 000 transcripts) data from each sample. We used the Kruskal–Wallis test to determine associations between gene expression and genotype. To confirm findings, we measured allelic expression imbalance (AEI) of complementary DNA compared with DNA in liver specimens from subjects who were heterozygous for rs3077 and rs9277535. On a genome-wide basis, rs3077 was the SNP most strongly associated with HLA-DPA1 expression ($p = 10^{-48}$), and rs9277535 was strongly associated with HLA-DPB1 expression ($p = 10^{-15}$). Consistent with these gene expression associations, we observed AEI for both rs3077 ($p = 3.0 \times 10^{-7}$; 17 samples) and rs9277535 ($p = 0.001$; 17 samples). We conclude that the variants previously associated with chronic hepatitis B are also strongly associated with mRNA expression of HLA-DPA1 and HLA-DPB1, suggesting that expression of these genes is important in control of HBV.
Genes and Immunity (2011) **12**, 428–433; doi:10.1038/gene.2011.11; published online 24 February 2011

Introduction

Chronic infection with hepatitis B virus (HBV) is the most common cause of liver cancer worldwide, as well as a major risk factor for development of cirrhosis and end-stage liver disease.[1] The hepatitis B vaccine is highly effective in preventing new infections, but $\sim$360 million people still suffer from chronic hepatitis B and there are $\sim$600 000 annual deaths from HBV-related causes.[1] Recent genome-wide association studies identified single nucleotide polymorphisms (SNPs) located within the human leukocyte antigen (HLA) class II genes *HLA-DPA1* and *HLA-DPB1* to be associated with chronic hepatitis B.[2,3] Replication studies performed on two of these SNPs (rs3077 and rs9277535) confirmed and strengthened results from the genome-wide association studies.[2] Class II HLA genes encode proteins expressed on the surface of antigen-presenting cells such as macrophages, dendritic cells and B cells, and thereby have a critical role in presentation of antigens to CD4+ T-helper lymphocytes. HLA genes have many structural variants that have been linked to immune response to infectious agents,[4] but genetic variants that influence HLA mRNA expression might also affect antigen presentation and many 'gene expression-associated SNPs' (eSNPs) have been found for HLA genes.[5,6] An integrated approach combining genotype information with genome-wide gene expression data in relevant tissues can identify genetic variations that are both regulatory and disease causing.[6,7] In this study, we examined whether SNPs implicated in chronic hepatitis B are associated with mRNA expression in liver samples from the Human Liver Cohort, a large study of genotype and gene expression in normal liver.[6] Additionally, we performed confirmatory allelic expression imbalance

Correspondence: Dr TR O'Brien, Division of Cancer Epidemiology and Genetics, National Cancer Institute 6120 Executive Blvd., EPS/5052, MSC 7246, Bethesda, MD 20892, USA.
E-mail: obrient@mail.nih.gov

(AEI) studies in human liver and peripheral blood. We found that SNPs associated with chronic hepatitis B are strongly associated with decreased expression of important antigen-presenting molecules, *HLA-DPA1* and *HLA-DPB1*.

Results

Genetic variants and mRNA expression

A published genome-wide association study for chronic hepatitis B identified 11 disease-associated SNPs; confirmatory replication studies were performed on rs3077 and rs9277535 (ref. 2). SNP rs3077 is located in the 3′ untranslated region (3′ UTR) of *HLA-DPA1*, rs2395309 lies ~6.5 kb downstream of *HLA-DPA1* and rs2301220 is

found in the first intron of *HLA-DPA1*. These SNPs are in complete linkage disequilibrium (LD) in Europeans (CEU HapMap samples, $r^2 = 1.0$; Figure 1a) and in near complete LD in Asians (JPT + CHB HapMap samples, $r^2 = 0.93$–0.97, Figure 1b). Among the ~650 000 SNPs that were genotyped and examined for association with expression of *HLA-DPA1*, rs2395309, rs3077 and rs2301220 were the variants most strongly associated with mRNA expression of this gene ($p \sim 10^{-48}$; Table 1). Of the eight remaining SNPs that were associated with chronic hepatitis B,[2] five were associated with expression of *HLA-DPA1* as well (Table 1).

