nature genetics

*HLA-B*5701* genotype is a major determinant of drug-induced liver injury due to flucloxacillin

Ann K Daly[1], Peter T Donaldson[1], Pallav Bhatnagar[1], Yufeng Shen[2], Itsik Pe'er[2], Aris Floratos[2], Mark J Daly[3], David B Goldstein[4], Sally John[5], Matthew R Nelson[6], Julia Graham[1], B Kevin Park[7], John F Dillon[8], William Bernal[9], Heather J Cordell[1], Munir Pirmohamed[7], Guruprasad P Aithal[10,11] & Christopher P Day[1,11], for the DILIGEN study[12] and International SAE Consortium[12]

Drug-induced liver injury (DILI) is an important cause of serious liver disease. The antimicrobial agent flucloxacillin is a common cause of DILI, but the genetic basis for susceptibility remains unclear. We conducted a genome-wide association (GWA) study using 866,399 markers in 51 cases of flucloxacillin DILI and 282 controls matched for sex and ancestry. The GWA showed an association peak in the major histocompatibility complex (MHC) region with the strongest association ($P = 8.7 \times 10^{-33}$) seen for rs2395029[G], a marker in complete linkage disequilibrium (LD) with *HLA-B*5701*. Further MHC genotyping, which included 64 flucloxacillin-tolerant controls, confirmed the association with *HLA-B*5701* (OR = 80.6, $P = 9.0 \times 10^{-19}$). The association was replicated in a second cohort of 23 cases. In *HLA-B*5701* carrier cases, rs10937275 in *ST6GAL1* on chromosome 3 also showed genome-wide significance (OR = 4.1, $P = 1.4 \times 10^{-8}$). These findings provide new insights into the mechanism of flucloxacillin DILI and have the potential to substantially improve diagnosis of this serious disease.

Hepatotoxicity remains the most common cause of clinical trial termination of new therapeutic agents (~33%) and a main cause of postmarketing withdrawals[1]. The standardized incidence rate of symptomatic hepatic adverse drug reaction (ADR) is 8 per 100,000 inhabitants, according to a French study[2]. Idiosyncratic hepatotoxicity accounts for 13% of acute liver failure cases in the United States, and 75% of affected individuals either die or require liver transplantation[3].

Flucloxacillin is widely used in many European countries and Australia for treatment of staphylococcal infection. Its use has been associated with a characteristic cholestatic hepatitis that is more common in females[4] and the elderly and with prolonged treatment courses[5,6]. In the UK, the incidence of flucloxacillin-induced DILI has been estimated at 8.5 in every 100,000 new users in days 1 to 45 after starting treatment[6]. The genetic basis of susceptibility to flucloxacillin DILI has not been examined previously, although studies of DILI due to other drugs have reported notable associations. Two independent studies on co-amoxiclav DILI, which has a similar cholestatic phenotype to that induced by flucloxacillin, reported associations with the class II *HLA-DRB1*1501* allele[7,8]. There are isolated reports of HLA associations for DILI due to other drugs, and ximelagatran-related transaminitis is associated with *HLA-DRB1*0701* (ref. 9). At present, the mechanism for flucloxacillin DILI is poorly understood. *In vitro* studies showed that the CYP3A4-generated 5′-hydroxymethylflucloxacillin metabolite was toxic to human biliary epithelial cells but not to hepatocytes, whereas flucloxacillin was nontoxic to both cell types[10]. Limited evidence for specific T-cell reactivity has also been reported[11].

The DILIGEN study seeks to identify genetic determinants of DILI. A variety of drugs are under investigation in this UK-wide study, but flucloxacillin was the most frequently reported cause of DILI in the initial cohort. To identify the genetic basis of flucloxacillin DILI, we carried out genome-wide and candidate gene association studies on cases from the DILIGEN study and two control groups.

Clinical details of the 51 eligible cases are summarized in **Supplementary Table 1** online. As in previous studies of this ADR, our cases showed a higher proportion of females than males[4]. The average age at onset was 63 years. Most cases (86%) were classified as having suffered cholestatic flucloxacillin-induced liver injury. Most international consensus criteria (ICC) causality scores (92%) were consistent with the disease being either probably or highly probably due to flucloxacillin exposure[12]. More detailed phenotype and genotype data are available at http://www.saeconsortium.org/.

