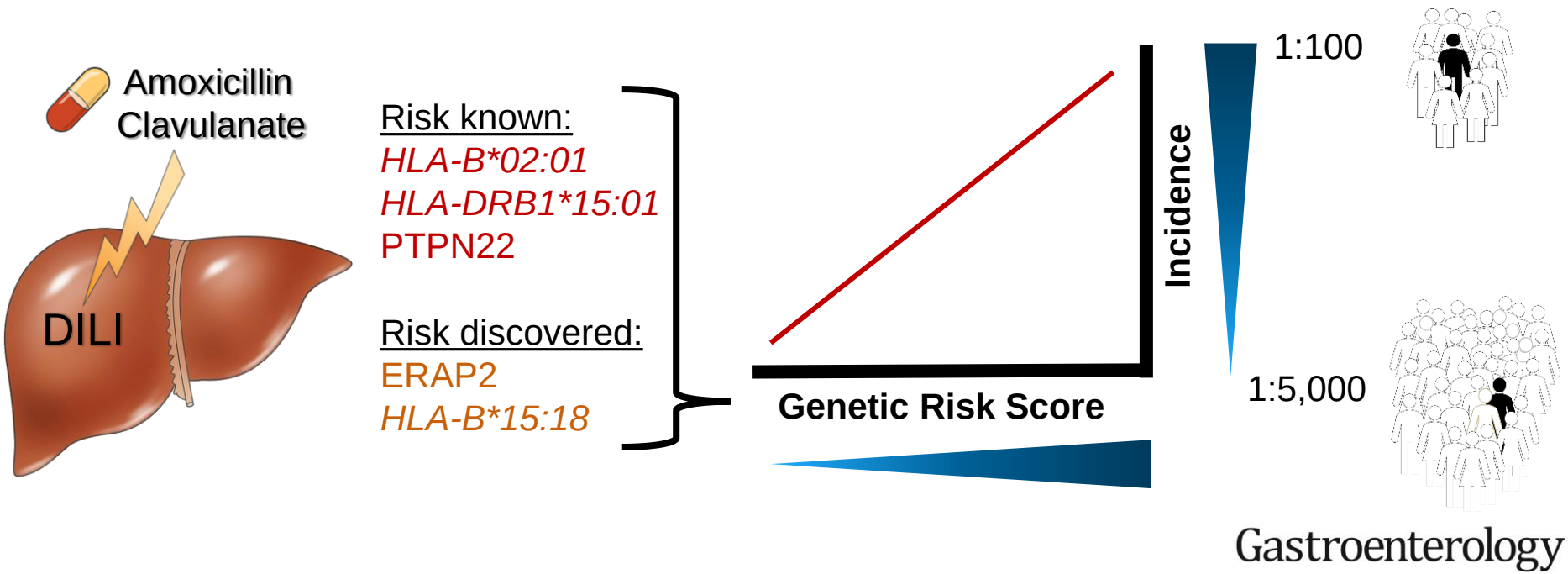
What You Need to Know:

BACKGROUND AND CONTEXT: The diagnosis of amoxicillin-clavulanate drug-induced liver injury (AC-DILI) can be challenging.
NEW FINDINGS: A variant in a gene involved in antigen processing, ERAP2, and a novel Class 1 HLA allele were identified as risk factors and incorporated into a polygenic risk score highly specific for AC-DILI.
LIMITATIONS: The risk score accounts for ~13% of the overall AC-DILI risk but out-performed clinical risk factors.
IMPACT: The polygenic risk score should assist in AC-DILI causality assessment and provides an example of how this approach could improve causality assessment and ultimately risk management of DILI due to other drugs.

Lay summary:

We describe a way to use genetic testing to help identify the rare patients susceptible to liver injury due to treatment with the antibiotic amoxicillin-clavulanate (Augmentin).

Graphical Abstract



Identification of reduced ERAP2 expression and a novel HLA allele as components of a risk score for susceptibility to liver injury due to amoxicillin-clavulanate

Short Title: *ERAP2*, *HLA-B*15:18* and a DILI risk score

Paola Nicoletti[1#], Andrew Dellinger[2], Yi Ju Li[2,3], Huiman X. Barnhart[2,3], Naga Chalasani[4], Robert J. Fontana[5], Joseph A. Odin[6], Jose Serrano[7], Andrew Stolz[8], Amy S. Etheridge[9], Federico Innocenti[9], Olivier Govaere[10], Jane I. Grove[11], Camilla Stephens[12], Guruprasad P. Aithal[11], Raul J. Andrade[12], Einar S. Bjornsson[13,14], Ann K. Daly[10], M. Isabel Lucena[12] and Paul B. Watkins[9,14 15 #]

on behalf of Drug-Induced Liver Injury Network (DILIN), iDILIC and ProEuroDILI investigators

[1] Department of Genetics and Genomic Science, Icahn School of Medicine at Mount Sinai, New York, New York;
[2]Duke Molecular Physiology Institute, Duke University, Durham, North Carolina
[3]Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina
[4]Division of Gastroenterology and Hepatology, Indiana University School of Medicine, Indianapolis, Indiana;
[5]University of Michigan, Ann Arbor, Michigan;
[6] Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, New York;
[7]National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, Maryland;
[8] University of Southern California, Los Angeles, California;
[9] UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina
[10] Translational & Clinical Research Institute, Newcastle University, Newcastle upon Tyne, UK;
[11]Nottingham Digestive Diseases Centre and National Institute for Health Research (NIHR) Nottingham Biomedical Research Centre at the Nottingham University Hospital NHS Trust and University of Nottingham, Nottingham, UK;
[12]UGC Digestivo, Servicio de Farmacología Clínica, Instituto de Investigación Biomédica de Málaga (IBIMA), Hospital Universitario Virgen de la Victoria, Universidad de Málaga, Málaga, Spain; Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd);
[13] Department of Internal Medicine, Landspitali University Hospital, Reykjavik, Iceland;
[14]Faculty of Medicine, University of Iceland, Reykjavik Iceland
[14,15]University of North Carolina Institute for Drug Safety Sciences, University of North Carolina at Chapel Hill.

[#]Authors share co-senior authorship

Formatted: Superscript

Grant support

The DILIN (https://dilin.dcri.duke.edu/) is supported by the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health (NIH) as a Cooperative

Agreement (U01s) under grants: U01-DK065176 (Duke), U01-DK065201 (UNC), U01-DK065184 (Michigan), U01-DK065211 (Indiana), U01DK065193 (UConn), U01-DK065238 (UCSF/CPMC), U01-DK083023 (UTSW), U01-DK083027 (TJH/UPenn), U01-DK082992 (Mayo), U01-DK083020 (USC), U01-DK100928 (Icahn). Additional funding is provided by CTSA grants UL1 RR025761 (Indiana), UL1 RR025747 (UNC), and UL1 UL1-RR024986 (UMich). The University of Nottingham sample collection was funded by the National Institute for Health Research (NIHR) Nottingham Digestive Diseases Biomedical Research Unit at the Nottingham University Hospitals NHS Trust and University of Nottingham. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. GPA is the gastrointestinal and liver disorder theme lead for the NIHR Nottingham BRC (Reference no: BRC-1215-20003). The Spanish DILI Registry is partly funded by the Spanish Medicine Agency, Fondo Europeo de Desarrollo Regional - FEDER (FIS PI16_01748, PI15 _01440). CIBERehd is funded by Instituto de Salud Carlos III. iDILIC and ProEuroDILI collections were also supported by the COST Action "CA17112 - Prospective European Drug-Induced Liver Injury Network" supported by COST (European Cooperation in Science and Technology)(www.cost.eu).

Abbreviations: Allele frequency (AF); Amoxicillin-clavulanate (AC), DILI (drug-induced liver injury); genome wide association study (GWAS); Genetic Risk Score (GRS); Homozygote (HH); Human leukocyte antigen (HLA); Major Histocompatibility Complex (MHC); Odd Ratio (OR); principal component analysis (PCA); Roussel Uclaf Causality Assessment Method (RUCAM); Single Nucleotide Polymorphism (SNP);

Correspondence to: Paola Nicoletti, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai One Gustave Levy Place, New York, NY, USA 10029. Email: paola.nicoletti@mssm.edu

Disclosures: The authors disclose the following: Drs. Chalasani, Fontana, Nicoletti and Watkins report consulting agreements and/or research grants with several pharmaceutical companies but none represent potential conflicts for this paper. The DILIN causality committee considers potential conflicts while assigning cases for adjudication to individual investigators. Dr. Innocenti is an AbbVie employee, receives stocks from AbbVie, this work was conceived when he was a

faculty at the University of North Carolina at Chapel Hill, and this work does not represent a potential conflict of interest. The remaining authors disclose no conflicts.

Author contribution:

Study concept and design: Paul B. Watkins, Paola Nicoletti

Case recruitment and data acquisition: Paul B. Watkins, Naga Chalasani, Robert J. Fontana, Jose Serrano, Andrew Stolz, Huiman X. Barnhart, Guruprasad P. Aithal, Einar Bjornsson, Raul J. Andrade, M. Isabel Lucena, Joseph A. Odin, Ann K. Daly and others from DILIN and the European consortia.

Sample preparation and laboratory analysis: Oliver Govaere, Jane I. Grove, Camilla Stephens

Case adjudication: Paul B. Watkins, Robert J. Fontana, Andrew Stolz, Naga Chalasani, Joseph A. Odin, Jose Serrano, and the DILIN Investigators, Guruprasad P. Aithal, Einar Bjornsson and Raul J. Andrade for the other European consortia;

Data analysis and interpretation: Paola Nicoletti, Andrew Dellinger, Amy S. Etheridge, Federico Innocenti, Yi-Ju Li, Ann K. Daly, Paul B. Watkins

Writing the manuscript: Paola Nicoletti, Paul B. Watkins

Abstract

BACKGROUND: Drug-induced liver injury (DILI) due to amoxicillin–clavulanate (AC) has been associated with *HLA-A*02:01*, *HLA-DRB1*15:01* and rs2476601, a missense variant in *PTPN22*. The aim of this study was to identify novel risk factors for AC-DILI and to construct a genetic risk score (GRS).

METHODS: Transcriptome-wide association (TWAS) and genome-wide association (GWAS) analyses were performed on 444 AC-DILI cases and 10,397 population-based controls of European descent. Associations were confirmed in a validation cohort (n=133 cases and 17,836 population-based controls). Discovery and validation AC-DILI cases were also compared to 1358 and 403 non-AC-DILI cases.

RESULTS: TWAS revealed a significant association of AC-DILI risk with reduced liver expression of *ERAP2* ($P = 3.7x10^{-7}$), coding for an aminopeptidase involved in antigen presentation. The lead eQTL SNP, rs1363907 (G), was associated with AC-DILI risk in the discovery (OR[95%CI] = 1.68 [1.23-1.66] $P = 1.7x10^{-7}$) and validation cohorts (OR[95%CI] = 1.2 [1.04-2.05] P = 0.03), following a recessive model. We also identified *HLA-B*15:18* as novel AC-DILI risk factor in both discovery (OR[95%CI] = 4.19 [2.09-8.36)] $P = 4.9x10^{-5}$) and validation cohorts (OR[95%CI] = 7.78 [2.75-21.99] P = 0.0001). GRS, incorporating rs1363907, rs2476601, *HLA-B*15:18*, *HLA-A*02:01*, *HLA-DRB1*15:01*, was highly predictive of AC-DILI risk when cases were analyzed against both general population and non-AC-DILI control cohorts. GRS was the most significant predictor in a regression model containing known AC-DILI risk clinical characteristics and significantly improved the predictive model.

CONCLUSIONS: We identified novel associations of AC-DILI risk with *ERAP2* low expression and with *HLA-B*15:18*. GRS based on the five risk variants may assist AC-DILI causality assessment and risk management.

Keywords: Amoxicillin-clavulanate; DILI, ERAP2; *HLA-B*15:18*; GWAS.

Introduction

Idiosyncratic drug-induced liver injury (DILI) is a rare adverse drug reaction that is an important cause of morbidity and acute liver failure[1, 2]. Diagnosing DILI and identifying which drug is responsible in the patient receiving multiple medications is often challenging[3]. Amoxicillin-clavulanate (AC) is a very common cause of idiosyncratic DILI[1, 4]. Many patients with AC-DILI present 1 to 4 weeks after discontinuation of AC making it difficult to recognize AC as a causative drug[5]. While most cases resolve with discontinuation of AC treatment, full recovery may require many months and liver failures have been reported[4]. Furthermore, invasive medical testing including liver biopsies and endoscopic retrograde cholangiopancreatography (ERCP) are frequently undertaken in older individuals presenting with jaundice and weight loss that is eventually attributed to AC-DILI[4].

Prior GWAS studies identified *HLA-A*02:01* and *HLA-DRB1*15:01* as genetic risk factors for AC-DILI among individuals of European descent[6-8]. A subsequent GWAS study identified a variant (rs2476601) of *PTPN22* to be a risk factor for all cause DILI, with AC-DILI as a major contributor to this association[8]. *PTPN22* is believed to be involved in regulating T-cell receptor signaling and the variant associated with DILI has also been associated with risk for certain autoimmune diseases[9], further supporting the involvement of the immune system in AC-DILI.

The aim of the current study was to identify novel genetic risk variants for AC-DILI and to develop a sensitive and specific multi-locus genetic risk score that could help in causality assessment and ultimately risk management. We applied both transcriptome-wide association (TWAS) and genome-wide association (GWAS) analysis approaches on the largest AC-DILI cohort published up to date. We provide additional support for the role of adaptive immunity in AC-DILI by discovering and validating novel associations between AC-DILI risk and the expression of *ERAP2*, a gene involved in antigen processing[10], as well as with a new HLA class I allele. The predictive risk associated with the combination of the five identified risk factors was also evaluated, and specificity assessed by comparing to DILI risk due to other drugs.

Materials and Methods

Study cohorts

European AC-DILI Discovery cohort

444 AC-DILI cases collected by DILIN, the Spanish DILI network, DILIGEN, Eudragene, and iDILIC consortia (**Table S1**) were compared to 10,397 European population control samples as included in Cirulli et al.[8] As AC-DILI has 45 in 100,000 prevalence in general population[11], we used a large set of general population controls.

European AC-DILI Validation cohort

133 AC-DILI cases (**Table S2**) were newly recruited for this study by DILIN, iDILIC, Prospective European DILI cohort study (ProEuro-DILI) and the Spanish DILI network. The eligibility criteria and causality assessment were performed as previously described.[12, 13, 8, 14] All participants provided written informed consent, and each study was approved by the appropriate national or institutional ethical review boards. Validation AC-DILI cases were compared to 17,836 European population controls from NCBI dbGaP cohorts, as inferred by principal component analysis (PCA) (**Figure S1**). Details on the dbGaP cohorts are reported in **Tables S2** and their age and sex information in **Table S3**.

European non-AC-DILI discovery and validation cohorts

We gathered two large independent European cohorts of non-AC-DILI individuals to use as controls to compare respectively to the discovery and validation AC-DILI cases. 1,358 non-AC-DILI individuals from Cirulli et al.[8] served as controls in the discovery cohort and 403 newly genotyped non-AC-DILI individuals recruited by DILIN served as controls in the validation cohort. Clinical information is reported in **Table S3**.

Hispanic and African American cohorts

13 AC-DILI cases of African American (AA) and 7 of Hispanic descent were recruited from DILIN. Based on PCA cases were compared with 2919 Hispanic and 5816 AA ancestry-matched controls from PAGE BioME (phs000925.v1.p1).

Genome-wide genotyping.

Genotype data was available for all controls, AC-DILI European discovery cases, 1358 non-AC DILI cases, 10 Hispanic and AA DILIN AC-DILI cases and 9 iDILIC validation cases. [6, 8, 14] The remaining 124 European validation cases, 8 Hispanic and AA DILIN cases and 403 European non-AC-DILI cases were genotyped using the MEGAX Illumina platform at Vanderbilt University (**Table S2)**.

Transcriptome-wide analysis (TWAS).