Figure 2a shows odds ratios for chronic hepatitis B[2] and differences in mean-log gene expression by rs3077 genotypes. Gene expression was measured relative to a pool of control liver samples,[6] and the risk genotype
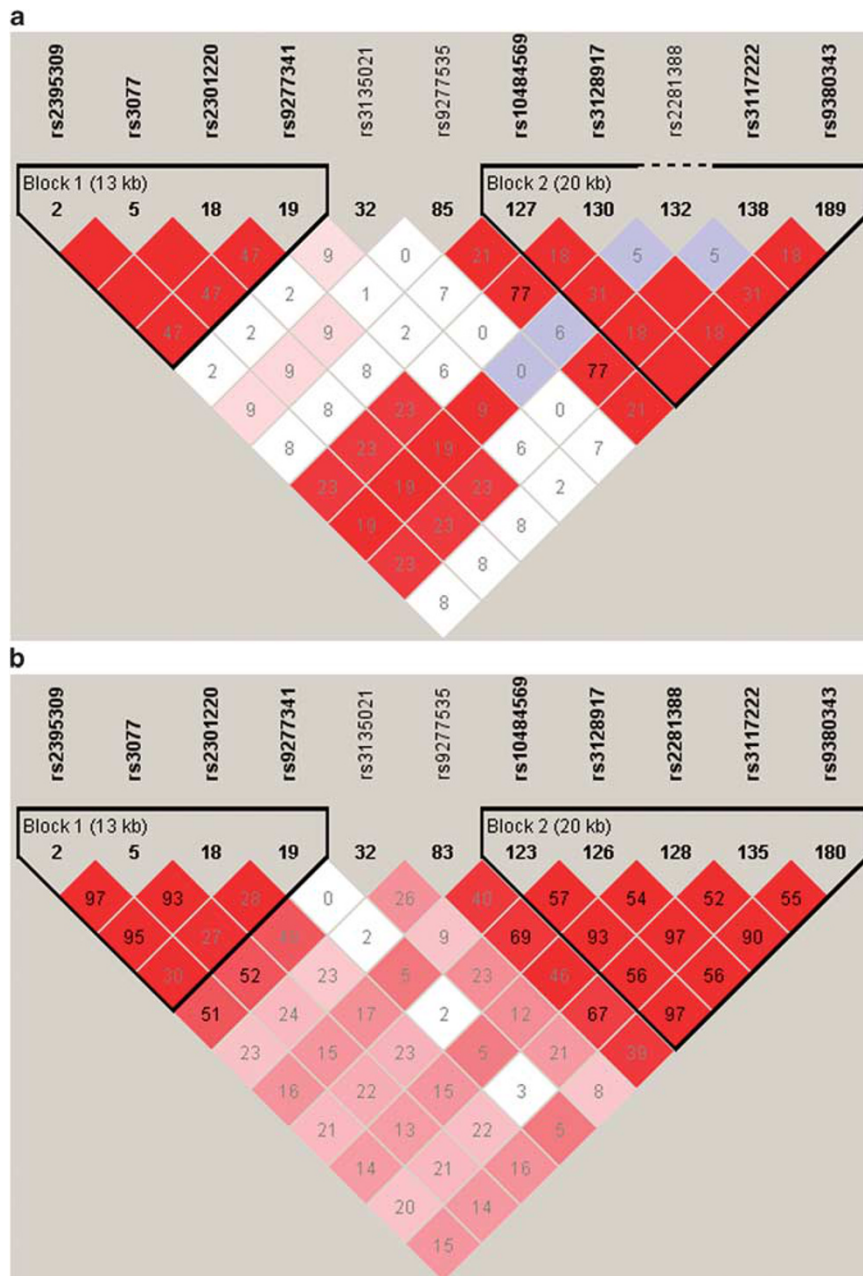

Figure 1 LD (r^2) between SNPs associated with chronic hepatitis B: (**a**) European samples (CEU, HapMap); (**b**) Asian samples (JPT and CHB, HapMap).