We typed the 51 flucloxacillin DILI cases together with seven additional DILI cases with exposure to both flucloxacillin and co-amoxiclav and 487 controls using the Illumina Human1M BeadChip

[1]Institute of Cellular Medicine and Institute of Human Genetics, Newcastle University, Newcastle upon Tyne, UK. [2]Columbia University, New York, New York, USA. [3]Harvard Medical School, Boston, Massachusetts, USA. [4]Duke University, Durham, North Carolina, USA. [5]Pfizer Inc., New London, Connecticut, USA. [6]GlaxoSmithKline, Research Triangle Park, North Carolina, USA. [7]Department of Pharmacology and Therapeutics, University of Liverpool, Liverpool, UK. [8]Ninewells Hospital and Medical School, Dundee, UK. [9]Kings College Hospital, London, UK. [10]Nottingham Digestive Diseases Centre: Biomedical Research Unit, Nottingham, UK. [11]These authors contributed equally to this work. [12]A full list of members is provided in the **Supplementary Note** online. Correspondence should be addressed to A.K.D. (a.k.daly@ncl.ac.uk).

Received 28 January; accepted 10 April; published online 31 May 2009; doi:10.1038/ng.379

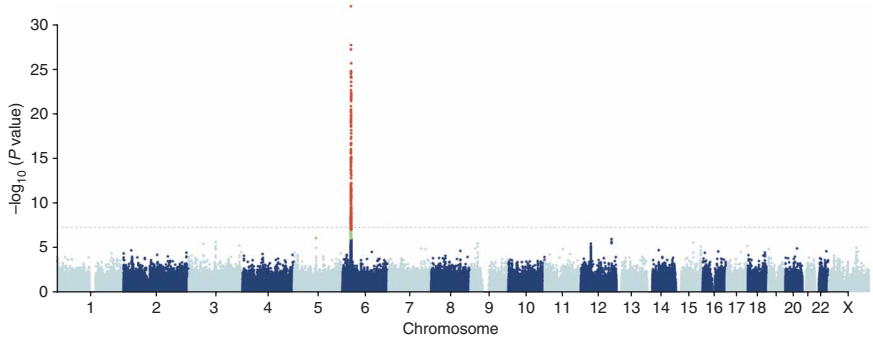

Figure 1 Flucloxacillin DILI genome-wide association result. Each dot represents a SNP. The *x* axis represents the position of the SNP on chromosomes. The *y* axis represents the $-\log_{10}$ of Cochran-Armitage trend *P* value of the SNP in the case-control association study. We included 51 DILI cases and 282 population controls in the study. SNPs with *P* values $<10^{-6}$ and 10^{-7} are highlighted in green and red, respectively. The strong signal in chromosome 6 lies in the MHC region.

(see Methods). Principal component analysis revealed samples with non-European origin. When restricted to European samples, a first component corresponded to a geographic cline[13] separating clusters of North versus South European origin (**Supplementary Fig. 1** online). We therefore restricted analysis to the 58 DILI cases that were all of Northern European origin along with 282 controls of similar origin. From 1,072,820 BeadChip markers, including SNPs and copy number variation (CNV) probes, we discarded 206,421 SNP and CNV markers that failed to meet standard quality control criteria, including filtering criteria for minor allele frequency (98% of the filtered markers), SNP call rate and Hardy-Weinberg equilibrium in the controls (**Supplementary Table 2** online). We then tested each SNP for association. The estimated genomic inflation factor[14] was 1.02, indicating that overall population structure had negligible impact on the case-control association. This was also confirmed by EIGENSTRAT analysis (**Supplementary Results** and **Supplementary Fig. 2a**,**b** online). The GWA revealed one genome-wide significant signal in the MHC region on chromosome 6 (**Figs. 1** and **2a**). The top SNP associated with flucloxacillin DILI was rs2395029, with a *P* value of 8.7×10^{-33} (trend test) and estimated odds ratio (OR) of 45 (95% confidence interval (CI) = 19.4–105). Among cases, 43 (84%) carried the risk allele (G), which has a frequency of approximately 5% in the population controls and in European populations generally (**Supplementary Table 3** online). rs2395029 is a missense polymorphism in the *HCP5* gene which was previously found to be in complete LD with *HLA-B*5701* in subjects of European origin[15]. As expected, owing to strong LD in the MHC region, there were many other strongly associated SNPs (**Supplementary Fig. 3a** online). Association analysis conditioned on rs2395029 indicated that none of these other SNPs in this region was independently significant (**Supplementary Fig. 3b**).