PrediXcan[15] was used to predict liver gene expression in cases and controls based on GTEX V8 liver gene expression dataset using the default model and parameters. Association testing was performed for each gene using logistic regression correcting for population stratification in R. Based on number genes expressed in the liver GTEX V8, transcriptome-wide statistical significance threshold was estimated as $1.1x10^{-5}$, using Bonferroni correction.

Genetic analysis.

Quality control (QC) checks on the genotype data were performed as described.[8, 14] Genetic ancestry was inferred by principal components (PCs) computed from EIGENSTRAT as described in Li et al[16]. SNP imputation was performed using Michigan Imputation Server[17] (see Supplementary Materials). Four-digit *HLA* alleles were inferred using HIBAG[18] using genotyped data in the MHC region. For both *HLA* alleles and SNPs, association analyses were performed using logistic regression under the additive genetic model with covariate adjustment for PCs (ten PCs in the discovery and four PCs in the validation cohort) in PLINK[19]. When indicated, age and sex were included in the model as covariates. Due to the limited size, we applied Fisher's Exact test for association analysis between AA and Hispanic cases and controls. We set the GWAS and MHC significance *p*-value threshold to $5.0x10^{-8}$ and $4.0x10^{-4}$ (0.05/115 total predicted HLA alleles) to correct for multiple testing. We used conditional analysis to determine which variants are independent within the susceptibility locus by including the most associated variant/s as covariate/s. As published,[8] to assess the joint effect of risk variants discovered to date, multi-marker logistic regression analysis was conducted with adjustment of PCs, where the joint negative carriers were used as the reference group. *ERAP1* haplotypes were reconstructed and tested in Plink using the set of variants and the nomenclature reported by Ombrello and collegues[20]. Meta-analysis was performed using a fixed-effect model in the *metafor* R package. Manhattan plots were

performed with R (Version 3.0.2). Regional plots were drawn by LocusZoom[21]. We investigated the relevance of the significant genetic associations with AC-DILI clinical phenotypes (see Supplementary Materials).

Pairwise SNP-SNP interaction association among the top associated signals was assessed using logistic regression to model the significant risk markers (expressed as dichotomic variables based on carriage status), and their interaction term with adjustment of PCs and Bonferroni correction was applied to correct for multiple comparisons, as proposed[22]. SNP interaction analysis, multivariable, multi-marker logistic regression models including the relevant genetic risk factors and Likelihood-ratio test were carried out using STATA15.

Genetic Risk Score (GRS).

We used the five variants associated with AC-DILI risk to construct a Genetic Risk Score (GRS). We used the discovery dataset to estimate the variant weights, which were the beta-coefficients from the multivariable logistic regression model on these five variants with adjustment of population stratification. The GRS was then formulated as the sum of the product of the estimated beta-coefficient and the number of risk alleles at each locus as shown in the equation below:

$$\mathrm{GRS} = (0.77 \text{ x allele dosage of HLA } - \mathrm{A} * 0201) + (1.16 \text{ x allele dosage of HLA} - \mathrm{DRB1} * 15{:}01) + (1.37 \text{ x allele dosage of HLA} - \mathrm{B} * 15{:}18) + (0.47 \text{ x ERAP2 rs1363907 GG genotype} + (0.48 \text{ x allele dosage of PTPN22 rs2476601 A})$$

Examples of how to calculate the GRS are reported in **Table S4**. The GRS values were divided into ten equal quantiles and the AC-DILI relative risk of each quantile was calculated to estimate the absolute risk of developing DILI after a treatment course with AC, assuming the overall incidence is 1:2,350 AC treated patients.[11] We tested the GRS associations with AC-DILI in the validation dataset against both population controls and both non-AC-DILI cohorts. The Youden's index[23] was used to determine the optimal diagnostic threshold. The predictive power of the model was evaluated by area under receiver operator characteristic (ROC) curve (AUC) using STATA15

HLA genotyping.

Genotypes of the most significant HLA alleles were confirmed using high resolution genotyping on the carriers by Histogenetics (Ossining, New York) for the iDILIC samples[14], and at Vanderbilt University using Illumina MiSeq for the DILIN cases[16].

Results

Genome-wide association study identifies a novel HLA risk allele for AC-DILI

GWAS was performed on 7,747,658 well-imputed or genotyped SNPs comparing 444 European ancestry cases and 10,397 matched population controls (**Figure S1A**). Clinical characteristics of the AC-DILI cases are reported in **Table 1**. Only variants in the MHC region were found to have genome-wide significant p-values, led by rs9268927 (A, OR = 3.13 95%CI [2.68-3.65] $P=6.3x10^{-47}$ which is a proxy of *HLA-DRB1*15:01* ($r^2 = 0.96$) and rs2734966 (G, OR = 2.11 95%CI [1.83-2.42] $P = 4.2x10^{-26}$ which is a proxy of *HLA-A*02:01* ($r^2 = 0.96$). (**Figure S2** and **Table S5**). Both are previously reported risk factors for AC-DILI[6]. Subsequent genome-wide analysis conditional on the genotypes of *HLA-DRB1*15:01* and *HLA-A*02:01* did not reveal additional genome-wide significant marker associations (**Figure S3).** For imputed HLA genotypes, *HLA-DRB1*15:01* and *HLA-A*02:01* together with their associated known haplotype alleles *HLA-DQB1*06:02* and *HLA-B*07:02*, respectively were the most significant HLA alleles as expected. Conditional analysis on *HLA-DRB1*15:01* and *HLA-A*02:01* identified a novel and independent association with *HLA-B*15:18* which was MHC-significant (OR = 4.19, 95%CI [2.09-8.36)], $P = 4.9x10^{-5}$, **Table S5**). Among the AC-DILI cases, ten cases (2%) carried this rare *HLA-B* allele when just three were expected based on reported European population frequency. All nine *HLA-B*15:18* carriers, with DNA available for confirmatory HLA typing had 100% concordance between the predicted and typed *HLA-B*15:18* allele. Finally, the previously reported association[8] between DILI from multiple drugs and *PTPN22* rs2476601 missense variant showed a similar independent association trend (OR = 1.43 95%CI [1.23-1.65] $P = 2.2 x10^{-6}$).

We also analyzed the most associated GWAS markers reported in **Table 2** in an independent validation cohort of 133 AC-DILI cases that we compared with a new cohort of 17,836 European population-based controls (**Figure S1B**). This analysis confirmed the role of *HLA-B*15:18* as an

AC-DILI risk factor (OR = 5.24[1.85-14.81] P= 0.002, **Table 2**) with 2% of the AC-DILI validation cases carrying this allele. All carriers were confirmed by direct HLA typing. The association was still significant when correcting for age and sex (OR = 7.31 [2.24-23.83] P = 0.0009) or when correcting for *HLA-DRB1*15:01* and *HLA-A*02:01* (OR = 7.78 [2.75-21.99] P = 0.0001). The association with rs2476601 (PTPN22) was not confirmed in this smaller cohort (**Table 2**).

***ERAP2* expression is associated with AC-DILI risk**

We next imputed genome-wide gene expression profiles of each individual in our discovery cohort using the PrediXscan algorithm,[15] with the liver specific eQTL genetic profile GTEX V8 as the reference panel. We then performed a Transcriptome Wide Association Study (TWAS) comparing the predicted gene expression between cases and controls. TWAS analysis revealed that several genes within the MHC region were differentially expressed, which mirrored the long LD block associated with the known HLA risk alleles (**Figure S4**), as demonstrated by the abolition of the PrediXcan MHC signals upon conditioning on *HLA-DRB1*15:01* and *HLA-A*02:01* (**Figure 1**). Outside the MHC region, only the predicted expression of one gene, *ERAP2*, was significantly different in controls compared with cases ($P = 3.7x10^{-7}$, **Figure S4**). This significant difference in predicted *ERAP2* gene expression was independent of the effect of the HLA risk alleles as tested in the conditional analysis ($P_{conditional} = 4.5 x10^{-6}$, **Figure 1**).

We identified rs1363907 (G) as the single SNP with the strongest association with AC-DILI in GWAS analysis in the *ERAP2* region, driving the association with *ERAP2* expression. rs1363907 (G) was associated with increased risk of AC-DILI (**Table 2**). rs1363907 (G) is part of a long LD region encompassing *ERAP2* and near LNPEP genes. The long LD region includes both intronic and exonic variants (**Figure S5**). The genomic region in *ERAP2* has two major haplotypes called Hap A and Hap B. Evidence suggests that Hap B codes for a truncated ERAP2 mRNA that is rapidly degraded mainly although not uniquely due to a splicing variant (rs2248374, G) located in the ERAP2 LD region[24]. rs1363907 is in linkage with rs2248374 and other suggested causal variants responsible for the mRNA truncation (r^2 => 0.80, **Table S6).** Therefore, rs1363907 (G) was considered a proxy for Hap B.

We confirmed the association of rs1363907 (G) and ERAP2 expression in an independent liver eQTL dataset derived from 1,183 liver samples[25]. Using this large eQTL liver-specific dataset, we identified that only individuals carrying two copies of the rs1363907 (G) risk allele showed a consistently lower liver expression of *ERAP2* mRNA compared to the other genotype groups (**Figure S5**), suggesting a recessive model (**Figure S6**). Consistent with this observation, the increased risk of AC-DILI was associated only with homozygous status of rs1363907 (G) in the multivariable analysis (**Table S7**). The recessive model was also a better model when all the SNPs in the LD region were considered (Figure S7). rs1363907 (G) was not correlated with the expression of any HLA loci in GTEX V8 dataset (**Figure S8**). The association between homozygous status of rs1363907 (G) and risk of AC-DILI was confirmed in the European validation cohort (OR = 1.43, 95%CI [1.01-2.03] P = 0.04, **Table 3**) with an overall significance of $4.3x10^{-7}$ in the meta-analysis. In this cohort, the rs1363907 (G) association was still significant when corrected by sex and age (OR = 1.40, 95%CI [1.04-1.88] P = 0.03). Moreover, allele frequency rs1363907 was similar between males and females in the GnomAD database (not shown).

The relationship between patients homozygous for rs1363907 (G) and AC-DILI was evaluated further focusing upon clinical phenotypes. The homozygote status of rs1363907 (GG) was not different among the three DILI clinical presentation phenotypes (e.g. hepatocellular, cholestatic or mixed) as well as other clinical characteristics (see Supplemental Materials) but was associated with a longer time to DILI onset after starting AC treatment (latency) than those carrying one or no rs1363907(G) allele (25 vs 21 days, P = 0.03). A similar trend was observed in the independent European validation cohort (26 vs 24 days).

From our previously published DILI cohort,[8] we examined 13 DILI drug groups with at least 20 European cases to determine whether rs1363907 (G) was an AC-specific risk factor. Each drug group was compared to the same set of controls. A significantly higher frequency of the rs1363907 (G) risk allele was observed only in DILI cases attributed to amoxicillin (n = 20), but not in DILI cases attributed to other causal drugs, including flucloxacillin (n = 195, **Figure S9**). Risk of

amoxicillin DILI was associated more strongly with *HLA-B*58:01* than with *HLA-A*02:01* (AF = 0.35) or with *HLA-DRB1*15:01*, as was reported previously in an amoxicillin DILI cohort[26].

Evaluation of *ERAP1* association with AC-DILI

ERAP1 encodes an enzyme with a similar function to *ERAP2*. *ERAP1* has 10 common haplotypes which were identified in cases and controls (see Methods). In the discovery cohort, *ERAP1* Hap2 and Hap8 were significantly associated with AC-DILI risk and Hap10 was marginally associated. Hap2 was a risk factor for AC-DILI (OR = 1.41; P = 4.9×10^{-5}) whereas Hap10 and Hap8 were protective (**Table 2**). Similar allele frequencies and effect sizes for those haplotypes were observed in the validation cohort with Hap2 approaching statistical significance (**Table 2**). Due to the strong linkage disequilibrium between the two genes, conditioning on rs1363907 (*ERAP2*) reduced the *ERAP1* associations with AC-DILI, and only Hap2 still showed a marginally significant residual effect (OR = 1.26; P = 0.01; **Table S8)**.

Predictive value for AC-DILI of the five markers and their combinations in a multiple regression analysis

The predictive value of three validated HLA risk alleles, rs2476601 (A) (*PTPN22*) and homozygous rs1363907 (G) (*ERAP2*) and their combinations were further evaluated. A multivariable logistic regression analysis established that each of the five markers was independently associated with AC-DILI risk and, except for rs1363907 and *HLA-B*15:18*, each marker showed a co-dominant effect (**Table 3**). The same model was applied to the validation cohort and supported the independent and codominant effects of the markers, including a trend for the PTPN22 variant (**Table 3**). We also applied a multi-marker logistic regression model to estimate the effect size of combinations of risk alleles compared to the baseline joint-negative group. Among most significant combinations (**Table S9**), co-occurrence of one allele of *HLA-A*02:01* and *HLA-DRB1*15:01* appears in 7% of the cases and confers a 3.77-fold increase in odds of developing AC-DILI. The presence of both HLA risk alleles and the *PTPN22* variant doubled the odds ratio to about 6-fold and a comparable increase in risk (about 8-fold) was

associated with the presence of homozygous *ERAP2*. Heterozygous status for all four of the risk markers was associated with a 17-fold increase in AC-DILI risk. The risk increases further in cases homozygous for at least one of the risk factors.

The predictive value of the multi-marker logistic regression model was assessed by ROC curve using first the discovery cases. The predictive model had comparable AUC values when these AC-DILI cases were compared to discovery population controls (AUC = 0.76, **Table S10**) and to 1358 European non-AC-DILI cases (AUC = 0.76, **Table S10**), confirming the drug-specificity of the model. Then, the same model was tested using the 133 validation AC-DILI cases compared to validation population controls, and to 403 European non-AC-DILI cases confirming that the model was predictive and specific for DILI (AUC = 0.79 and AUC = 0.70, **Table S10**). The model accounted for about 13% of the AC-DILI susceptibility (R^2 of the model, **Table S10**).

Risk allele interaction analysis

To assess possible interaction effects among the five risk factors, a systematic pairwise snp-snp interaction analysis[22] was carried out (**Table S11)**. A more than additive effect between *HLA-DRB1*15:01* and *HLA-A*02:01* was the only interaction (OR = 2.0; P = 0.0002) statistically significant after multiple testing correction. However, both the rs2476601 (*PTPN22*) and homozygous rs1363907 (*ERAP2, HH*) showed consistent trends for more than additive risk effects when paired with any of the HLA risk alleles, with OR = 2.0 P = 0.03 for rs2476601 and OR = 1.7; P = 0.04 for rs1363907. Notably, homozygous rs1363907 (*ERAP2* Hap B) demonstrated a stronger trend for more than additive risk effect in association with *HLA-A*02:01* (OR = 1.62; P = 0.03) than with *HLA-DRB1*15:01* (OR = 1.18; P = 0.40). There was no interaction effect detected between the *PTPN22* variant and *ERAP2* Hap B. In the validation cohort, the interaction analysis supported similar trends for these interactions (**Table S11)**.

Genetic Risk Score and AC-DILI risk

In order explore optimal use of the identified genotypes to prospectively assess AC-DILI risk in patients, we created a Genetic Rick Score (GRS). The AC-DILI GRS was estimated for each

individual based on the weighted sum of the coefficients from a multivariable analysis of the five validated AC-DILI risk markers (see Methods). GRS ranged from 0 to 4.81 in the discovery and validation cohorts (**Figure S10**). We explored the predictive value of GRS for AC-DILI risk in the validation cohort where the strong association was confirmed (OR = 2.2, 95%CI [1.88-2.69] P = $9.5x10^{-19}$). To quantify the risk for individuals carrying an increasing burden of risk alleles, we divided discovery and validation subjects into 10 quantiles based on the GRS score ranges. We estimated the relative risk of each quantile of developing DILI as a result of being treated with AC assuming that 1:2,350 patients treated with AC develop DILI[11] . While only about 1:5,000 patients with GRS in the with lowest quantiles are predicted to experience DILI when treated with AC, the incidence increases to greater than 1:500 patients in the highest GRS quantiles (**Table 4**).