Table 1 SNPs associated with chronic infection with HBV[2] and with mRNA expression of *HLA-DPA1* and *HLA-DPB1* in 651 human liver tissue samples

SNPs	HBV GWAS[2]	HLA-DPA1 Expression in Liver		HLA-DPB1 Expression in Liver		Location within HLA region
	P-value	Rank	P-value	Rank	P-value	
rs2395309	1.16E-13	1	1.24E-48		NS	DPA1, downstream
rs3077	1.59E-13	1	1.24E-48		NS	DPA1, 3' UTR
rs2301220	1.81E-13	5	4.13E-47		NS	DPA1, intron 1
rs9277535	2.14E-12	17	1.92E-08	2	2.28E-15	DPB1, 3' UTR
rs3117222	1.29E-10	14	9.76E-10		NS	DPB1, downstream
rs3128917	1.98E-10	15	9.76E-10		NS	DPB1, downstream
rs3135021	1.63E-08		NS		NS	DPB1, intron 1
rs9380343	1.66E-08		NS		NS	DPB2, upstream
rs9277341	1.84E-08	7	1.32E-22		NS	DPA1, intron 1
rs10484569	2.34E-08		NS		NS	DPB1, downstream
rs2281388	3.62E-08	21	1.98E-06		NS	DPB1, downstream

Abbreviations: GWAS, genome-wide association study; HBV, hepatitis B virus; HLA, human leukocyte antigen; SNPs, single nucleotide polymorphisms.
Rank indicates position among all *cis* SNPs that were tested for expression; P-values are for association of SNPs with chronic hepatitis B based on the GWAS[2] or with gene expression; NS indicates a non-significant P-value ($> 5.0 \times 10^{-5}$ for *cis*-effect (probe to SNP distance < 1 Mb)).
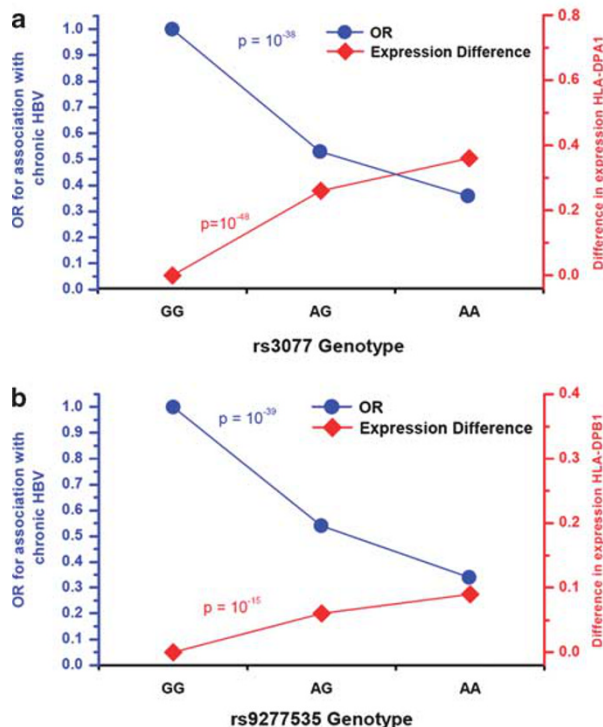
Figure 2 Association of rs3077 and rs9277535 with risk of chronic HBV infection[2] (blue) and mRNA expression of *HLA-DPA1* and *HLA-DPB1* in 651 human liver tissue samples from the Human Liver Cohort (red).[6] (**a**) Odds ratios for chronic HBV[2] and difference in expression of *HLA-DPA1* in liver tissue, by rs3077 genotype. (**b**) Odds ratios for chronic HBV[2] and difference in expression of *HLA-DPB1* in liver tissue, by rs9277535 genotype. Gene expression values represent the arithmetic increase in mean-log gene expression compared with the risk genotype (rs3077GG or rs9277535GG) groups.

for chronic HBV (GG) is set to zero as a reference. The rs3077-G allele was associated with both higher risk of chronic hepatitis B and lower expression of *HLA-DPA1* in a pattern consistent with an additive genetic model.

Similarly, SNP rs9277535, located within the 3' UTR of *HLA-DPB1* (~ 22 kb from rs3077), was strongly associated with both increased risk of chronic hepatitis B and decreased expression of *HLA-DPB1* ($p = 10^{-15}$; Table 1, Figure 2b). There is only weak LD between rs9277535 and rs3077 in Europeans ($r^2 = 0.09$, D' $= 0.50$; HapMap CEU; Figure 1a) and Asians ($r^2 = 0.24$, D' $= 0.54$; HapMap JPT + CHB; Figure 1b), suggesting that effects of these SNPs are likely to be independent.