Cases and a separate drug-exposed control group were genotyped directly for *HLA B*5701* (**Table 1**). For the cases, *HLA-B*5701* and rs2395029 genotypes correlated perfectly, consistent with previous reports. *HLA-B*5701* showed a highly significant association with DILI (**Table 1**), with possession of this allele associated with an 80-fold increased risk of developing disease (OR = 80.6; 95% CI = 22.8–284.9). Cases and controls were also genotyped for *HLA* class II and for a range of SNPs in *TNF* and *HSPA1L*, as these MHC-region genes had been associated previously with ADRs due to other drugs and with liver disease generally[7–9,16–18]. For *HLA-DRB1*, *HLA-DRB1*0701* alleles were more common in cases than in controls (OR = 7.2; 95% CI = 3.15–16.45; **Supplementary Table 4** online). *HLA-DRB1*0701* alleles cosegregate with two common *DQB1* alleles, so we were able to identify two common haplotypes on the basis of LD. One of these, the *HLA-DRB1*0701-DQB1*0303* haplotype, was significantly more common among cases compared to controls (**Supplementary Table 5** online). We also observed significant differences between cases and controls for various *TNF* and *HSPA1L* SNPs (*TNF* rs361525 $P = 4.33 \times 10^{-13}$; *TNF* rs1799964 $P = 5.64 \times 10^{-8}$; *HSPA1L* rs2227956 $P = 3.33 \times 10^{-6}$) (**Supplementary**

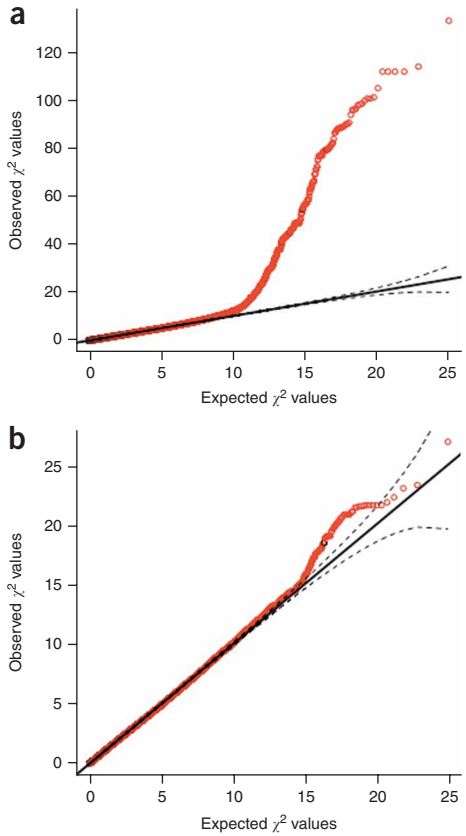

Figure 2 Quantile-quantile plot of χ^2 statistics in case-control study. (**a**) The theoretical quantile-quantile plot of χ^2 statistics in the case-control study. The black solid line represents the null model where observed χ^2 values match the expected values. The dashed lines represent 95% confidence interval of the null model. The red dots represent observed χ^2 values versus the expected values from the case-control study. This quantile-quantile plot indicates that there are many SNPs with genome-wide significance in association with the cases. (**b**) The same quantile-quantile plot as **a**, except that all SNPs from chromosome 6 are removed to show the remaining signals. The significant deviation from null model between expected χ^2 values 15 to 20 suggests that there are additional associations from other chromosomes. We note that, although deviations of observed scores from their expected distribution may also reflect failure to properly correct for population structure, genotype errors or similar artifacts such systematic deviations are predicted to affect the entire distribution rather than its tail end, as in the case of additional associated variants[28,29].

Table 1 Distribution of *HLA-B*5701* genotypes

	Positive	Negative	*P* value	OR (95% CI)
Controls ($n = 64$)	4 (6.3)	60 (93.7)		
Cases ($n = 51$)	43 (84.3)	8 (15.7)	8.97×10^{-19}	80.6 (22.8–284.9)
Replication cases ($n = 23$)	20 (87.0)	3 (13.0)	6.62×10^{-13}	100.0 (20.6–485.8)

Replication cases were those exposed to co-amoxiclav concurrently with flucloxacillin ($n = 7$) and flucloxacillin DILI cases recruited after the GWA study ($n = 16$). Percentages are shown in parentheses. *P* values are uncorrected.