Drug-specificity of AC-DILI Genetic Risk Score compared to DILI due to other drugs

Finally, we explored the predictive value of GRS for AC-DILI risk when AC-DILI cases were compared with two independent cohorts of DILI due to other drugs (non-AC DILI). We found that the GRS was strongly associated with AC-DILI when our 444 AC-DILI discovery cases were compared to 1558 non-AC-DILI subjects (OR = 2.68, 95%CI [2.34-3.07] P = $5.7x10^{-46}$). A significant association was also found when we compared the 133 AC-DILI validation cases to a second independent set of 403 non-AC-DILI subjects (OR = 2.06, 95%CI [1.65-2.56] P=$9.7x10^{-11}$). The predictive model had comparable AUC values in both the AC-DILI case/non-AC-DILI control cohorts (AUC = 0.76 and 0.70, respectively, **Table S10**). Finally, we determined the GRS in the 13 largest European DILI drug groups in the Cirulli et al[8] cohort. We found that after multiple testing correction, the score was not significantly correlated with the risk of DILI due to any of the 13 specific causal agents (**Table S12**). Moreover, GRS distinguished high versus low causality cases among cases where AC was the lead implicated drug, and this was not the case for DILI attributed to other drugs (**Figure S11**).

AC-DILI Genetic Risk Score in the causality assessment compared to DILI due to other drugs

We next examined the predictive value of the GRS in addition to clinical risk factors associated with AC-DILI.[4] For this purpose, we compared the 444 discovery AC-DILI cases to 1358 non-AC-DILI DILI cases (called non-AC-DILI discovery cohort) and the 133 replication cases against 403 newly genotyped non-AC-DILI cases (called non-AC-DILI validation cohort). As clinical risk factors for AC-DILI we considered patient age, being male, latency (time [days] from onset of AC treatment to recognition of DILI), and injury type (cholestatic or mixed vs hepatocellular). In the multivariable logistic model of the AC-DILI cases/non-AC-DILI discovery cohort, GRS and age, latency, sex and injury type and GRS were significantly associated with AC-DILI (**Table 5**). In the AC-DILI cases/non-AC-DILI validation cohort only age, latency and GRS were significant (**Table 5**). Among all the risk factors, the GRS was the most significant risk predictor in both multivariable models (**Table 5**). In both AC-DILI case/non-AC-DILI discovery and validation cohorts, the multivariable logistic model including clinical variables and GRS (AUC using the discovery cohort = 0.84 and AUC using the validation cohort = 0.85, **Table S10** and **Figure S12**) was significantly more predictive than the model including the only clinical variables (AUC = 0.76 and AUC = 0.80) when tested by Likelihood-ratio test ($P = 3.1x10^{-44}$ and $P = 1.1x10^{-10}$ respectively).

Clinical Application of the GRS

To show the applicability of GRS in clinical practice, we estimated the optimal diagnostic cut off for the GRS by the Youden index commonly used for setting thresholds for diagnostic utility on medical tests, at 1.58 (95%CI [1.21-1.95]). GRS Youden index had similar specificity (SP) and sensitivity (SN) in the discovery and validation cohorts when cases were compared to population controls (SP = 79% and SN = 56% for discovery and SP = 77% and SN = 47% for the validation cohorts) and when AC-DILI cases were compared to non-AC-DILI cohorts (SP = 76% and SN = 56% for discovery and SP = 77% and SN = 47% for the validation cohorts), confirming the score specificity for AC-DILI. In fact, 56% of the AC cases had GRS > 1.58 compared to 21% and 23% of the controls and non-AC-DILI cases, respectively (**Table S12**). Moreover, GRS was determined in the 20 cases in the DILIN registry in which co-exposure to AC was noted, but AC was not considered to be the cause of the DILI (and therefore not included in our AC DILI cohorts). The GRS was > 1.58 in 5 of these 20 cases (25%) (**Table S13)**. Each of these cases underwent repeat

adjudication by hepatologists experienced in DILI (PBW, RJF, JAO, JS and AS). In the case with the highest GRS, knowledge of the GRS resulted in unanimous consensus that the causal drug was AC and not azithromycin (see footnote in **Table S13**).

The five AC-DILI risk markers in African American and Hispanic populations

We evaluated the association of the five markers in AA and HSP populations, which are reported in **Table S14**. Although with limited power, we found that *HLA-DRB1*15:01* risk allele had a higher frequency in cases across the three major ethnicities compared to population-based controls. Compared to population controls, the 13 AA cases had higher frequencies of *HLA-A*02:01* (AF 0.12 vs 0.19), ERAP2 rs1363907 (AF 0.65 vs 0.77) and *HLA-DRB1:15:03* (AF 0.23 vs 0.12) which has a similar peptide binding profile to *HLA-DRB1:15:01*.[27] The 7 Hispanic cases only revealed a trend for carrying the *PTPN22* risk allele compared to controls. All these differences were not statistically significant due to the low number of cases. For non-European individuals, predictions for the *ERAP1* haplotypes could not be estimated because of the missing data of several relevant proxy SNPs.

Genetic evaluation of European AC-DILI cases not carrying the five risk factors

Since no clinical differences were detected between the AC-DILI cases not carrying any the five risk AC-DILI markers (non-carriers) and the rest of the AC-DILI cases, we evaluated if the negative carriers would be enriched in additional novel HLA risk alleles. We compared the 43 AC-DILI non-carriers to the control population. The cases were enriched in *HLA-B*18:01* (AF 0.14 vs AF 0.05) and *HLA-B*35:01* (AF 0.13 vs AF 0.05, **Table S15**). The trend association with *HLA-B*18:01* was also observed in the validation cohort with *HLA-B*18:01* frequency higher among the 18 non-carriers than controls (AF 0.08 vs AF 0.04).

Discussion

In the current study, TWAS and GWAS analyses were performed on the largest existing cohort of European AC-DILI cases resulting in the identification of novel independent genetic associations

of AC-DILI with rs1363907 in *ERAP2*, a gene known to be involved in antigen processing[10], and with an additional HLA allele, *HLA-B*15:18*. These associations were confirmed in a validation cohort of 133 additional European AC-DILI cases. We demonstrated that the predicted risk associated with the combination of the five genetic risk loci identified in this and previous studies[6, 8] was specific for AC-DILI and could be helpful in the diagnosis of AC-DILI.

The *ERAP2* association was identified using the TWAS approach where the predicted liver expression of *ERAP2* was significantly lower in the AC-DILI cases than controls. *ERAP2* has two major haplotypes (Hap A and Hap B) and our lead SNP rs1363907 appears to tag Hap B. Hap B undergoes differential splicing of exon 10, leading to generation of a premature stop codon resulting in a truncated *ERAP2* mRNA that is rapidly degraded[24]. This rapid degradation presumably accounts for the reduction of total *ERAP2 mRNA* we detected as a risk factor in the PrediXscan approach. Interestingly, for 1,183 livers that have undergone eQTL mapping[25], the expression of *ERAP2* mRNA was significantly lower only in individuals carrying two copies of the Hap B, supporting a recessive model. Consistent with this observation, the increased risk of AC-DILI was associated only with homozygous carriage of Hap B. In the heterozygous state, compensatory mechanisms may exist to maintain *ERAP2* expression in the liver. From a clinical perspective, Hap B homozygous state seemed to be associated only with a small increase in latency to onset of DILI and not with other clinical characteristics of the liver injury.

ERAP2 is present in the endoplasmic reticulum of all cells where it functions to trim peptides generated by the proteasome so that they will fit into the peptide binding pocket of newly synthesized Class I HLA molecules prior to their transit to the cell surface[10] (**Figure 2**). It has been reported that ERAP2 may over-trim some peptides such that they are unable to bind HLA molecules, supporting a role for *ERAP2* in controlling which peptides have antigenic potential[10]. *ERAP2* haplotype B has been associated with risk of several autoimmune diseases, including birdshot chorioretinopathy, ankylosing spondylitis, and some forms of psoriasis, all of which have strong risk associations with specific Class I HLA alleles[10]. However, unlike the case with AC-DILI, the risk of these diseases is generally lower in those carrying Hap B rather than higher[10]. In the case of birdshot chorioretinopathy, it has been recently shown that ERAP2 generates a profile of peptides that specifically bind to the class I allele associated with risk of this disease (HLA-

A29)[28]. The lower risk of this disease in those carrying ERAP2 Hap B may therefore result from reduced production and presentation of the culprit neoantigens by HLA-A29. The association of Hap B with increased AC-DILI risk may suggest that the culprit neoantigens are preferentially created by the other endoplasmic reticulum enzyme that trim peptides, ERAP1 **(Figure 2)**.

Among the drugs most frequently associated with DILI, only AC and amoxicillin DILI were significantly associated with *ERAP2* Hap B. Amoxicillin DILI risk was not associated with our GRS and the clinical presentation is not identical to that of AC-DILI[4]. Although pharmacy records were not checked in most cases, it seems unlikely that many cases were erroneously attributed to amoxicillin and actually due to AC. Clavulanate has long been considered the main component causing AC-DILI (LiverTox https://www.ncbi.nlm.nih.gov/books/NBK548517/). However, the recent observation that amoxicillin-modified peptides bind *HLA-DRB1*15:01* and/or *DQB1*06:02* and activate T cells obtained from an AC-DILI patients[29] may support a role of amoxicillin in at least some of the AC-DILI cases.

Various haplotypes of *ERAP1* have also been associated with certain autoimmune diseases[30]. We found that increased AC-DILI risk was associated *ERAP1* Hap 2 which produces an enzyme with relatively high activity, while the lower activity products of *ERAP1* Hap 8 and 10 showed a protective effect. These observations further support the notion that the culprit neoantigens sparking AC-DILI are likely produced by ERAP1. However, when conditioned on carriage of ERAP2 Hap B, only the risk association with *ERAP1* Hap 2 remained marginally significant, reflecting the long LD block encompassing the two genes. Additional studies will be needed to identify independent risk associations with ERAP1 haplotypes.

We identified a gene-dose effect *HLA-A*02:01* and *DRB1*15:01* for AC-DILI risk, and our data support this also for the *PTPN22* missense variant. In addition, we confirmed the more than additive effect of carrying both the *HLA-A*02:01* and *DRB1*15:01* risk alleles[6], doubling the sum of risks associated by each individual allele. We also found a trend for a more than additive increase in DILI risk when *ERAP2* Hap B co-occurred with *HLA-A*02:01* but not with *HLA-DRB1*15:01*. Although not significant after multiple comparison correction, these observations

are consistent with the known role of ERAP enzymes to provide suitable peptides for binding to Class 1 but not Class 2 molecules[10].

Finally, we developed a GRS based on the five identified risk alleles. We included the PTPN22 variant although the risk association was not observed in our validation cohort. This is because the association was GWAS significant in a prior study of all cause DILI, and because a trend towards a gene-dose effect was observed in the validation cohort. We assume that the association would have been validated in a larger cohort. To our knowledge, specific genes involved in absorption, distribution, metabolism or excretion (ADME) of either amoxicillin or clavulonic acid have not been identified, and no pharmacogenetic variants located in or near any major ADME genes were genome-wide significant.

Our observations support the future value of the GRS for identifying AC as the culprit drug in DILI cases, and potentially a role in risk management. ~~In~~ Among European ~~spopulation~~, the GRS was found to be predictive of AC-DILI risk when tested against population controls and was highly specific for AC-DILI when tested against non-AC-DILI cohorts. We estimated that a GRS value above 1.58 would be a minimum value to support the diagnosis of AC-DILI, although up to half true AC DILI cases will have GRS values below this cutoff (~50% sensitivity). The value to health care professionals in making an AC-DILI diagnosis increases the higher the GRS value (increasing specificity although reducing sensitivity). Because the risk of developing DILI during treatment in an average Caucasian has been estimated to be about 1:2,350[11], we were able to use the GRS to estimate an individual risk to be both lower than this average (>1:5,000) or much higher (approaching 1:100). To apply the GRS to causality assessment, we examined cases in the DILIN registry where AC was a concomitant treatment but not felt to be the cause of the DILI. Five of these cases had GRS above the 1.58 threshold, and re-adjudication of the patient with the highest GRS (corresponding to at risk of 1:600) resulted in confident attribution of the DILI event to AC in one of these cases.

AC-DILI has been shown to be more common in elderly men, and to characteristically present as a mixed or cholestatic injury several weeks after initiating AC treatment.[4] Accordingly, we found

that a regression model including age, sex, latency and injury type did differentiate AC-DILI from non-AC-DILI in two independent large case/non-AC-DILI control cohorts. Nonetheless, addition of the GRS to this model significantly improved its sensitivity and specificity in differentiating AC-DILI from DILI due to other drugs.

It is of note that our multi-marker model based on the five risk alleles predicted about 13% of the total risk for AC-DILI, which is about one third of the total DILI risk that has been previously attributed to common genetic variants (approximately 40%[31]). This is a large proportion of total genetic risk relative to most diseases[32]. This finding suggests that DILI, as is the case for other severe adverse drug events, might have an oligogenic architecture with few highly penetrant genetic risk factors. Under this hypothesis, identification of a limited number of additional risk alleles might lead to substantial improvements in the GRS performance, increasing the sensitivity of the GRS score.

Unfortunately, our GRS is unlikely to be useful in AA and Hispanic populations, which appear to have a lower incidence of AC-DILI compared to Europeans[33]. Even though *HLA-DRB1*15:01* is less common in both minority populations compared to Europeans (AF = 0.12), both AA and Hispanic AC-DILI cases showed double the frequency of *HLA-DRB1*15:01* in AC-DILI cases *vs* controls (AF 0.04 vs 0.02 and 0.14 vs 0.06), supporting this HLA Class II allele as a general risk factor for AC-DILI. Interestingly, *HLA-DRB1*15:03* showed an even stronger enrichment in the AA AC-DILI cases (0.23 in cases vs 0.13 in controls). In AA, *HLA-DRB1*15:03* is the most common *HLA-DRB1*15* allele and has a similar peptide binding affinity as *HLA-DRB1*15:01*.[27] Interestingly, *HLA-DRB1*15:03* has an established role in the susceptibility of autoimmune diseases[27] and drug-induced DRESS[34]. AA cases were also enriched in *HLA-DRB1*15:02*, a known risk factor for AC-DILI among people of Indian or Pakistani origin[35]. Due to limited number of cases, no conclusion could be drawn for rs2476601and rs1363907 associations. Larger association studies are needed to confirm whether differences in HLA profiles accounts for differences in AC-DILI susceptibility.

In summary, a common variant in *ERAP2*, a proxy of Hap B, was identified as a novel risk factor for AC-DILI, implicating for the first-time antigen processing as a locus of DILI risk. A rare class

1 HLA risk allele, *HLA-B*15:18* was also identified as a risk factor for AC-DILI. When combined with three previously established genetic risk factors, the resulting five-locus GRS was highly specific for AC-DILI, supporting a future role for genetic testing in DILI causality assessment and potentially risk management.

Acknowledgments

We acknowledge the contributors to sample collection via DILIN, iDILIC, the Spanish DILI registry, EUDRAGENE, and DILIGEN and ProEuroDILI. We thank Jay Hoofnagle and Ludmila Pawlikowska for feedback that helped to improve the manuscript. We thank Daniele Cusi (Hypergenes), Patrik K. Magnusson (Swedish Twin Registry) and Javier Martin (Spanish DNA bank) for provision of control data and Arthur Holden (iSAEC) for establishing collaborations with their institutions. We also thank NCBI/dbGaP at http://www.ncbi.nlm.nih.gov/gap, who approved our application to obtain additional control datasets. In particular, eMERGE GWAS datasets were from dbGaP accession number *phs000360.v3.p1*. The eMERGE network is a consortium of five participating sites (Group Health Seattle, Marshfield Clinic, Mayo Clinic, Northwestern University, and Vanderbilt University) funded by the National Human Genome Research Institute (NHGRI). Funding supports for the cohort and genotyping of eMERGE samples are as follows: (1) Group Health Seattle: U01AG006781 and U01HG004438 with genotyping performed at Johns Hopkins University; (2) Mayo Clinic: UOIHG004599 and U01HG004424 with genotyping performed at The Broad Institute; (3) Marshfield Clinic: U01HG004608 and U01HG004438 with genotyping performed at Johns Hopkins University; (4) Northwestern University: U01HG04603 and U01HG004424 with the genotyping performed at The Broad Institute; (5) Vanderbilt University: U01HG004603 and U01HG004424 with the genotyping performed at The Broad Institute. The PAGE dataset was from the Charles Bronfman Institute for Personalized Medicine (IPM) BioMe BioBank at the Icahn School of Medicine at Mount Sinai (New York). Phenotype data collection was supported by The Andrea and Charles Bronfman Philanthropies. Funding support for genotyping, which was performed at The Center for Inherited Disease Research (CIDR), was provided by the NIH (U01HG007417). The GWAS data were from dbGaP accession number *phs000925.v1.p1*. The Ashkenazi GWAS data were from dbGaP accession number *phs000448.v1.p1*. The dataset was provided by Dr. Todd Lencz and Ariel Darvasi, Ph.D. Support for collection and analysis was provided by RC2MH089964, R01

MH084098, the North Shore - LIJ Health System Foundation, and the Hebrew University Genetic Resource. The kidney GWAS data were from dbGaP accession number *phs001271.v1.p1*. The dataset was funded by grant NCI U01CA155309 to G Scelo. The KORA GWAS data were from dbGaP accession number *phs000303.v1.pl*. The dataset was obtained from the NEI Refractive Error Collaboration (NEIREC) Database. Funding support for NEIREC was provided by the National Eye Institute. We thank NEIREC participants and the NEIREC Research Group for their valuable contribution to this research.