To confirm that rs3077 and rs9277535 are associated with increased mRNA expression of their respective genes, we also examined genotype and gene expression data from a publically available database of 400 Epstein-Barr virus-transformed lymphoblastoid cell lines obtained from British children with asthma.[5] Among these subjects, rs3077 was strongly associated with *HLA-DPA1* expression ($p = 10^{-21}$) and rs9277535 with *HLA-DPB1* expression ($p = 10^{-12}$).

Allelic expression imbalance
In the Human Liver Cohort, mRNA expression was measured by a microarray. It has been suggested that genetic and structural variations within the HLA region could affect efficiency of hybridization of microarray probes,[8,9] and, if so, differences in mRNA levels could reflect differences in detection efficiency rather than true differences in gene expression. We reasoned that if a genetic variation is truly associated with altered gene expression in *cis*, samples from heterozygous individuals should display AEI in complementary DNA (cDNA) when compared with DNA.[10] The location of rs3077 and rs9277535 in transcribed (3' UTR) regions of *HLA-DPA1* and *HLA-DPB1* makes these SNPs good candidates for AEI testing. Furthermore, AEI Taqman assays for these SNPs were designed to avoid any other underlying genetic variation in the amplicons. Therefore, AEI is an independent test for possible differences in gene expression associated with genetic variants. Consistent with the mRNA expression associations found in the Human Liver Cohort, we observed AEI for both rs3077 (liver tissue ($n = 17$), $p = 3.0 \times 10^{-7}$, Figure 3a) and monocytes (($n = 22$), $p = 2.0 \times 10^{-8}$, Figure 3b) and rs9277535 (liver
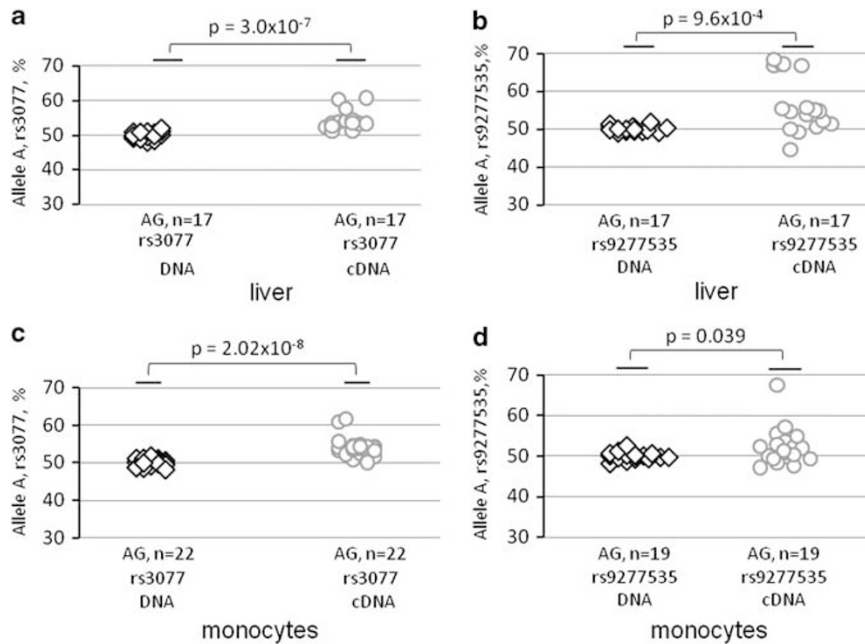
Figure 3 AEI in heterozygous DNA and cDNA in human samples: (**a**) liver ($n = 17$) for rs3077; (**b**) liver ($n = 17$) for rs9277535; (**c**) monocytes ($n = 22$) for rs3077; (**d**) monocytes ($n = 19$) for rs9277535.

tissue ($n = 17$), $p = 0.001$, Figure 3c) and monocytes (($n = 17$), $p = 0.04$, Figure 3d). For both SNPs, the proportion of non-risk allele A was significantly increased in cDNA compared with DNA of heterozygous samples, indicating reduced expression of the risk alleles. AEI for rs3077 was not explained by rs9277535 and vice versa (Supplementary Figure 1), although the statistical power for those comparisons was more limited than for the primary comparisons. Therefore, it is likely that rs3077 and rs9277535 (or variants in high LD with these SNPs) exhibit independent effects on the expression of *HLA-DPA1* and *HLA-DPB1*, respectively.