Table 6 online)[19]. Overall, the single-locus analysis conducted on the various MHC genes showed the strongest signals for *TNF* rs361525, *HSPA1L* rs2227956, *HLA-DRB1*0701* and *HLA-B*5701*. There is previous evidence that some of these loci may have a role in liver disease[18], but conditioning analysis of the relevant alleles indicated that *HLA-B*5701* was the most significant for flucloxacillin DILI and that none of the other three loci was conditionally statistically significant (**Supplementary Table 7** online), suggesting that the effects at these loci can be accounted for by (that is, they are not independent of) the association at *HLA-B*5701*.

We compared disease severity on the basis of *HLA-B*5701* genotype (**Supplementary Table 8** online) and found that there were no significant differences between those negative for this allele and those carrying at least one copy. Most of the subjects (65%) had suffered at least one extrahepatic manifestation of disease. In subjects where no extrahepatic symptoms were reported, only two (11%) were negative for *HLA-B*5701*, suggesting that *B*5701* carriage did not increase the incidence of these symptoms. *HLA-B*5701* genotype did not correlate with pattern of injury. The four subjects for whom ICC causality scores[12] were low were all *HLA-B*5701* positive.

For replication purposes, we genotyped *HLA-B*5701* on seven individuals with DILI who had been exposed to co-amoxiclav in addition to flucloxacillin and on 16 recently recruited individuals with flucloxacillin DILI. As summarized in **Table 1**, 20 of these 23 additional subjects (87%) were *HLA-B*5701* positive ($P = 6.62 \times 10^{-13}$).

Despite the strong association with *HLA-B*5701*, only 1 in every 500 to 1,000 individuals with this genotype will develop DILI when treated with flucloxacillin. We estimated that the population attributable fraction (PAF)[20] of *HLA-B*5701* is 0.64. In addition, given the observed distribution of the test statistics, we speculate that there might be other genetic factors that also contribute. After removing all chromosome 6 SNPs from the GWA, we observed that the number of SNPs with small association *P* values ($<5 \times 10^{-6}$) was in excess of chance expectation (**Fig. 2b**, *P* value <0.01), suggesting additional associations elsewhere. A few regions on chromosomes 3, 9 and 12 were potential sources of this excess (**Supplementary Table 9** online).

For analysis of CNV data, we removed 11 cases and 10 population controls that failed quality control. No common CNV was found to be associated with DILI, and the analysis of rare CNVs was inconclusive (**Supplementary Results**, **Supplementary Figs. 4** and **5** and **Supplementary Table 10** online).

To test whether there is an interaction effect between *HLA-B*5701* and other markers, we attempted to fit a general interaction model, but we did not find any genome-wide significant association. We then conducted a separate analysis including only those 48 cases that carry the risk alleles (43 flucloxacillin only and 5 with exposure to both flucloxacillin and co-amoxiclav), and did GWA with all population controls. We found one SNP (rs10937275) on chromosome 3 significantly associated with flucloxacillin DILI (*P* value $= 1.4 \times 10^{-8}$, allelic OR = 4.1, minor allele frequency = 0.29 versus 0.09; **Supplementary Table 11** online). This SNP does not overlap with the chromosome 3 regions highlighted in the analysis of the entire sample. rs10937275 is an intronic SNP in *ST6GAL1* (MIM109675), which encodes an enzyme involved in transfer of sialic acid to cell-surface and serum glycoproteins[21].

Our study of genetic association with DILI linked to a licensed drug done on a genome-wide scale shows that *HLA-B*5701* is the main common genetic risk factor for flucloxacillin DILI and provides suggestive evidence for additional contributions from other loci. *HLA-B*5701* is relatively common in Northern Europe but rarer in Africa and Asia (see URLs section in Online Methods). There is at present no data available on the incidence of flucloxacillin DILI outside the UK, Sweden and Australia, but it seems possible that it may be less common in ancestry groups where *B*5701* frequency is low. Detailed data on MHC haplotypes that include *B*5701* is limited, but there seems to be a common *B*5701*-containing haplotype also including *A*0101*, *Cw*0602*, *DRB1*0701* and *DQB1*0303* (ref. 19). Our data suggest this is the 'at-risk' haplotype for flucloxacillin DILI, but studies to determine whether any other *B*5701*-containing haplotypes are also risk factors would be of interest.

The association described here is stronger than any previously reported genetic association for DILI on the basis of odds ratio and size of *P* value. The role of MHC genotype in susceptibility to DILI is consistent with previous weaker observations for co-amoxiclav DILI and ximelagatran transaminitis but with different associated haplotypes[7–9]. It remains possible that *HLA-B*5701* carriage may be a risk factor for other forms of DILI, but this requires further investigation.