References

1. Chalasani N, Bonkovsky HL, Fontana R, et al. Features and Outcomes of 899 Patients With Drug-Induced Liver Injury: The DILIN Prospective Study. Gastroenterology 2015;148:1340-52 e7.
2. Reuben A, Koch DG, Lee WM. Drug-induced acute liver failure: results of a U.S. multicenter, prospective study. Hepatology 2010;52:2065-76.
3. Watkins PB. Idiosyncratic drug-induced liver injury in patients: Detection, severity assessment, and regulatory implications. Adv Pharmacol 2019;85:165-193.
4. deLemos AS, Ghabril M, Rockey DC, et al. Amoxicillin-Clavulanate-Induced Liver Injury. Dig Dis Sci 2016;61:2406-16.
5. Lucena MI, Andrade RJ, Fernandez MC, et al. Determinants of the clinical expression of amoxicillin-clavulanate hepatotoxicity: a prospective series from Spain. Hepatology 2006;44:850-6.
6. Lucena MI, Molokhia M, Shen Y, et al. Susceptibility to amoxicillin-clavulanate-induced liver injury is influenced by multiple HLA class I and II alleles. Gastroenterology 2011;141:338-47.
7. Stephens C, Lopez-Nevot MA, Ruiz-Cabello F, et al. HLA alleles influence the clinical signature of amoxicillin-clavulanate hepatotoxicity. PLoS One 2013;8:e68111.
8. **Cirulli ET, Nicoletti P**, Abramson K, et al. A Missense Variant in PTPN22 is a Risk Factor for Drug-induced Liver Injury. Gastroenterology 2019;156:1707-1716.e2.
9. Stanford SM, Bottini N. PTPN22: the archetypal non-HLA autoimmunity gene. Nat Rev Rheumatol 2014;10:602-11.
10. de Castro JAL, Stratikos E. Intracellular antigen processing by ERAP2: Molecular mechanism and roles in health and disease. Hum Immunol 2019;80:310-317.
11. Bjornsson ES, Bergmann OM, Bjornsson HK, et al. Incidence, presentation, and outcomes in patients with drug-induced liver injury in the general population of Iceland. Gastroenterology 2013;144:1419-25, 1425 e1-3; quiz e19-20.
12. Fontana RJ, Watkins PB, Bonkovsky HL, et al. Drug-Induced Liver Injury Network (DILIN) prospective study: rationale, design and conduct. Drug Saf 2009;32:55-68.
13. Aithal GP, Watkins PB, Andrade RJ, et al. Case Definition and Phenotype Standardization in Drug-Induced Liver Injury. Clinical Pharmacology and Therapeutics 2011;89:806-815.
14. **Nicoletti P, Aithal GP**, Bjornsson ES, et al. Association of Liver Injury From Specific Drugs, or Groups of Drugs, With Polymorphisms in HLA and Other Genes in a Genome-Wide Association Study. Gastroenterology 2017;152:1078-1089.
15. Gamazon ER, Wheeler HE, Shah KP, et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet 2015;47:1091-8.
16. Li YJ, Phillips EJ, Dellinger A, et al. Human Leukocyte Antigen B*14:01 and B*35:01 Are Associated With Trimethoprim-Sulfamethoxazole Induced Liver Injury. Hepatology 2021;73:268-281.
17. Das S, Forer L, Schonherr S, et al. Next-generation genotype imputation service and methods. Nat Genet 2016;48:1284-1287.
18. Zheng X, Shen J, Cox C, et al. HIBAG--HLA genotype imputation with attribute bagging. Pharmacogenomics Journal 2014;14:192-200.
19. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-75.

20. Ombrello MJ, Kastner DL, Remmers EF. Endoplasmic reticulum-associated amino-peptidase 1 and rheumatic disease: genetics. Curr Opin Rheumatol 2015;27:349-56.
21. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010;26:2336-7.
22. Cordell HJ. Detecting gene-gene interactions that underlie human diseases. Nat Rev Genet 2009;10:392-404.
23. Youden WJ. Index for rating diagnostic tests. Cancer 1950;3:32-5.
24. Andres AM, Dennis MY, Kretzschmar WW, et al. Balancing selection maintains a form of ERAP2 that undergoes nonsense-mediated decay and affects antigen presentation. PLoS Genet 2010;6:e1001157.
25. Etheridge AS, Gallins PJ, Jima D, et al. A New Liver Expression Quantitative Trait Locus Map From 1,183 Individuals Provides Evidence for Novel Expression Quantitative Trait Loci of Drug Response, Metabolic, and Sex-Biased Phenotypes. Clin Pharmacol Ther 2020;107:1383-1393.
26. Nicoletti P, Aithal GP, Chamberlain TC, et al. Drug-Induced Liver Injury due to Flucloxacillin: Relevance of Multiple Human Leukocyte Antigen Alleles. Clin Pharmacol Ther 2019;106:245-253.
27. Zumelzu C, Le Roux-Villet C, Loiseau P, et al. Black patients of African descent and HLA-DRB1*15:03 frequency overrepresented in epidermolysis bullosa acquisita. J Invest Dermatol 2011;131:2386-93.
28. Kuiper JJW, Setten JV, Devall M, et al. Functionally distinct ERAP1 and ERAP2 are a hallmark of HLA-A29-(Birdshot) Uveitis. Hum Mol Genet 2018;27:4333-4343.
29. Tailor A, Meng X, Adair K, et al. HLA DRB1*15:01-DQB1*06:02-Restricted Human CD4+ T Cells Are Selectively Activated With Amoxicillin-Peptide Adducts. Toxicol Sci 2020;178:115-126.
30. Alvarez-Navarro C, Lopez de Castro JA. ERAP1 structure, function and pathogenetic role in ankylosing spondylitis and other MHC-associated diseases. Mol Immunol 2014;57:12-21.
31. Overby CL, Hripcsak G, Shen Y. Estimating heritability of drug-induced liver injury from common variants and implications for future study designs. Sci Rep 2014;4:5762.
32. O'Connor LJ. The distribution of common-variant effect sizes. Nat Genet 2021;53:1243-1249.
33. Chalasani N, Reddy KRK, Fontana RJ, et al. Idiosyncratic Drug Induced Liver Injury in African-Americans Is Associated With Greater Morbidity and Mortality Compared to Caucasians. Am J Gastroenterol 2017;112:1382-1388.
34. Hollenbach JA, Ombrello MJ, Tremoulet AH, et al. IL-1 and IL-6 inhibitor hypersensitivity link to common HLA-DRB1*15 alleles. medRxiv 2021:2020.08.10.20172338.
35. Kaliyaperumal K, Grove JI, Delahay RM, et al. Pharmacogenomics of drug-induced liver injury (DILI): Molecular biology to clinical applications. J Hepatol 2018;69:948-957.
36. Robinson PC, Brown MA. The genetics of ankylosing spondylitis and axial spondyloarthritis. Rheum Dis Clin North Am 2012;38:539-53.

Author names in bold designate shared co-first authorship.

Figure Legends:

Figure 1. Manhattan plot displaying difference between discovery AC-DILI cases and controls in predicted liver mRNA expression (TWAS) conditioned on the presence of known HLA risk alleles. Only *ERAP2* mRNA showed a significant difference ($p < 1.1x10^{-5}$).

Figure 2. Schematic representation of the role of ERAP1 and ERAP2 in the presentation of antigens. ERAP1 and ERAP2 function to shorten peptides generated by the proteosome so that they can fit in the peptide binding domain of Class 1 HLA molecules as they are synthesized in the endoplasmic reticulum and before they are trafficked to the cell surface. ERAP1 and ERAP2 generate different profiles of peptides. Treatment with AC presumably alters some peptides produced by the proteosome. Reduced function of ERAP2 as a result of carriage of two HAP B alleles presumably shifts peptide production to ERAP1. AC altered peptides (neoantigens) presumably created by ERAP1 may stimulate an adaptive immune attack on the liver if presented on the liver cell surface by the risk associated Class 1 HLA molecules. Carriage of two Class I HLA risk alleles presumably doubles the potential load of neoantigens presented on the cell surface, increasing the risk of an adaptive immune response (Figure modified from [36]). TCR = T-cell receptor.

Table

Table 1: Clinical characteristics of subjects in the four AC-DILI case cohorts

CHARACTERISTICS	European discovery	European validation	African American	Hispanic
Clinical information				
Number of patients	444	133	13	7
Mean age, *years*	59	64	60	51
Female, %	44	46	69	14
Mean alanine aminotransferase (SD), *U/L*	485 (618)	465 (425)	705 (804)	337 (217)
Mean alkaline phosphatase (SD), *U/L*	445 (330)	459 (395)	593 (365)	350 (151)
Mean latency (SD), *days*	22 (19.7)	27 (19.3)	34 (23.6)	28 (18.6)
Injury Phenotype				
Cholestatic –R value < 2.0 (%)	162 (38%)	44 (33%)	7 (54%)	2 (29%)
Hepatocellular - R value > 5.0 (%)	101 (22%)	40 (30%)	3 (23%)	0 (0%)
Mixed - R value between 2 and 5 (%)	158 (35%)	46 (35%)	3 (23%)	5 (71%)
Not available (%)	23 (5%)	3 (2%)	0 (0%)	0 (0%)

Table 2. Summary statistics for the single marker analysis of the most relevant associated variants

		Discovery cohort (444 vs 10397)				Validation cohort (133 cases 17836)				Meta-analysis	
Variant of interest	**GENE**	**OR (95%CI)**	**P**	**AF CA**	**AF CO**	**OR (95%CI)**	**P**	**AF CA**	**AF CO**	**OR (95%CI)**	**P**
HLA alleles											
HLA-DRB1*15:01	HLA	3.14 (3.66-2.69)	$1.4x10^{-47}$	0.30	0.12	2.46 (1.87-3.22)	$1.1x10^{-10}$	0.27	0.13	2.96 (2.59-3.39)	$3.8x10^{-56}$
HLA-A*02:01	HLA	2.11 (2.42-1.84)	$4.5x10^{-26}$	0.44	0.28	1.97 (1.54-2.52)	$7.1x10^{-8}$	0.42	0.27	2.08 (1.84-2.34)	$2.2x10^{-32}$
HLA-B*15:18	HLA	3.1 (6.08-1.58)	0.001	0.01	0.004	5.24 (1.85-4.81)	0.002	0.02	0.002	3.61 (2.05-6.37)	0.000009
rs1363907	ERAP2	1.68 (1.23-1.66)	$1.7x10^{-7}$	0.67	0.58	1.2 (1.04-2.05)	0.03	0.66	0.6	1.38 (1.22-1.57)	$5.0x10^{-7}$
rs2476601	PTPN22	1.62 (1.32-1.98)	$4.0x10^{-6}$	0.13	0.08	1.01 (0.66-1.54)	0.95	0.09	0.09	1.4 (1.18-1.67)	0.0001
ERAP1 haplotypes											
Haplotype 2 (Hap2)	ERAP1	1.41 (1.20- 1.71)	$4.9x10^{-5}$	0.19	0.14	1.42 (1.02-1.94)	0.01	0.19	0.14	1.41 (1.23-1.63)	$2.0x10^{-6}$
Haplotype 8 (Hap8)	ERAP1	0.78 (0.62-0.88)	0.005	0.19	0.24	0.95 (0.70-1.28)	0.63	0.22	0.23	0.84 (0.74-0.97)	0.014
Haplotype 10 (Hap10)	ERAP1	0.87 (0.73-1.05)	0.01	0.18	0.20	0.88 (0.62-1.21)	0.22	0.18	0.2	0.87 (0.79-0.96)	0.004

Odds ratios (OR); confidence intervals (95%CI); p-values (P) are presented for multivariate analyses correcting for population stratification with EIGENSTRAT axes within each cohort; AF CA= minor allele frequency in cases; AF CO = minor allele frequency in controls.

Table 3. Multivariable model for AC-DILI risk incorporating HLA-A*02:01, HLA-DRB1*15:01, HLA-B*15:18, ERAP2 rs1363907 and PTPN22 rs2476601 in the European population

		Discovery cohort		*Validation cohort*		*Meta-analysis*	
Marker	**Genotype**	**OR (95%CI)**	**P**	**OR (95%CI)**	**P**	**OR (95%CI)**	**P**
HLA-A*02:01	HZ	2.23 (1.78-2.79)	$1.6x10^{-12}$	2.55 (1.73-3.77)	$2.6x10^{-6}$	2.31 (1.9-2.8)	$2.5x10^{-17}$
	HH	4.59 (3.42-6.16)	$2.3x10^{-24}$	3.62 (2.09-6.26)	$4.3x10^{-6}$	4.35 (3.36-5.64)	$7.2x10^{-29}$
HLA-DRB1*15:01	HZ	3.26 (2.66-4.04)	$6.8x10^{-29}$	2.48 (1.7-3.61)	$2.2x10^{-6}$	3.06 (2.55-3.67)	$1.9x10^{-33}$
	HH	9.83 (6.70-14.44)	$1.8x10^{-31}$	6.76 (3.57-12.78)	$4.3\ x10^{-9}$	8.9 (6.40-12.40)	$8.5x10^{-39}$
HLA-B*15:18	HZ	3.95 (1.97-7.89)	$1.1x10^{-4}$	7.71 (2.73-21.81)	$1.2\ x10^{-4}$	4.86 (2.72-8.66)	$8.7x10^{-8}$
rs1363907 (G) ERAP2	HH	1.60 (1.31-1.96)	$3.3x10^{-6}$	1.43 (1.01-2.03)	0.04	1.56 (1.31-1.85)	$4.3x10^{-7}$
rs2476601(A) PTPN22	HZ	1.53 (1.20-1.95)	$5.6x10^{-4}$	0.92 (0.57-1.49)	0.73	1.38 (1.11-1.71)	0.003
	HH	3.56 (1.68-7.46)	$7.8x10^{-4}$	1.69 (0.41-7.00)	0.47	3.04 (1.57-5.86)	0.0009

Odds ratios (OR), 95% confidence intervals (95%CI) and p-values (P) are presented after correcting for population stratification; HZ = risk allele heterozygote, HH = risk allele homozygote.

Table 4 Relative risk and estimate of incidence of AC DILI predicted from GRS in the discovery and validation cohorts.