Discussion

On the basis of findings for both genome-wide mRNA expression and allelic expression imbalance, it seems that SNPs rs3077 and rs9277535 are strongly associated with regulation of *HLA-DPA1* and *HLA-DPB1*, respectively. Previous studies found that these SNPs were highly associated with chronic HBV infection.[2,3] Together, these independent studies strongly implicate lower expression of *HLA-DPA1* and *HLA-DPB1* with increased risk of chronic HBV.

HLA-DPA1 and HLA-DPB1 are expressed on the surface of antigen-presenting cells. In the liver, expression of these proteins is likely to be limited to a small population of Kupffer cells, the resident macrophages of the liver. Kupffer cells are derived from blood monocytes, and we observed evidence for AEI in both monocytes and liver samples. We found that mRNA expression of both *HLA-DPA1* and *HLA-DPB1* was low in total human liver tissue, suggesting that it might be limited to a specific sub-population of cells. We attempted to quantify protein expression in lysates from total liver tissue, but failed to detect a measurable signal (data not shown). Similarly, HLA-DPA1 and HLA-DPB1

protein expression was undetectable in liver tissue examined by antibody-based proteomic methods.[11] It is intriguing that the genetic risk variants associated with chronic hepatitis B affect expression of the alpha (DPA) and beta (DPB) chains of the same antigen-presenting complex, as this suggests that insufficient expression of either or both of these chains might result in the same phenotype.

Kamatani et al.[2] reported that chronic hepatitis B was associated with haplotypes comprised of rs3077-G, rs9277535-G and certain structural variants of *HLA-DPA1* and *HLA-DPB1*. Our findings do not exclude a role for structural variants of *HLA-DPA1* and *HLA-DPB1* in chronic hepatitis B, as viral control of HBV could be affected by both regulatory and structural variants. For example, functional variants of *MBL2*, which has an important role in host defense against HIV-1 and other infectious agents, include both regulatory and structural variants.[12] On a genome-wide basis, rs3077 is the variant most strongly associated with risk of chronic hepatitis B and the variant most strongly associated with expression of *HLA-DPA1*. It seems highly likely, therefore, that expression of *HLA-DPA1* has a role in the clearance of HBV. Similarly, the association of rs9277535 with chronic HBV and with *HLA-DPB1* expression suggests that *HLA-DPB1* expression also contributes to chronic HBV infection.

Kamatani et al.[2] noted that the HBV risk alleles for rs3077 and rs9277535 are more common in Asians than Europeans (fitting the global pattern for the epidemiology of chronic hepatitis B). LD patterns in this genomic region are very similar for populations of Asian or European ancestry groups (Figures 1a and b), indicating that the relationships between SNPs in this region are comparable between these two ancestral groups even though the allele frequencies differ.

We could not determine why the rs3077 and rs9277535 variants are associated with decreased mRNA expression of *HLA-DPA1* and *HLA-DPB1* in liver, but variants within

432

the 3′ UTR may affect mRNA stability through binding of regulatory factors or regulation by microRNAs.[13] SNP rs3077 was found to be associated with methylation level of *HLA-DPB1* and *HLA-DPA1* in adult cerebellum samples studied by Zhang *et al*.[14] (see Supplementary Table 5 of that paper). It may also be noteworthy that rs9277535 was linked to two different copy-number variation regions in European HapMap subjects.[15] We did not observe any obvious factors that could explain differential expression of *HLA-DPA1* and *HLA-DPB1* mRNA, but future studies should address this question.

SNPs rs3077and rs9277535 have been associated with other diseases. The rs9277535 variant has been linked to primary biliary cirrhosis.[16] Both rs3077 and rs9277535 have been associated with rheumatoid arthritis,[17] although other SNPs in the major histocompatibility complex class II region were more strongly associated with that disease.