The mechanism for the *HLA-B*5701* association in abacavir hypersensitivity is well studied. Abacavir hypersensitivity affects a number of organs, although specific effects on the liver have generally not been reported[22]. An abacavir metabolite seems to interact with cytotoxic $CD8^+$ T cells, with recognition dependent on the presence of the *HLA-B*5701* antigen, which has several unique amino acids compared with other *HLA-B* gene products[23]. There is no obvious structural similarity between abacavir and flucloxacillin to explain why both ADRs are associated with *HLA-B*5701*. *HCP5* encodes an endogenous retroviral element mainly expressed in immune cells, and rs2395029 has been found to be protective against HIV-1 infection[24]. It remains possible that *HCP5* genotype could affect risk of flucloxacillin DILI directly, possibly by differential effects on inflammation within the liver.

This GWA study has also provided evidence that additional genes may be involved in disease susceptibility. *ST6GAL1* seems to be an additional susceptibility factor in *HLA-B*5701*-positive cases only, suggesting that both genes might interact in the underlying mechanism. Increased hepatic expression of ST6GAL1 has been demonstrated during acute inflammation[25]. In mice lacking this gene, an exaggerated acute neutrophilic response in thioglycolate-induced peritonitis occurs[26]. It therefore seems plausible that *ST6GAL1* genotype could determine susceptibility to an immunoinflammatory event. However, the effect for *ST6GAL1* is modest and needs investigation in a larger group, as does the inconclusive data on CNV.

Clinically, *HLA-B*5701* genotyping of suspected cases of flucloxacillin DILI, possibly using *HCP5* rs2395029 as a convenient alternative to direct typing[15], may prove to be a useful diagnostic test. If a prompt test is available, substituting flucloxacillin with alternative antistaphylococcal agents such as cloxacillin and dicloxacillin in these suspected cases should be feasible[27]. According to our data, a test of suspected DILI cases for *HLA-B*5701* would show that an estimated

85% (based on the case definition as used in this study and subject to substantial uncertainty about the estimate given the small numbers) carry *HLA-B*5701*. A prospective (before treatment) test based on genotype for those who might develop DILI would have a very high false-positive rate, despite the large effect of the variant, because the proportion of carriers that develop DILI on treatment is low. The overall clinical utility of *B*5701* genotyping can, however, be assessed accurately only by a prospective study on the risk of liver disease in *B*5701*-positive versus negative subjects prescribed flucloxacillin.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

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

ACKNOWLEDGMENTS

The DILIGEN sample collection and MHC genotyping was funded by the UK Department of Health (Pharmacogenetics Initiative ref PHGX10A). The GWA study and additional case collection during 2008 was funded by the International Serious Adverse Events Consortium with support from Abbott, Daiichi-Sankyo, GlaxoSmithKline, Johnson & Johnson, Novartis, Pfizer, F. Hoffmann-La Roche, Sanofi-Aventis, Takeda, Wellcome Trust and Wyeth. We extend special thanks to the research committees of British Association for Study of the Liver and British Society for Gastroenterology for help with publicizing the study, to the UK Medicines and Healthcare Products Regulatory Agency for enabling access to a yellow card case and to the subjects for their participation.

AUTHOR CONTRIBUTIONS

A.K.D., P.T.D., B.K.P., M.P., G.P.A. and C.P.D. participated in the conception, design and coordination of the DILIGEN study and in writing the paper; P.B., Y.S., I.P., A.F., M.R.N. and H.J.C. analyzed the data and participated in writing the paper; M.J.D., D.B.G., S.J. and the International Serious Adverse Events Consortium advised on the GWA study and participated in writing the paper; J.G. performed DNA extraction and genotyping; W.B., J.F.D. and the DILIGEN study members recruited study subjects and provided clinical data.