		Discovery cohort (444 vs 10397)				Validation cohort (133 cases 17836)			
Q	**GRS-range**	**N cases (%)**	**N controls (%)**	**RR**	**Estimated cases in general population**	**N cases (%)**	**N controls (%)**	**RR**	**Estimated cases in general population**
1	0-0.49	71 (16%)	3918 (38%)	0.43	1:5407	22 (17%)	6740 (38%)	0.44	1:5362
2	0.5-0.99	43 (10%)	1979 (19%)	0.52	1:4525	19 (14%)	3189 (18%)	0.80	1:2945
3	1.0-1.499	67 (15%)	1925 (18%)	0.82	1:2861	24 (18%)	3199 (18%)	1.00	1:2343
4	1.5-1.99	77 (17%)	1442 (14%)	1.24	1:1898	23 (17%)	2555 (14%)	1.20	1:1955
5	2.0-2.499	82 (19%)	715 (7%)	2.51	1:935	20 (15%)	1364 (8%)	1.95	1:1207
6	2.5-2.99	39 (9%)	193 (2%)	4.10	1:572	9 (7%)	395 (2%)	3.00	1:783
7	3.0-3.499	38 (9%)	119 (1%)	5.91	1:397	10 (8%)	219 (1%)	5.88	1:399
8	3.5-3.99	21 (5%)	46 (0.4%)	7.65	1:307	2 (2%)	99 (0.6%)	2.67	1:881
9	4.0-4.499	5 (1%)	11 (0.1%)	7.63	1:307	4 (3%)	16 (0.09%)	26.94	1:87
10	4.5-4.81	1 (0.2%)	0	24.42	1:96	0	4 (0.02%)		

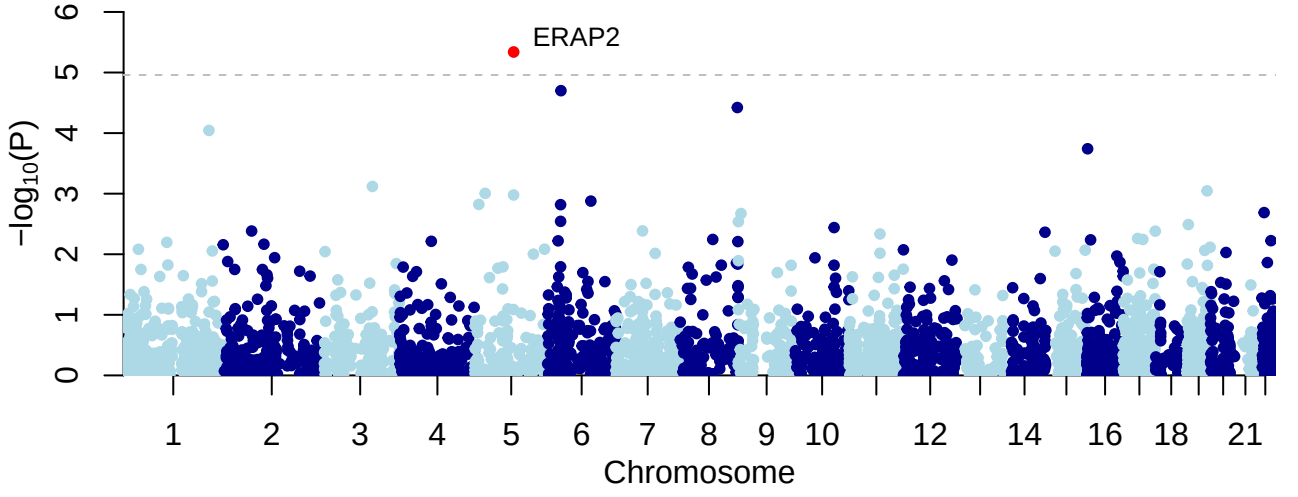
Q = groups of GRS; GRS range = GRS max and min in the group; N (%) = number of subjects and proportion of cases or controls in each group out of the total number; RR= Relative Risk; Estimated incidence calculated based on relative risk and 1:2,350 incidence of AC-DILI in patients treated with AC[11].

Table 5: Summary statistics of age, sex, GRS, latency, and injury type in the multivariable logistic model developed to aid in the diagnosis of AC-DILI. The model was developed comparing the discovery European AC-DILI cohort to DILI due to drugs other than AC, and then applied to the validation AC-DILI cases using the same comparator non-AC-DILI cohort.

	Discovery (444 AC-DILI vs 1358 non-AC-DILI)		**Validation (133 AC-DILI vs 403 non-AC-DILI)**	
Factor	**OR (95%CI)**	**P**	**OR (95%CI)**	**P**
Age	1.01 (1.00-1.02)	$3.7x10^{-5}$	1.05 (1.03-1.07)	$1.7x10^{-7}$
Sex (M)	1.97 (1.50-2.59)	$9.5x10^{-7}$	1.07 (0.62-1.84)	0.82
GRS	2.90 (2.48-3.40)	$1.6x10^{-40}$	2.39 (1.80-3.17)	$1.7 x10^{-9}$
Latency	0.39 (0.30-0.51)	$7.3x10^{-12}$	0.48 (0.39-0.63)	$1.1 x10^{-7}$
Injury type				
cholestatic	2.95 (208-4.18)	$1.1x10^{-9}$	1.19 (0.59-2.38)	0.62
mixed	3.28 (2.32-4.64)	$1.4x10^{-11}$	1.53 (0.79-2.98)	0.20

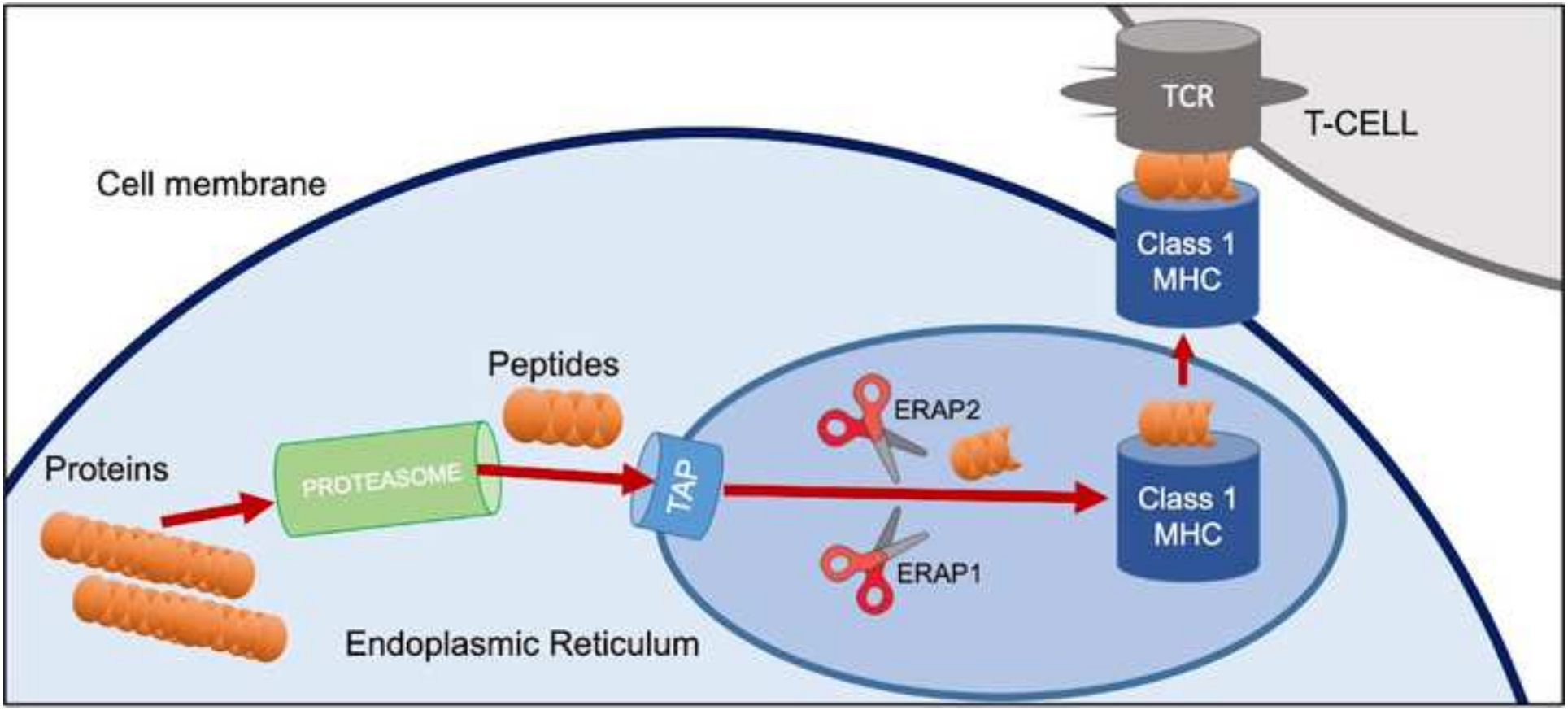
Odds ratios (OR), 95% confidence intervals (95%CI); P =logistic regression p-value correcting for population stratification. Age, GRS and Latency are quantitative variables. Their OR is referred to a unit increase. For Injury type, hepatocellular was the baseline group.

Figure



Figure

Click here to access/download;Figure;Figure2.jpg



Identification of reduced ERAP2 expression and a novel HLA allele as components of a risk score for susceptibility to liver injury due to amoxicillin-clavulanate

Paola Nicoletti[1#], Andrew Dellinger[2], Yi Ju Li[2,3], Huiman X. Barnhart[2,3], Naga Chalasani[4], Robert J. Fontana[5], Joseph A. Odin[6], Jose Serrano[7], Andrew Stolz[8], Amy S. Etheridge[9], Federico Innocenti[9], Olivier Govaere[10], Jane I. Grove[11], Camilla Stephens[12], Guruprasad P. Aithal[11], Raul J. Andrade[12], Einar S. Bjornsson[13], Ann K. Daly[10], M. Isabel Lucena[12] and Paul B. Watkins[9,14 #]

on behalf of Drug-Induced Liver Injury Network (DILIN), iDILIC and ProEuroDILI investigators

Supplementary materials

Imputation pipeline

SNP imputation was performed in batches dividing the samples according to genotyping platforms and ethnicity. For each batch, imputation was carried out using Michigan Imputation Server.[1] We used Haplotype Reference Consortium as the reference for European ancestry samples and the TOPMED as the reference for other ancestries[2, 3]. Sex chromosomes and mitochondria genes were not imputed. As we previously described,[4] well-imputed SNPs with $R^2 > 0.6$ were retained and genotypes were discretized based on the probability (PP) > 0.9.

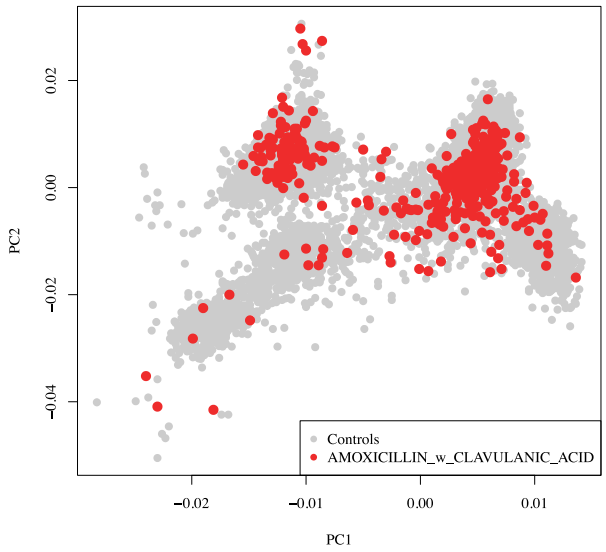
Clinical characteristics of the DILI cases

We collected additional clinical information to further investigate the relevance of the significant associations with AC-DILI risk. Time from start of medication to DILI recognition (latency), age, gender and maximum serum levels of alkaline phosphatase (ALP) and (ALT) were available for both DILIN and iDILIC cases. Also available on all cases was whether the injury was hepatocellular, cholestatic or mixed (based on the initial R value). After transforming the quantitative clinical variables (latency, maximum ALP and maximum ALT) to improve normality, we applied a linear regression model to test differences among the significant risk genotypes in the AC cohorts. Between group differences were tested using the Fisher's exact test or chi-square test for categorical variables and the t-test or Wilcoxon/Kruskal-Wallis test for the continuous variables as appropriate. All statistical analyses were performed with STATA15.

Supplementary Figures

Figure S1: Matching genetic ancestry of cases and controls. Principal components (PCs) were derived for cases and controls using a set of 38,697 shared SNPs across cohorts. Shown are scatterplots representing the first two principal components of the discovery study cohort (A) and replication cohort (B). They showed the homogenous distribution between cases and controls across major European clusters. Cases are in red and controls are in gray.

A)



B)

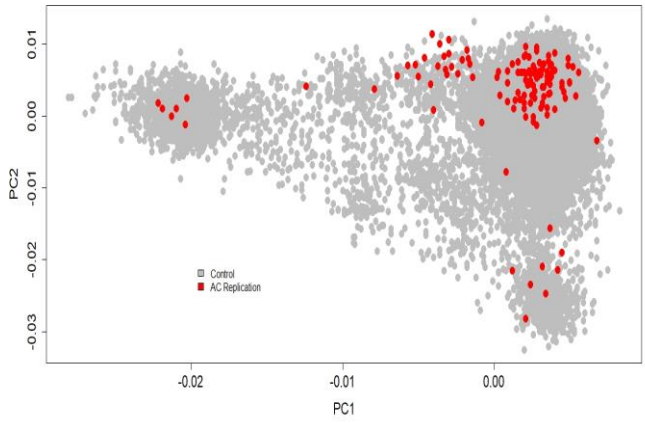
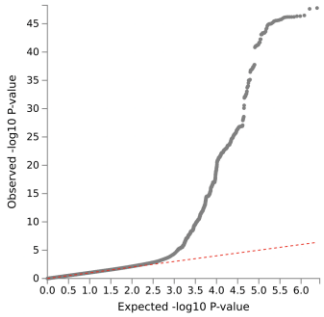


Figure S2. QQ plot (A) and Manhattan plot (B) displaying the association results of each SNP tested in AC-DILI GWAS analysis. Lambda = 1.05

(A)



(B)

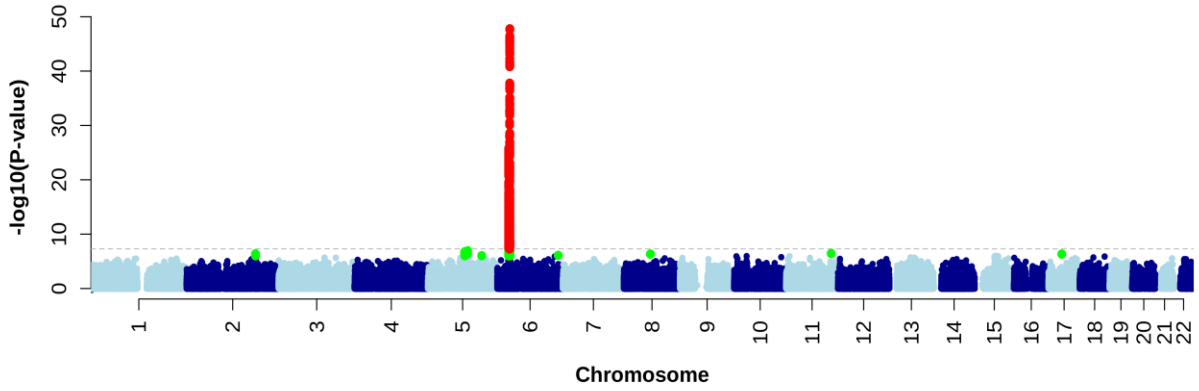


Figure S3. Manhattan plot and QQ plot displaying the association results of each SNP tested in AC-DILI GWAS analysis after conditioning for the genotypes of HLA-A*02:01 and HLA-DRB1*15:01. SNP shown in green have a significance level less than $5x10^{-6}$. Lambda = 1.02

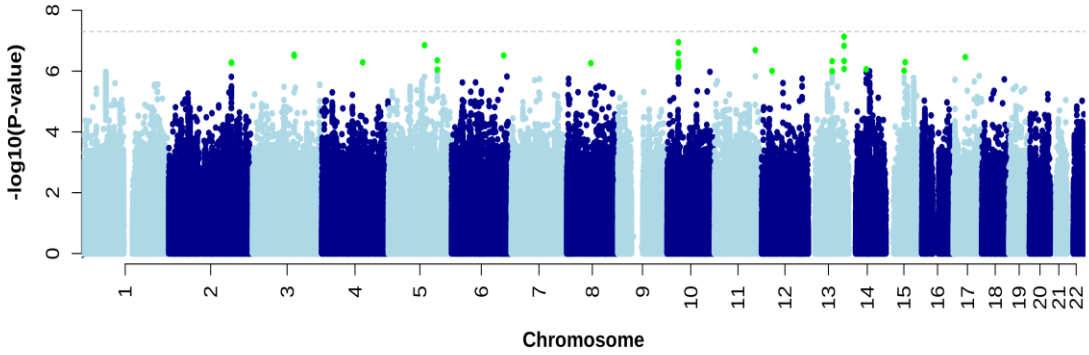


Figure S4. Manhattan plot displaying the association results of differential gene expression between AC-cases and controls. In this analysis SNPs were used to predict liver mRNA expression in AC cases (discovery cohort) and controls using the Predixscan approach. Genes shown in green have a significance level less than $5x10^{-6}$ (significant for liver gene expression) and in red less than $5x10^{-8}$.