In conclusion, genetic variants previously associated with chronic hepatitis B[2] are also associated with decreased expression of an antigen-presenting complex that consists of *HLA-DPA1* and *HLA-DPB1* chains. Together, these independent studies strongly implicate lower expression of *HLA-DPA1* and *HLA-DPB1* as a factor for increased risk of chronic hepatitis B. Other recent studies have demonstrated a relationship between expression of *HLA-C* and control of HIV.[18,19] It is possible, therefore, that HLA expression has a role in control of a range of viruses. More broadly, our findings lend support to the concept that SNPs identified through an integrated genomic approach can provide both confirmation and functional insights for disease associations.[6,20] If greater expression of *HLA-DPA1* and *HLA-DPB1* facilitate clearance of HBV infection, development of therapies that increase expression of these genes might aid in treatment of chronic hepatitis B.

Methods

Genome-wide association of SNPs with mRNA expression
As previously described,[6] the Human Liver Cohort identifies eSNPs by integrating genotype and gene expression data for liver samples without evidence of liver disease. The current analysis is based on 651 liver samples obtained from patients of non-Hispanic European ancestry. Genome-wide gene expression was measured with custom-designed Agilent (Agilent Technologies, Inc., Santa Clara, CA, USA) arrays ($>39\,000$ transcripts), and genotyping was performed with the Illumina (San Diego, CA, USA) Sentrix HumanHap650Y genotyping chips.[6] The Kruskal–Wallis test was used to determine association between adjusted expression traits according to an additive genetic model. False discovery rate threshold was set at 10%. Associations were considered significant at $p < 5.0 \times 10^{-5}$ for *cis*–factors (probe to SNP distance <1 Mb) and $p < 1.0 \times 10^{-8}$ for *trans* –factors.[6] Only SNPs used both in genome-wide association studies for chronic HBV[2] and the mRNA expression study[6] were used in this analysis. For each of the SNPs, gene expression values were compared with the referent non-risk genotypes.

Allelic expression imbalance
Liver tissue samples for AEI studies were non-cancerous specimens provided by the Liver Tissue Cell Distribution

System (LTCDS). Anonymized fresh peripheral blood samples from healthy donors were provided by the Blood Bank at the NIH. Monocytes were purified from fresh blood with CD14^{+}-coated magnetic MicroBeads with AutoMacs (Miltenyi Biotec, Auburn, CA, USA) as previously described.[21] DNA from all samples was prepared with DNAeasy kit (Qiagen, Valencia, CA, USA) and RNA was prepared with Trizol reagent (Invitrogen, Carlsbad, CA, USA) followed by RNAeasy kit (Qiagen). The AEI was quantified with the allelic discrimination genotyping assays C__11916951_10 for rs3077 and a custom-designed assay for rs9277535. To ensure that the signal was specific for cDNA and was not derived from residual DNA in these samples, a 'no reverse-transcriptase' control reaction was tested initially. Genotyping for all DNA and cDNA samples was performed in duplicate on the 7900 Sequence Detection System (ABI, Foster City, CA, USA).

AEI experiments were carried out as previously described.[22,23] A standard curve was prepared based on 10 dilutions of two homozygous DNA samples prepared from liver tissue, representing 5, 15, 25, 35, 45, 55, 65, 75, 85 and 95% of allele 'X', based on the standard curve. The potential AEI in samples heterozygous for rs3077 and rs9277535 was evaluated by comparing the proportion of allele 'X' in DNA with that in cDNA with a paired two-sided *T*-test. To evaluate the potential effect of one SNP on AEI for the other (for example, effect of rs3077 on AEI for rs9277535), we tested whether the proportion of allele 'X' in samples heterozygous for rs9277535 differed by the genotype of rs3077. Using an F-test, we compared the proportion of allele 'A' in all samples that were heterozygous for rs9277535 with the proportion of allele 'A' in samples that were heterozygous for rs9277535, but homozygous for rs3077.

References

1 Shepard CW, Simard EP, Finelli L, Fiore AE, Bell BP. Hepatitis B virus infection: epidemiology and vaccination. *Epidemiol Rev* 2006; **28**: 112–125.