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/

1. Mann, R. & Andrews, E. (eds.). *Pharmacovigilance* (John Wiley and Sons, Chichester, 2006).
2. Sgro, C. *et al.* Incidence of drug-induced hepatic injuries: a French population-based study. *Hepatology* **36**, 451–455 (2002).
3. Ostapowicz, G. *et al.* Results of a prospective study of acute liver failure at 17 tertiary care centers in the United States. *Ann. Intern. Med.* **137**, 947–954 (2002).
4. Olsson, R. *et al.* Liver damage fron flucloxacillin, cloxacillin and dicloxacillin. *J. Hepatol.* **15**, 154–161 (1992).
5. Fairley, C.K., McNeil, J. & Desmond, P. Risk factors for development of flucloxacillin associated jaundice. *Br. Med. J.* **306**, 233–235 (1993).
6. Russmann, S., Kaye, J.A., Jick, S.S. & Jick, H. Risk of cholestatic liver disease associated with flucloxacillin and flucloxacillin prescribing habits in the UK: cohort study using data from the UK General Practice Research Database. *Br. J. Clin. Pharmacol.* **60**, 76–82 (2005).
7. Hautekeete, M.L. *et al.* HLA association of amoxicillin-clavulanate-induced hepatitis. *Gastroenterology* **117**, 1181–1186 (1999).
8. O'Donohue, J. *et al.* Co-amoxiclav jaundice: clinical and histological features and HLA class II association. *Gut* **47**, 717–720 (2000).
9. Kindmark, A. *et al.* Genome-wide pharmacogenetic investigation of a hepatic adverse event without clinical signs of immunopathology suggests an underlying immune pathogenesis. *Pharmacogenomics J.* **8**, 186–195 (2008).
10. Lakehal, F. *et al.* Indirect cytotoxicity of flucloxacillin toward human biliary epithelium via metabolite formation in hepatocytes. *Chem. Res. Toxicol.* **14**, 694–701 (2001).
11. Maria, V.A. & Victorino, R.M. Diagnostic value of specific T cell reactivity to drugs in 95 cases of drug induced liver injury. *Gut* **41**, 534–540 (1997).
12. Danan, G. & Benichou, C. Causality assessment of adverse reactions to drugs–I. A novel method based on the conclusions of international consensus meetings: application to drug-induced liver injuries. *J. Clin. Epidemiol.* **46**, 1323–1330 (1993).
13. Novembre, J. *et al.* Genes mirror geography within Europe. *Nature* **456**, 98–101 (2008).
14. de Bakker, P.I. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* **17**, R122–R128 (2008).
15. Colombo, S. *et al.* The HCP5 single-nucleotide polymorphism: a simple screening tool for prediction of hypersensitivity reaction to abacavir. *J. Infect. Dis.* **198**, 864–867 (2008).
16. Pachkoria, K. *et al.* Analysis of IL-10, IL-4 and TNF-α polymorphisms in drug-induced liver injury (DILI) and its outcome. *J. Hepatol.* **49**, 107–114 (2008).
17. Aithal, G.P. *et al.* Hepatic adducts, circulating antibodies, and cytokine polymorphisms in patients with diclofenac hepatotoxicity. *Hepatology* **39**, 1430–1440 (2004).
18. Donaldson, P.T. Genetics of liver disease: immunogenetics and disease pathogenesis. *Gut* **53**, 599–608 (2004).
19. Ahmad, T. *et al.* Haplotype-specific linkage disequilibrium patterns define the genetic topography of the human MHC. *Hum. Mol. Genet.* **12**, 647–656 (2003).
20. Browner, W.S. & Newman, T.B. Sample size and power based on the population attributable fraction. *Am. J. Public Health* **79**, 1289–1294 (1989).
21. Bast, B.J. *et al.* The HB-6, CDw75, and CD76 differentiation antigens are unique cell-surface carbohydrate determinants generated by the β-galactoside α 2,6-sialyltransferase. *J. Cell Biol.* **116**, 423–435 (1992).
22. Hetherington, S. *et al.* Hypersensitivity reactions during therapy with the nucleoside reverse transcriptase inhibitor abacavir. *Clin. Ther.* **23**, 1603–1614 (2001).
23. Chessman, D. *et al.* Human leukocyte antigen class I-restricted activation of CD8+ T cells provides the immunogenetic basis of a systemic drug hypersensitivity. *Immunity* **28**, 822–832 (2008).
24. Fellay, J. *et al.* A whole-genome association study of major determinants for host control of HIV-1. *Science* **317**, 944–947 (2007).
25. Kaplan, H.A., Woloski, B.M., Hellman, M. & Jamieson, J.C. Studies on the effect of inflammation on rat liver and serum sialyltransferase. *J. Biol. Chem.* **258**, 11505–11509 (1983).
26. Nasirikenari, M., Segal, B.H., Ostberg, J.R., Urbasic, A. & Lau, J.T. Altered granulopoietic profile and exaggerated acute neutrophilic inflammation in mice with targeted deficiency in the sialyltransferase ST6Gal I. *Blood* **108**, 3397–3405 (2006).
27. Sutherland, R., Croydon, E.A. & Rolinson, G.N. Flucloxacillin, a new isoxazolyl penicillin, compared with oxacillin, cloxacillin, and dicloxacillin. *BMJ* **4**, 455–460 (1970).
28. Saxena, R. *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331–1336 (2007).
29. Willer, C.J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* **41**, 25–34 (2009).