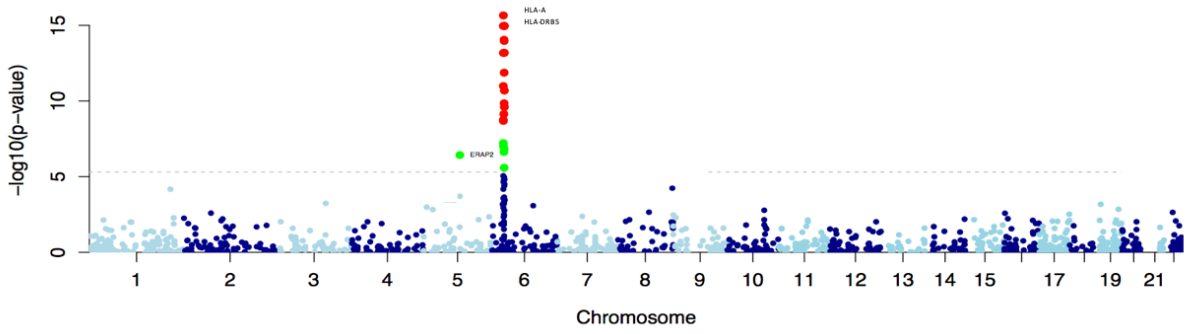


Figure S5 Locus zoom plots of ERAP2 region displaying AC-DILI GWAS results (A) and ERAP2 eQTL results (B). In panel C the box plot displays ERAP2 mRNA expression in 1,183 liver samples[5] across rs1363907 genotype groups, showing how the heterozygote group (G/A) has similar expression as homozygotes (A/A).

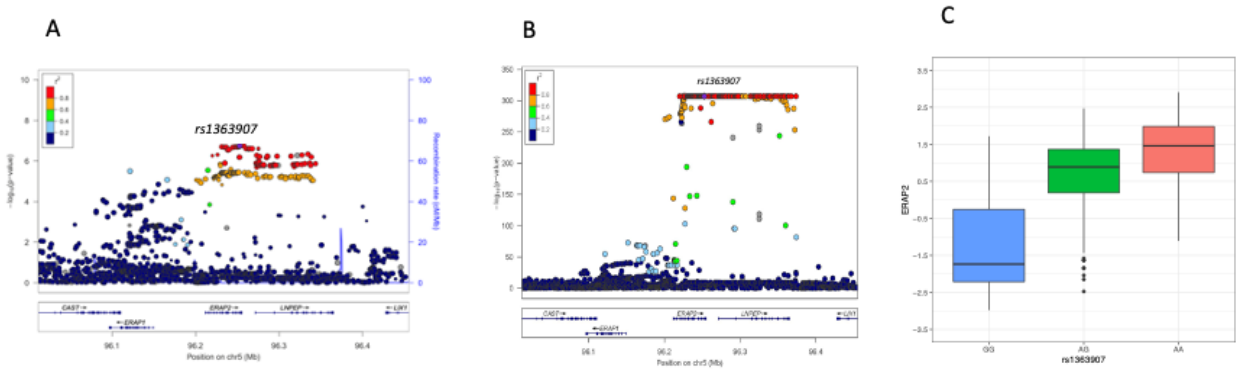


Figure S6 Box plot shows the expression of rs1363907 G allele following recessive (A), dominant (B) models using the meta-analysis liver specific eQTL dataset.

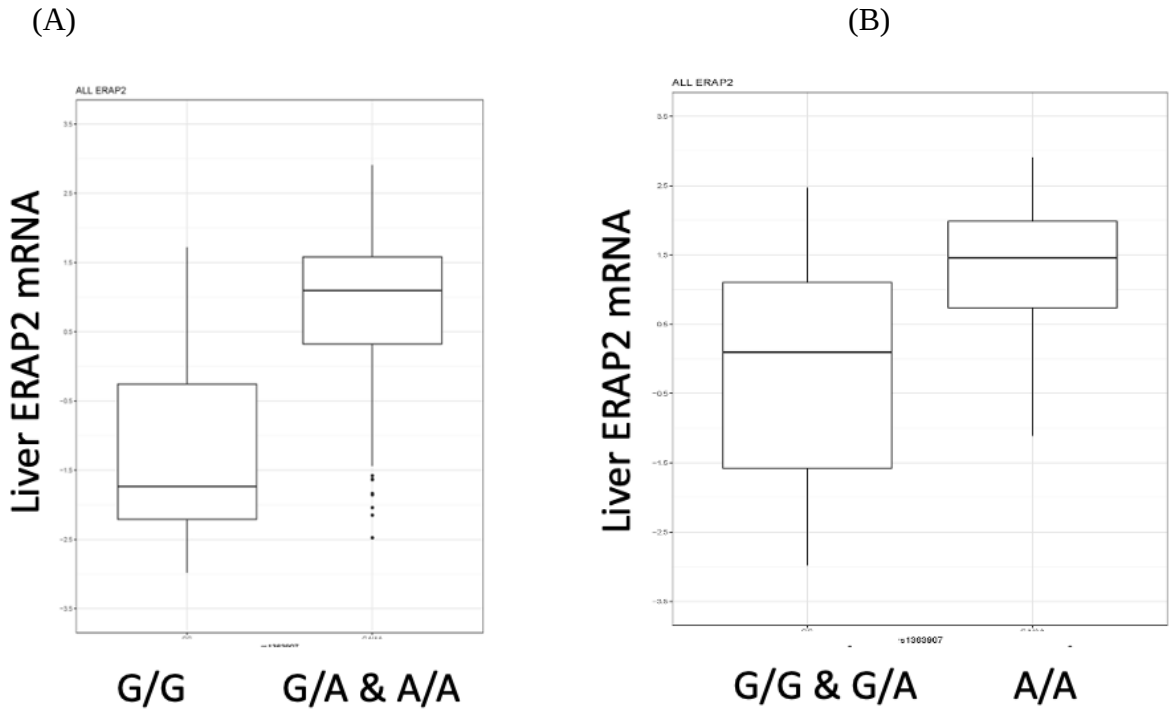


Figure S7 Locus zoom plots of ERAP2 region displaying AC-DILI GWAS results from logistic regression under recessive model (A) and dominant model (B).

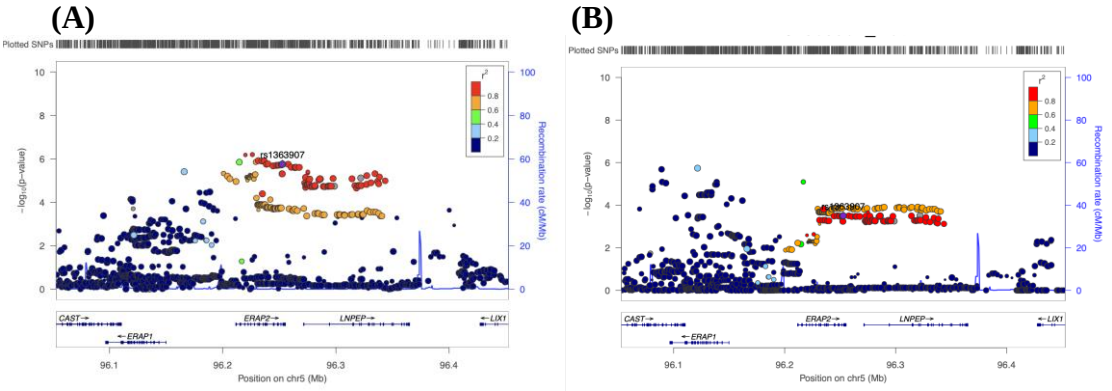


Figure S8. The box plot shows the expression of each HLA locus across the three ERAP2 genotype groups using GTEX.

A

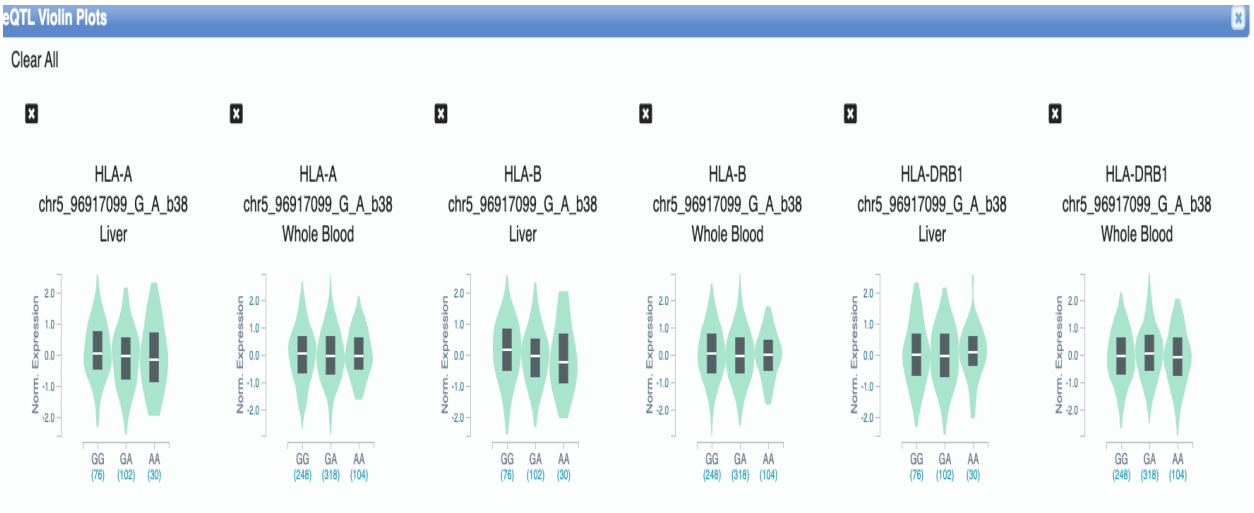


Figure S9. Forest plot showing the summary statistics for rs1363907 (ERAP2, G allele) association across causal agent groups containing at least 20 cases from the largest published DILI cohort (Cirulli et al.[4]).

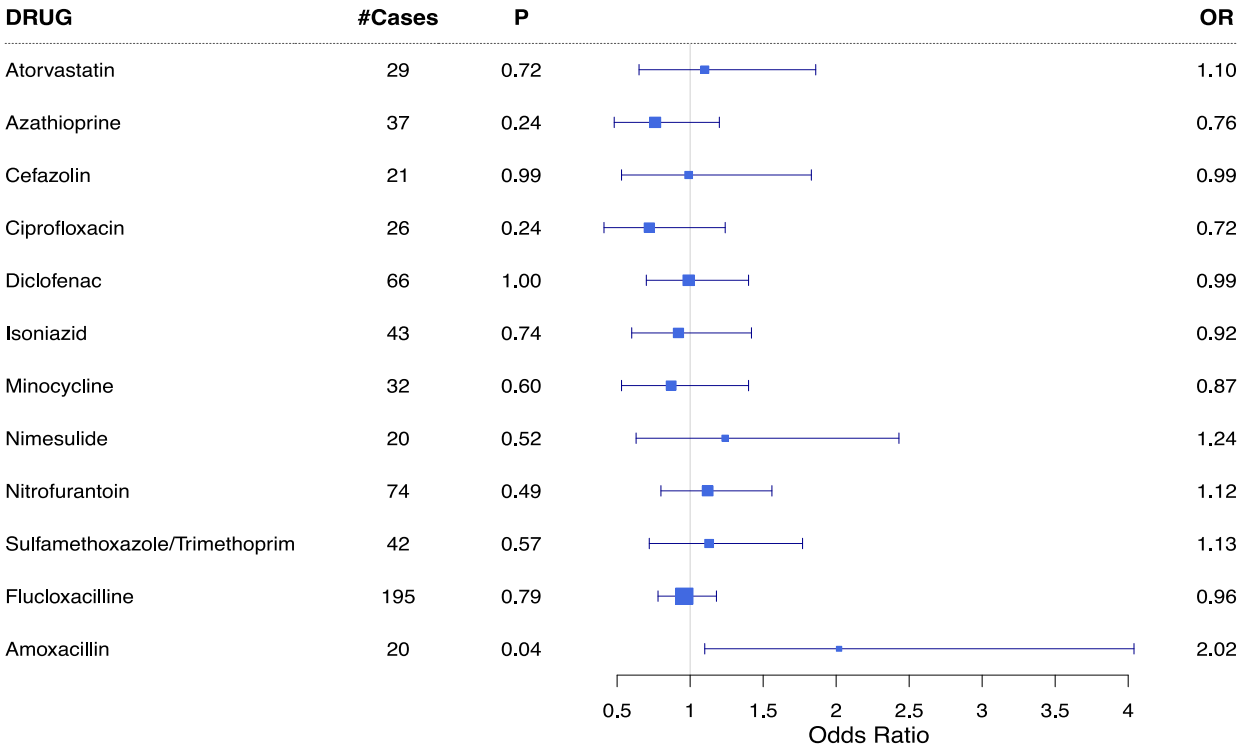


Figure S10. Distribution of Genetic Risk Score (GRS) among AC cases and controls (discovery cohorts).

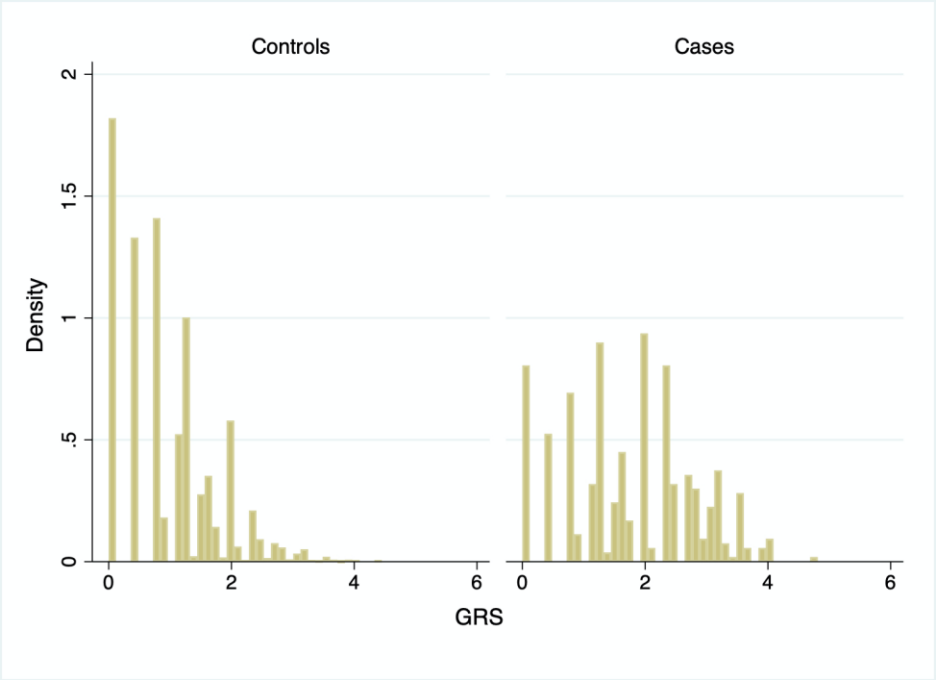
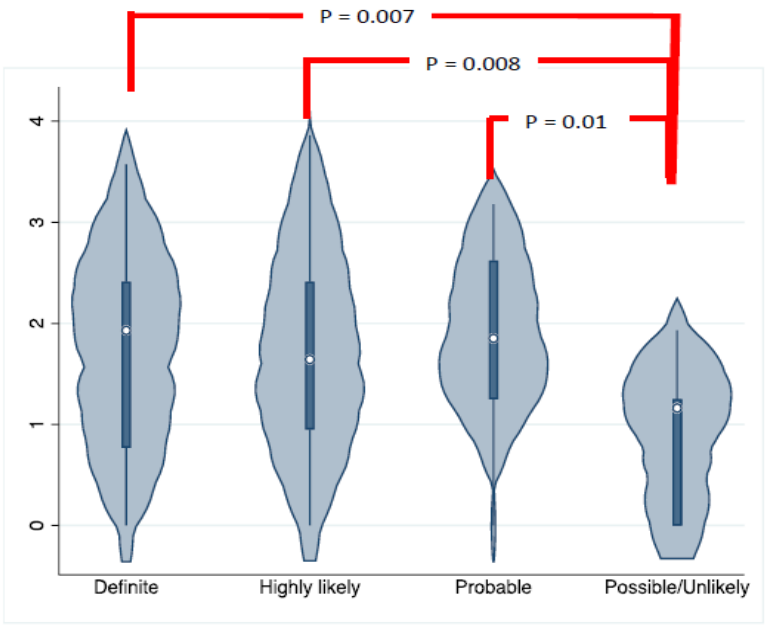


Figure S11. Violin plots showing the distribution of GRS across the four categories from DILIN causality assessment by expert opinion for (A) AC-DILI cases including all DILIN discovery and validation cases (79 definite, 94 highly likely, 29 probable cases plus 13/6 possible/unlikely cases excluded from the main analysis since were not high confident cases) and (B) DILI cases due to other causal drugs (142 definite, 364 highly likely, 179 probable, and 170 possible/unlikely). (Difference in GRS mean ($P<0.05$) among groups was tested by one-way ANOVA using Bonferroni correction as post doc test).