2 Kamatani Y, Wattanapokayakit S, Ochi H, Kawaguchi T, Takahashi A, Hosono N *et al*. A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. *Nat Genet* 2009; **41**: 591–595.

3 Matsuura K, Tanaka Y, Nishida N, Hige S, Asahina Y, Ito K *et al*. Genome-wide association study identifies genetic variants in the HLA-DP locus associated with chronic hepatitis B. *J Hepatol* 2010; **52**(Supplement 1): S282–S283.

4 MHC sequencing consortium. Complete sequence and gene map of a human major histocompatibility complex. *Nature* 1999; **401**: 921–923.

5 Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC *et al*. A genome-wide association study of global gene expression. *Nat Genet* 2007; **39**: 1202–1207.

6 Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY *et al*. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol* 2008; **6**: e107.

7 Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J *et al*. Genetics of gene expression and its effect on disease. *Nature* 2008; **452**: 423–428.

8 Alberts R, Terpstra P, Li Y, Breitling R, Nap J-P, Jansen RC. Sequence polymorphisms cause many false cis eQTLs. *PLoS ONE* 2007; **2**: e622.

9 Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. Mapping complex disease traits with global gene expression. *Nat Rev Genet* 2009; **10**: 184–194.

10 Johnson AD, Zhang Y, Papp AC, Pinsonneault JK, Lim JE, Saffen D *et al*. Polymorphisms affecting gene transcription and mRNA processing in pharmacogenetic candidate genes: detection through allelic expression imbalance in human target tissues. *Pharmacogenet Genomics* 2008; **18**: 781–791.

11 Uhlen M, Bjorling E, Agaton C, Szigyarto CA, Amini B, Andersen E *et al*. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteomics* 2005; **4**: 1920–1932.

12 Garred P, Larsen F, Seyfarth J, Fujita R, Madsen HO. Mannose-binding lectin and its genetic variants. *Genes Immun* 2006; **7**: 85–94.

13 Prokunina L, Alarcon-Riquelme ME. Regulatory SNPs in complex diseases: their identification and functional validation. *Expert Rev Mol Med* 2004; **6**: 1–15.

14 Zhang D, Cheng L, Badner JA, Chen C, Chen Q, Luo W *et al*. Genetic control of individual differences in gene-specific methylation in human brain. *Am J Hum Genet* 2010; **86**: 411–419.

15 Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y *et al*. Origins and functional impact of copy number variation in the human genome. *Nature* 2010; **464**: 704–712.

16 Hirschfield GM, Liu X, Xu C, Lu Y, Xie G, Lu Y *et al*. Primary biliary cirrhosis associated with HLA, IL12A, and IL12RB2 Variants. *N Engl J Med* 2009; **360**: 2544–2555.

17 Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, Ding B *et al*. TRAF1-C5 as a risk locus for rheumatoid arthritis—a genomewide study. *N Engl J Med* 2007; **357**: 1199–1209.

18 Thomas R, Apps R, Qi Y, Gao X, Male V, O'HUigin C *et al*. HLA-C cell surface expression and control of HIV/AIDS correlate with a variant upstream of HLA-C. *Nat Genet* 2009; **41**: 1290–1294.

19 Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M *et al*. A whole-genome association study of major determinants for host control of HIV-1. *Science* 2007; **317**: 944–947.

20 Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R *et al*. Genome-wide associations of gene expression variation in humans. *PLoS Genet* 2005; **1**: e78.

21 Prokunina-Olsson L, Welch C, Hansson O, Adhikari N, Scott L, Usher N *et al*. Tissue-specific alternative splicing of TCF7L2. *Hum Mol Genet* 2009; **18**: 3795–3804.

22 Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, Burdick JT. Mapping determinants of human gene expression by regional and genome-wide association. *Nature* 2005; **437**: 1365–1369.

23 Dennis MY, Paracchini S, Scerri TS, Prokunina-Olsson L, Knight JC, Wade-Martins R *et al*. A common variant associated with dyslexia reduces expression of the KIAA0319 gene. *PLoS Genet* 2009; **5**: e1000436.

Supplementary Information accompanies the paper on Genes and Immunity website (http://www.nature.com/gene)