ONLINE METHODS

Case recruitment. We included cases with suspected liver injury due to flucloxacillin with either (i) clinically apparent jaundice or bilirubin >40 μmol/l (after exclusion of cases due to hemolysis), (ii) alanine aminotransferase (ALT) >5× ULN (upper limit of normal) or (iii) alkaline phosphatase (ALP) >2× ULN plus any raised bilirubin above ULN. We assessed the causal relationship of liver injury to flucloxacillin using international consensus criteria (ICC)[12]. Ethical approval was provided by the Leeds East Research Ethics committee. All participants gave prior written informed consent. Cases were identified by searching histological databases and discharge records at UK Regional Liver Units for cases of DILI or cholestasis/hepatitis of unknown etiology. Potentially eligible subjects were contacted by their own physicians and invited to participate. Once subjects had signed a consent form, we assessed their eligibility by reference to their case-notes and by causality scoring using ICC[12].

Between October 2004 and December 2007, we collected 51 retrospective and prospective cases of flucloxacillin DILI. Studies were also carried out on a separate group of seven cases who received concomitant treatment with co-amoxiclav in addition to flucloxacillin. A further 16 flucloxacillin cases were recruited after December 2007, including one case recruited from an ADR report to the UK MHRA with approval from their Independent Scientific Advisory Committee. All cases in the study were of self-reported white European ancestry.

Control recruitment. For the GWA study, we selected a control group of 282 matched sample controls using principal component analysis (PCA) (see separate section below) from a total collection of 468 country- and sex-matched controls (POPRES[30]). These controls were also intended for a larger DILI study that includes cases exposed to other drugs. Specifically, the first principal component of the genotypes of all cases and controls in the larger study corresponds to northern-southern separation of European populations. The subjects with negative values are of Northern European origin and all flucloxacillin cases had negative values; therefore, we chose the 282 controls that have negative values for the current study.

For candidate gene analyses, the control group ($n = 64$) consisted of individuals recruited from UK hospitals and general practices (27 male, 37 female, mean age 54.9 years, range 24 to 90 years) who had been prescribed flucloxacillin in the last 5 years but had not reported any symptoms consistent with DILI after completing the course. All were of self-reported white European ancestry.

DNA preparation. We prepared DNA as described previously[17] from cases and drug-exposed controls. Details for DNA preparation from the POPRES collection are described elsewhere[30].

Genome-wide genotyping. Genome-wide genotyping of the DILI cases and POPRES controls was conducted by Expression Analysis Inc. using the Illumina Human1M-Duo BeadChip containing 1,072,820 markers. A SNP was discarded from the case-control association if it had a minor allele frequency less than 0.01, genotype failure rate larger than 0.05, or P value from the test of Hardy-Weinberg equilibrium $<10^{-7}$ within controls. The quality control steps removed 206,421 SNPs. A sample was discarded if it had a genotype call rate <0.9. Sample duplicates and hidden relatedness were investigated on the basis of pairwise identity-by-state (IBS) analysis via PLINK[31]. Only seven pairs of control samples intentionally genotyped in duplicate as part of the quality control procedures were properly identified as duplicates. In the case of each duplicate, the sample with the smaller genotyping call rate was removed from later analysis.

Candidate gene genotyping. Most SNP genotyping was conducted using a fluorescent allele-specific system by Kbiosciences, except for *TNF* rs361525 and *HSPAIL* rs2227956. SNPs were genotyped blind to case-control status, with 20% of samples genotyped in duplicate. *TNF* rs361525 (−238G>A) was genotyped as described previously[32]. *HSPA1L* rs2227956 was genotyped by PCR-RFLP analysis using the primers GTCCCTGGGGCTGGAGACGG and GTGATGATAGGGTTACACATCTGCT and the general PCR methods described previously[33]. The 627-bp product was digested with 1 unit *Nco*I (New England Biolabs) at 37 °C overnight and then analyzed by agarose gel electrophoresis on a 1% gel with TBE as buffer. Samples positive for the T variant gave bands of 354 and 273 bp, whereas the C variant was not digested.