(A)



(B)

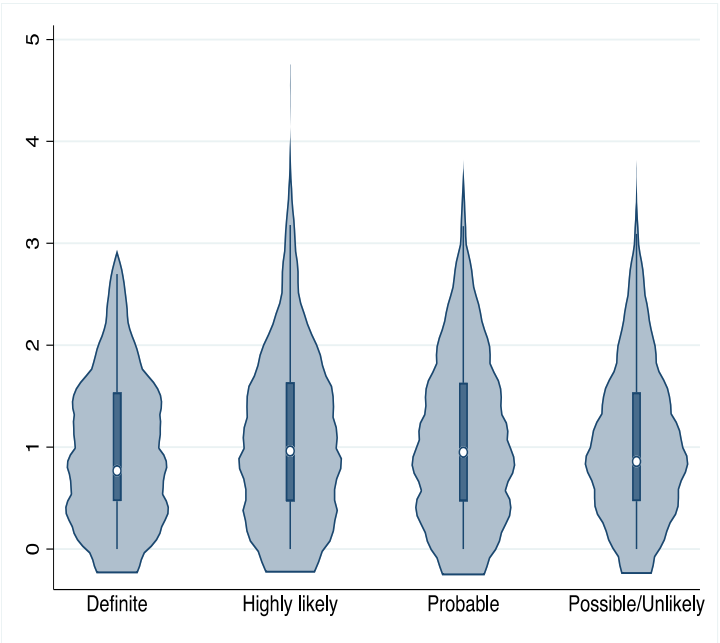
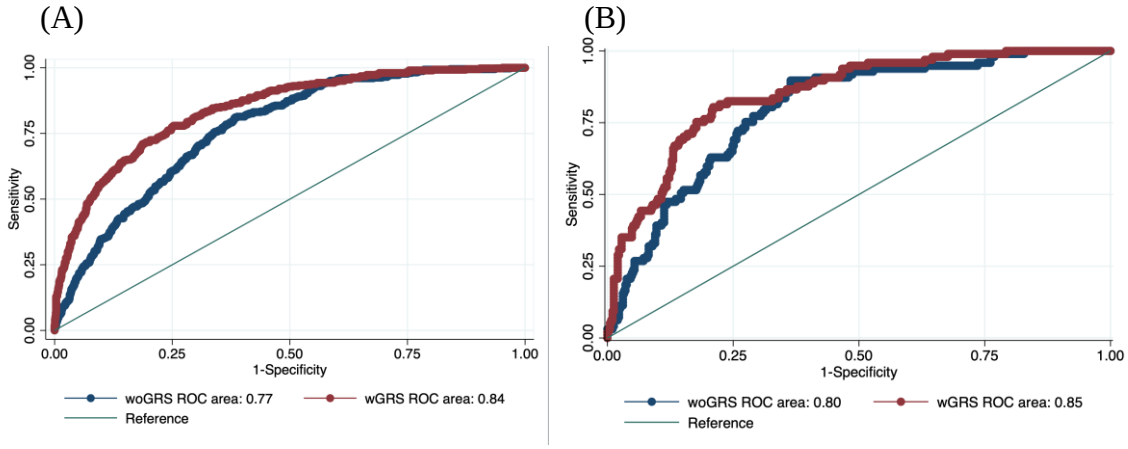


Figure S12: ROC curves comparison between the model considering solely clinical predictors (blue) and the model including also GRS score (red) in the (A) AC-DILI case/non-AC-DILI control discovery cohort and (B) AC-DILI case/non-AC-DILI control validation cohort



Supplementary Tables

Table S1: List of the DILI cases included in the AC DILI European discovery cohort

Cohort	European discovery	African American	Hispanic
Number of patients	444	13	7
DILIN	145	13	7
Diligen	99		
Eudragene	19		
iDILIC	130		
Spanish DILI network	51		

Table S2: List of datasets with correspondent genotyping platform included in the AC DILI European replication cohort

Cohort	Dataset	Sample Size	Platform
Case			
	DILIN	57	MEGAEX
	IDILIC	50	MEGAEX
	Nottingham	6	MEGAEX
	Spanish Network	20	MEGAEX
Control			
	PAGE (phs000925.v1.p1)	44	MEGA
	eMERGE1 (phs000360.v3.p1)	12121	Illumina 660v1
	Genetics of Schizophrenia in a Ashkenazi Jewish Case-Control Cohort (phs000448.v1.p1)	2052	Omni 1 Quad chip
	Genetic Epidemiology of Refractive Error in the KORA study (phs000303.v1.p1)	1810	Omni 2.5 chip
	Pooled Genome-wide Analysis of Kidney Cancer Risk (phs001271.v1.p1)	1809	Omni 5 chip

Table S3: Clinical characteristics of subjects in the two large independent European non-AC-DILI cohorts and in the validation control set

CHARACTERISTICS	European non-AC-DILI discovery cohort	European non-AC-DILI validation cohort	European Population-based validation cohort
Clinical information			
Number of patients	1358	403	17836
Mean age, *years*	53	52	61
Female, %	39	42	47
Mean latency (SD), *days*	127 (356)	127 (297)	-
Injury Phenotype			
Cholestatic –R value < 2.0 (%)	301 (22%)	88 (22%)	-
Hepatocellular - R value > 5.0 (%)	647 (48%)	227 (56%)	-
Mixed - R value between 2 and 5 (%)	307 (23%)	88 (22%)	-
Not available (%)	103 (7%)	- (0%)	-

Table S4 Examples of how to calculate the GRS according to the formula:

$$\begin{aligned}\text{GRS} = {} & (0.77 \text{ x allele dosage of HLA} - \text{A} * 0201) \\ & + (1.16 \text{ x allele dosage of HLA} - \text{DRB1} * 15{:}01) \\ & + (1.37 \text{ x allele dosage of HLA} - \text{B} * 15{:}18) \\ & + (0.47 \text{ x ERAP2 rs1363907 GG genotype} \\ & + (0.48 \text{ x allele dosage of PTPN22 rs2476601 A}\end{aligned}$$

	A*02:01		DRB1*15:01		B*15:18		PTPN22		ERAP2 (GG)		GRS
	Coeffcient	number of risk alleles	Coeffcient	number of risk alleles	Coeffcient	number of risk alleles	Coeffcient	number of risk alleles	Coeffcient	number of risk alleles	
individual, carrying one HLA-A*02:01 allele and one HLA-DRB1*15:01 allele and is HH for ERAP2	0.77	1	1.16	1	1.37	0	0.47	0	0.48	1	2.41
individual, carrying one risk allele for PTPN22	0.77	0	1.16	0	1.37	0	0.47	1	0.48	0	0.47

Table S5 Summary statistics for the univariate analysis of genome-wide associated variants and conditioned on HLA-A*02:01 and HLA-DRB1*15:01 for the European discovery cohort.

Variants		Discovery cohort (444 vs 10397)				Conditional analsyis Discovery cohort (444 vs 10397)		Validation cohort (133 cases 17886)				Meta-analysis	
GWAS significant variants	GENE	OR (95%CI)	P	AF CA	AF CO	OR (95%CI)	P	OR (95%CI)	P	AF CA	AF CO	OR (95%CI)	P
rs9268927	MHC, intergenic,	3.13 (2.68-3.65)	$6.3x10^{-47}$	0.3	0.12	3.08 (0.98-9.64)	0.05	2.51 (1.91-3.29	$2.4x10^{-11}$	0.31	0.15	2.96 (2.59-3.39)	$2.4x10^{-56}$
rs2734966	MHC, intergenic	2.11 (1.83-2.42)	$4.2x10^{-26}$	0.47	0.3	1.89 (1.28-2.79)	0.001	2.01 (1.58-2.57)	$1.2x10^{-8}$	0.45	0.29	2.08 (1.85-2.35)	$3.5x10^{-33}$
Other variants of interest													
rs2476601	PTPN22	1.62 (1.32-1.98)	$4.0x10^{-6}$	0.13	0.08	1.43 (1.23-1.65)	2.2x10-6	1.01 (0.66-1.54)	0.95	0.09	0.09	1.4 (1.18-1.67)	0.0001
rs1363907	ERAP2	1.68 (1.23-1.66)	$1.7x10^{-7}$	0.67	0.58	1.62 (1.30-1.98)	8.9x10-6	1.2 (1.04-2.05)	0.03	0.66	0.6	1.38 (1.22-1.57)	$5.0x10^{-7}$
HLA alleles													
HLA-DQB1*06:02[a]	HLA	3.19 (3.73-2.73)	$1.4x10^{-48}$	0.3	0.11	2.2 (0.97-4.98)	0.05	2.45 (1.86-3.21)	$1.4x10^{-10}$	0.27	0.13	2.99 (2.61-3.42)	$5.9x10^{-57}$
HLA-DRB1*15:01	HLA	3.14 (3.66-2.69)	$1.4x10^{-47}$	0.3	0.12	-	-	2.46 (1.87-3.22)	$1.1x10^{-10}$	0.27	0.13	2.96 (2.59-3.39)	$3.8x10^{-56}$
HLA-DQA1*01:02[a]	HLA	2.7 (3.12-2.34)	$5.3x10^{-42}$	0.38	0.18	1.54 (1.19-2.01)	0.001	2.17 (1.68-2.80)	$2.6x10^{-9}$	0.35	0.2	2.56 (2.26-2.9)	$2.9x10^{-49}$
HLA-A*02:01	HLA	2.11 (2.42-1.84)	$4.5x10^{-26}$	0.44	0.28	-	-	1.97 (1.54-2.52)	$7.1x10^{-8}$	0.42	0.27	2.08 (1.84-2.34)	$2.2x10^{-32}$
HLA-B*07:02[a]	HLA	1.82 (2.16-1.53)	$1.0x10^{-11}$	0.2	0.12	0.9 (0.73-1.1)	0.41	1.45 (1.05-2.01)	0.02	0.18	0.12	1.73 (1.48-2.01)	$1.4x10^{-12}$
HLA-B*15:18	HLA	3.1 (6.08-1.58)	0.001	0.01	0.004	4.19 (2.09-8.36)	4.9x10-5	5.24 (1.85-14.81)	0.002	0.02	0.002	3.61 (2.05-6.37)	0.000009

Odds ratios (OR), confidence intervals (95%CI) and p-values (P) are presented after correcting for population stratification with EIGENSTRAT axes and known HLA risk alleles. MHC-P-value threshold for significance is $P < 0.0002$

Table S6 The table shows the LD structure of the 4 informative SNPs associated with the two ERAP2 allotypes Hap A and Hap B[6]. r2 and D' values were calculated by LDmatrix Tool.

SNP	FUNCTION	SNP	FUNCTION	r^2	D'
rs2549782	Missense variant	rs2248374	Splice region variant	1	1
rs2549782	***Missense variant***	***rs1363907***	***Intron variant***	***0.80***	***1.00***
rs1363907	***Intron variant***	***rs10044354***	***Intron variant***	***0.96***	***1.00***
rs2248374	***Splice region variant***	***rs1363907***	***Intron variant***	***0.80***	***1.00***
rs2248374	Splice region variant	rs10044354	Intron variant	0.83	1.00
rs2549782	Missense variant	rs10044354	Intron variant	0.83	1.00

Table S7: Multivariate model for AC-DILI risk considering *HLA-DRB1*15:01, HLA-A*02:01* and rs2476601 (*PTPN22*) and considering all three genotype groups of rs1363907 (*ERAP2*) in the European discovery population.

		Discovery cohort	
Marker	**Genotype**	**OR (95%CI)**	**P**
HLA-A*02:01	HZ	2.23 (1.78-2.79)	2.10E-12
	HH	4.59 (3.42-6.16)	2.70E-24
HLA-DRB1*15:01	HZ	3.26 (2.65-4.02)	1.10E-28
	HH	9.68 (6.59-14.23)	5.70E-31
HLA-B*15:18	HZ	3.93 (1.96-7.87)	1.10E-04
rs1363907 (G, ERAP2)	HZ	1.26 (0.91-1.74)	1.60E-01
	HH	1.92 (1.39-2.65)	7.90E-05
rs2476601(A, PTPN22)	HZ	1.53 (1.2-1.94)	6.10E-04
	HH	3.5 (1.67-7.35)	9.20E-04

Odds ratios (OR), 95% confidence intervals (95%CI) and p-values (P), HZ = heterozygote, HH = homozygote.

Table S8 Summary statistics for the univariate of ERAP1 haplotype association conditioned on ERAP2 signal (rs1363907).

ERAP1 Haplotypes	**OR**	**P**
Hap2	1.26	0.01
Hap8	0.93	0.45
Hap10	0.84	0.07

Odds ratios (OR) and p-values (P)

Table S9. Summary statistics of the most significant (P < 0.001) combinations of HLA-DRB1*15:01, HLA-A*02:01 and rs2476601(A, *PTPN22*) and homozygote rs1363907 (GG, *ERAP2*) from the multi-marker logistic model using the European AC-DILI discovery cohort.

HLA-A*02:01	HLA-DRB1*15:01	HLA-B*15:18	ERAP2 (rs1363907)	PTPN22 (rs2476601)	OR 95%CI	P	N in cases	N in controls	AF in cases	AF in controls
					8.7 (2.08-13.9)	**$1.3x10^{-19}$**	**36**	**221**	**0.08**	**0.02**
					18.58 (6.27-35.99)	$4.6x10^{-18}$	16	47	0.04	0.005
					17.47 (5.99-34.21)	**$7.1x10^{-17}$**	15	44	0.03	0.004
					12.27 (3.69-22.12)	$8.0x10^{-17}$	19	94	0.04	0.01
					29.72 (13.67-73.21)	$1.7x10^{-13}$	9	15	0.02	0.001
					15 (5.96-32.68)	$9.3x10^{-12}$	10	38	0.02	0.004
					31.72 (17.5-93.52)	$3.7x10^{-10}$	6	10	0.01	0.001
					41.83 (26.5-144.82)	$3.8x10^{-9}$	5	6	0.01	0.001
					16.34 (8.06-42.97)	$1.5x10^{-8}$	6	22	0.01	0.002
					59.36 (43.68-251.11)	$2.9x10^{-8}$	4	5	0.01	0.0005
					3.77 (0.92-6.08)	**$5.4x10^{-8}$**	**31**	**463**	**0.07**	**0.04**
					6.53 (2.34-13.18)	**$1.6x10^{-7}$**	**11**	**91**	**0.02**	**0.01**
					15.2 (8.95-48.19)	$3.8x10^{-6}$	4	16	0.01	0.002
					4.12 (1.32-7.71)	$9.3x10^{-6}$	14	189	0.03	0.02
					7.56 (3.56-19.03)	$1.8x10^{-5}$	6	40	0.01	0.004
					17.41 (11.94-66.77)	$3.1x10^{-5}$	3	10	0.01	0.001
					57.73 (59.99-442.5)	$9.5x10^{-5}$	2	2	0.005	0.0002
					24.66 (21.39-134.98)	$2.2x10^{-4}$	2	5	0.005	0.0005
					4.01 (1.61-8.79)	$5.1x10^{-4}$	8	117	0.02	0.01
					2.77 (0.83-4.99)	$6.9x10^{-4}$	16	321	0.04	0.03
					2.13 (0.49-3.35)	$9.8x10^{-4}$	36	906	0.08	0.09
					13.5 (10.95-66.16)	1.3x10-3	2	8	0.005	0.001

Odds ratios (OR), 95% confidence intervals (95%CI) and p-values (P) are presented after correcting for population stratification and considering the five negative carriers as the reference group. The risk alleles status is represented by "light green " = heterozygote, "dark green" = homozygote and "gray" = absent. N = number of samples in the group; AF = allele frequency. In bold the combinations mentioned in the main text.