HLA-*DRB1*, *DQB1* and *B*5701* genotyping was conducted using commercial kits (Dynal AllSet SSP DR "low resolution," Dynal AllSet SSP DQ "low resolution," Dynal AllSet SSP DRB1*07/09 and Dynal AllSet Gold SSP B17 high resolution kits from Invitrogen). Initially, all samples were analyzed using the DR "low-resolution" kit. Samples positive for *DRB1*07* by low resolution were further subtyped to determine *DRB1*07* alleles (using the DRB1*07/09 kit) and were also genotyped to determine the *DRB1*07-DQB1* haplotype (using the SSP "DQ" low resolution kit). Samples were genotyped for *B*5701* using the Gold SSP B17 high-resolution kits. We assigned all HLA genotypes according to a standard proforma and checked gels by a second interpreter for validity of assignments.

GWA control selection and correction for population stratification. For the GWA study, we used a control group of 468 population country- and sex-matched controls from the Population Reference Sample (POPRES) collection[34]. The controls were originally selected for a larger DILI study that included cases exposed to other drugs and were chosen to match cases on sex and population origin in so far as possible. To control for population stratification, we used a principal components analysis (PCA)-based approach implemented in the EigenSoft suite of programs[35]. Specifically, the first principal component corresponds to northern-southern separation of European populations. By comparison to the POPRES sample of known origin, the PCA confirmed that the white European cases were all of Northern and Western European origin. To assess and correct for population stratification, we applied PCA to all subjects based on the genome-wide genotyping results (**Supplementary Fig. 1**). We observed two clusters of controls, separated by the sign of the first principal component. All cases had a negative projection on this axis, corresponding to northern and western European ancestry. We thus selected the subset of 282 controls with negative first principal component for single-marker association analysis.

Statistical analysis. For primary analysis of the GWA, we tested the statistical significance of association to each marker with Cochran-Armitage trend test. We further examined results of the following tests: the 1-degree-of-freedom (d.f.) Fisher's exact test for allelic association, 2-d.f. genotypic test, 1-d.f. dominant model test, 1-d.f. recessive model test and an additive model with logistic regression. The genomic inflation factor (1.02) was estimated from the median χ^2 statistics of trend test[14]. All genome-wide analyses were carried out with PLINK[31]. Additional markers with suggestive significance were selected by P-value cutoff of 5×10^{-6}, chosen based on an excess of markers with an association χ^2 score >20.83.

The distribution of HLA class II genotypes in cases and controls was assessed using Fisher's exact test. Following convention, P values for HLA class II genotypes were corrected (P_c) for multiple testing according to Svejgaard and Ryder[36]. We evaluated differences in genotype distribution in non-HLA class II candidate genes between cases and controls by using Fisher's exact test. P values for non-HLA class II genotyping were not corrected for multiple testing. To estimate the individual effect of one locus over others, we carried out conditioning analysis using the program Unphased[37] by adding each locus in turn with a likelihood ratio χ^2 test used to assess the significance of the locus-specific effect.

Logistic regression model to test interaction. To test whether there was an interaction effect between *HLA-B*5701* and other markers, we used a general logistic linear model:

$$Y = b_0 + b_1 \times G_S + b_2 \times G_H + b_3 \times G_S G_H + e$$

In the formula, Y is the logit of case-control status, G_H is the genotype of rs2395029, G_S is the genotype of other SNPs, b_1 and b_2 are additive effects of the SNPs respectively, and b_3 is the interaction effect. We found no genome-wide significant associations under this model.

URLs. Allele frequencies, http://www.allelefrequencies.net/.

30. Nelson, M.R. *et al.* Genome-wide approaches to identify pharmacogenetic contributions to adverse drug reactions. *Pharmacogenomics J.* **9**, 23–33 (2008).
31. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
32. Grove, J., Daly, A.K., Bassendine, M.F. & Day, C.P. Association of a tumor necrosis factor promoter polymorphism with susceptibility to alcoholic steatohepatitis. *Hepatology* **26**, 143–146 (1997).
33. Daly, A.K., King, B.P. & Leathart, J.B. Genotyping for cytochrome P450 polymorphisms. *Methods Mol. Biol.* **320**, 193–207 (2006).
34. Nelson, M.R. *et al.* The Population Reference Sample, POPRES: a resource for population, disease, and pharmacological genetics research. *Am. J. Hum. Genet.* **83**, 347–358 (2008).
35. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
36. Svejgaard, A. & Ryder, L.P. HLA and disease associations: detecting the strongest association. *Tissue Antigens* **43**, 18–27 (1994).
37. Dudbridge, F. Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Hum. Hered.* **66**, 87–98 (2008).