Table S10. Predictive value of the multi-marker model of the five risk variants and of the logistic regression considering GRS (estimated on the 5 risk variants) and other clinical risk factors for AC-DILI

		444 Discovery AC-DILI cases		133 validation AC-DILI cases	
Model	**Control group**	**R^2**	**AUC**	**R^2**	**AUC**
Multi-marker model					
HLA-DRB1*15:01 + HLA-A*02:01 + rs2476601 (A, PTPN22) + rs1363907 (GG, ERAP2)	Population controls[$]	0.13	0.76	0.12	0.79
	non-AC DILI set*	0.17	0.76	0.10	0.70
Logistic model					
GRS	Population controls[$]	0.12	0.75	0.09	0.77
GRS	Non-AC-DILI set*	0.16	0.76	0.09	0.68
Age + being Male + latency + R-value [#]	Non-AC-DILI set*	0.15	0.76	0.19	0.80
GRS + Age + being Male + latency + R-value [#]	Non-AC-DILI set*	0.27	0.84	0.28	0.85

[#]Cholestatic (R < 2.0) or Mixed (R 2-5) vs Hepatocellular (R>5.0) as baseline
[$]444 AC-DILI cases were compared to 10397 population controls and the 133 AC-DILI validation cases were compared to other independent 17836 non-AC-DILI individuals
[*]444 AC-DILI cases were compared to 1558 non-AC-DILI individuals and the 133 AC-DILI validation cases were compared to other independent 403 non-AC-DILI individuals

Table S11. Pairwise interaction analysis among identified AC-DILI risk factors, considering the carriage status of HLA-DRB1*15:01, HLA-A*02:01 and rs2476601 (A, PTPN22) and homozygote rs1363907 (GG, ERAP2) in the European discovery and validation cohorts

	Discovery	**Validation**	**Meta-analysis**
Model	**Interaction [OR 95%CI P]**	**Interaction [OR 95%CI P]**	**Interaction [OR 95%CI P]**
All 3 HLA risk alleles x rs1363907	***1.72 (1.02-2.89) P = 0.04***	***1.18 (0.49-2.830 P = 0.70***	**1.55 (0.998-2.41) P = 0.05**
*HLA-DRB1*15:01* x rs1363907	1.18 (0.80-1.74) P = 0.40	1.09 (0.54-2.20) P = 0.79	1.16 (0.831-1.61) P = 0.39
***HLA-A*02:01* x rs1363907**	***1.61 (1.05-2.47) P = 0.03***	***1.56 (0.72-3.34) P = 0.25***	**1.6 (1.1-2.32) P = 0.01**
*HLA-B*15:18* x rs1363907	0.5 (1.12-1.98) P = 0.33	1.29 (0.16-10.24) P = 0.81	0.671 (0.211-2.14) P = 0.5
All 3 HLA risk alleles x rs2476601	***2.07 (1.04-4.14) P = 0.03***	***1.44 (0.39-5.34) P = 0.58***	**1.92 (1.07-3.45) P = 0.03**
*HLA-DRB1*15:01* x rs2476601	1.23 (0.78-1.95) P = 0.30	0.08 (0.34-2.20) P = 0.77	1.23 (0.83-1.82) P = 0.3
*HLA-A*02:01* x rs2476601	1.58 (0.94-2.66) P = 0.08	0.50 (0.19-1.31) P = 0.16	1.23 (0.781-1.93) P = 0.37
*HLA-B*15:18* x rs2476601	1.88 (0.45-7.84) P = 0.38	1.45 (0.12-16.37) P = 0.76	1.76 (0.522-5.91) P = 0.36
HLA-DRB1*15:01* x *HLA-A*02:01	***2.01 (1.31-3.07) P = 0.001***	***1.88 (0.86-4.07) P = 0.10***	**1.98 (1.38-2.85) P = 0.0002**
*HLA-B*15:18* x *HLA-A*02:01*	1.02 (0.25-4.05) P =0.98	0.59 (0.07-4.69) P = 0.62	0.84 (0.242-2.91) P = 0.78
*HLA-DRB1*15:01* x *HLA-B*15:18*	0.81 (1.17-3.86) P = 0.80	0.76 (0.07-8.28) P = 0.82	0.794 (0.207-3.04) P = 0.74
rs1363907 x rs2476601	1.00 (0.63-1.60) P = 0.99	1.75 (0.69-4.43) P = 0.23	1.52 (0.691-3.34) P = 0.3

Odds ratios (OR), 95% confidence intervals (95%CI) and p-values (P) are presented after correcting for population stratification. In bold are reported the significant interaction terms.

Table S12. Summary statistics for the multivariate analysis of AC-GRS and risk of DILI due to other agents

DRUG	N	Mean GRS (SD)	% Samples with GRS > 1.58	OR (95%CI)	P
Controls	10,348	0.95 (0.78)	21%	-	-
Overall DILI (w/o AC cases)	1318	1.00 (0.79)	23%	1.02 (0.95-1.10)	0.44
Amoxicillin/Clavulanic Acid	444	1.76 (1.05)	56%	2.71 (2.44-3.01)	$9.7x10^{-179}$
Flucloxacillin	195	0.83 (0.73)	17%	0.94 (0.58-1.51)	0.02
Nitrofurantoin	74	1.13 (0.82)	24%	1.08 (0.73-1.61)	0.11
Diclofenac	66	1.15 (0.95)	31%	1.85 (1.17-2.93)	0.18
Isoniazid	43	0.83 (0.72)	18%	1.02 (0.63-1.67)	0.22
Sulfamethoxazole/Trimethoprim	41	0.95 (0.78)	24%	1.21 (0.91-1.62)	0.80
Azathioprine	37	1.06 (0.66)	29%	0.77 (0.51-1.16)	0.67
Minocycline	32	0.81 (0.67)	15%	0.71 (0.44-1.15)	0.17
Atorvastatin	29	0.93 (0.93)	31%	1.39 (0.79-2.42)	0.81
Ciprofloxacin	26	0.99 (0.70)	19%	1.23 (0.94-1.62)	0.91
Cefazolin	21	1.50 (0.70)	47%	0.95 (0.63-1.41)	0.008
Nimesulide	20	1.02 (0.86)	25%	0.66 (0.31-1.39)	0.24
Amoxicillin alone	20	1.05 (0.88)	20%	0.79 (0.66-0.96)	0.84
Azithromycin	17	0.92 (0.14)	6%	0.89 (0.48-1.65)	0.72

Odds ratios (OR) and p-values (P) LCI =lower 95% CI; UCI = upper 95% CI; N =number of cases. Case/controls genetic data was extracted from Cirulli and Nicoletti et al.[4]. Only causal drugs with 20 or more DILI cases were considered in this analysis. Each drug group was compared with the same discovery cohort controls. None of the comparisons other than AC passed the Bonferroni threshold for significance (P = 0.003).

Table S13 List of DILIN cases that were receiving AC as concomitant treatment, but AC was not considered causal, their calculated GRS, and the estimated risk of developing DILI as a result of treatment with AC.

Subject	GRS	~Risk	rs2476601 (PTPN22)	HH rs1363907 (ERAP2)	HLA-DRB1*15:01	HLA-A*02:01	HLA-B*15:18	Causality grade	GRS CUT OFF	CAUSAL DRUG	OTHER MEDICATIONS
1*	2.70	1:600	0	0	1	2	0	Probable	YES	Azithromycin	Amoxicillin W/Clavulanic Acid
2	2.40	1:900	0	2	1	1	0	Probable	YES	Methylprednisolone	Amoxicillin W/Clavulanic Acid
3	1.64	1:1,900	1	0	1	0	0	Highly likely	YES	Sulfamethoxazole W/Trimethoprim	Amoxicillin W/Clavulanic Acid
4	1.93	1:1,900	0	0	1	1	0	Probable	YES	Azathioprine	Nebivolol, Amoxicillin W/Clavulanic Acid
5	1.71	1:1,900	1	2	0	1	0	Highly likely	YES	Minocycline	Amoxicillin W/Clavulanic Acid
6	1.24	1:2,900	0	2	0	1	0	Probable	NO	Pravastatin	Amoxicillin W/Clavulanic Acid
7	1.24	1:2,900	0	2	0	1	0	Possible	NO	Cefalexin	Amoxicillin W/Clavulanic Acid
8	0.77	1:4,500	0	0	0	1	0	Probable	NO	Montelukast	Amoxicillin W/Clavulanic Acid
9	0.77	1:4,500	0	0	0	1	0	Probable	NO	Infliximab	Amoxicillin W/Clavulanic Acid, Melatonin
10	0.77	1: 4,500	0	0	0	1	0	Highly likely	NO	Disulfiram	Amoxicillin W/Clavulanic Acid
11	0.77	1: 4,500	0	0	0	1	0	Probable	NO	Sulfamethoxazole W/Trimethoprim	Amoxicillin W/Clavulanic Acid
12	0.77	1: 4,500	0	0	0	1	0	Possible	NO	Thiamazole	Amoxicillin W/Clavulanic Acid
13	0	1:5,500	0	0	0	0	0	Highly likely	NO	Gatifloxacin	Amoxicillin W/Clavulanic Acid
14	0	1:5,500	0	0	0	0	0	Possible	NO	Isoniazid	Amoxicillin W/Clavulanic Acid, Levofloxacin
15	0	1:4,000	0	0	0	0	0	Possible	NO	Moxifloxacin	Ciprofloxacin, Amoxicillin W/Clavulanic Acid

GRS = Genetic Risk Score; Risk – estimated from GRS based on Table 5 in the main manuscript.

*Subject 1 - 76 y.o. man who developed a respiratory infection and subsequent pneumonia. He was treated with AC, an Azithromycin dose pack, and a Medrol dose pack and he noted jaundice about 3 weeks later. He was hospitalized and developed pruritus, constipation, fecal impaction and urinary retention, fatigue and a 10lb unintentional weight loss. His bilirubin peaked at 15.5 mg/dl, his albumin

fell to 2.4 g/dl, and his R value was always less than 2.0. His laboratory values normalized over the next month, but it took two more months until he felt well. He did not recall ever taking azithromycin or AC prior to this event. Complete workup for other causes was negative. Upon completion of the DILIN causality assessment process, azithromycin and AC were considered equally likely to have caused the event. Re-review by PW, AS, JS, and RF was influenced by the GRS value that placed this patient in the quantile 6, which estimates he is among about 2% of patients with a risk of roughly 1:600 of developing DILI, a far higher incidence than DILI in patients receiving azithromycin [7,8]. They also noted that there was no trend for association between azithromycin DILI risk (Table S11) and the GRS, and based the model combining clinical factors with GRS (Table 6 in the manuscript) , the fact that the patient was a man and 76 years old was consistent with AC was the culprit. There was unanimous agreement that the causal agent was AC and not azithromycin.

Table S14 Summary statistics for the univariate association analysis of known HLA alleles and other alleles in the same family group in African American and Hispanic AC-DILI cohorts

	African American cohort (13 cases)					Hispanic cohort (7 cases)				
ALLELE	**AF cases**	**AF controls**	**P**	**AF Pop**	**# carriers**	**AF cases**	**AF controls**	**P**	**AF Pop**	**# carriers**
HLA-A*02:01	0.19	0.12	0.22	0.12	5	0.14	0.23	0.75	0.19	2
HLA-A*02:02	0.15	0.04	0.02	0.04	4	0	0.01		0.007	0
HLA-A*02:05	0.08	0.02	0.09	0.02	2	0.07	0.02	0.26	0.01	1
HLA-A*02:06	0.04	0	0.004	0.0002	1	0	0.01		0.04	0
HLA-A*02:13	0	0			0	0.07	0.003	0.04	0.0008	1
B*15:01	0	0		0.010	0	0	0.04		0.03	0
B*15:03	0.08	0.07	0.70	0.06	2	0	0.01		0.02	0
B*15:10	0.00	0.08		0.03	0	0	0.01		0.005	0
B*15:16	0.08	0.02	0.10	0.02	2	0	0.01		0.005	0
B*15:17	0	0.004		0.006	0	0	0.005		0.007	0
B*15:18	0	0		0.001	0	0	0		0.001	0
DRB1*15:01	0.04	0.02	0.39	0.03	1	0.14	0.06	0.22	0.07	2
DRB1*15:02	0.04	0	-	0.001	1	0	0.01		0.01	0
DRB1*15:03	0.23	0.13	0.13	0.12	6	0	0.01		0.01	0
rs1363907 (G, ERAP)	0.77	0.65	0.3	0.62	13	0.5	0.68	0.25	0.7	7
rs2476601 (A, PTPN22)	0	0.01		0.01	0	0.2	0.04	0.07	0.03	3

Odds ratios (OR), 95% confidence intervals (95%CI) and p-values (P) are presented after correcting for population stratification; AF = minor allele frequency; AF Pop = allele frequency reported in the largest cohort listed in http://www.allelefrequencies.net/. N carrier = number of individuals carrying an allele.

For non-European individuals, predictions for the *ERAP1* haplotypes could not be estimated based on the missing data of several relevant proxy SNPs.

Table S15 Summary statistics of the top associated HLA alleles for comparison between 43 discovery cases without any of the five known risk factors and discovery population controls.

Allele	OR	LCI	UCI	P	AF negative cases	AF negative controls	AF all AC cases	AF control all
*HLA-B*18:18*	208.00	11.74	3685.00	0.0003	0.01	0.0002	0.001	4.8×10^{-5}
*HLA-B*18:01*	2.88	1.54	5.40	0.001	0.14	0.05	0.07	0.05
*HLA-B*35:01*	2.85	1.51	5.38	0.001	0.13	0.07	0.05	0.05
*HLA-C*04:01*	2.24	1.35	3.73	0.002	0.22	0.14	0.09	0.11

Odds ratios (OR), 95% confidence intervals (95%CI) and p-values (P) are presented after correcting for population stratification, AF = allele frequency. "Negative case" = AC DILI cases not carrying any of the five risk markers; "Negative controls" = European population controls not carrying any of the five risk markers. With correction for multiple comparisons, a significant P value is < 0.0004.

References:

1. Das S, Forer L, Schonherr S, et al. Next-generation genotype imputation service and methods. Nat Genet 2016;48:1284-1287.
2. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 2016;48:1279-83.
3. Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature 2021;590:290-299.
4. Cirulli ET, Nicoletti P, Abramson K, et al. A Missense Variant in PTPN22 is a Risk Factor for Drug-induced Liver Injury. Gastroenterology 2019;156:1707-1716.e2.
5. Etheridge AS, Gallins PJ, Jima D, et al. A New Liver Expression Quantitative Trait Locus Map From 1,183 Individuals Provides Evidence for Novel Expression Quantitative Trait Loci of Drug Response, Metabolic, and Sex-Biased Phenotypes. Clin Pharmacol Ther 2020;107:1383-1393.
6. de Castro JAL, Stratikos E. Intracellular antigen processing by ERAP2: Molecular mechanism and roles in health and disease. Hum Immunol 2019;80:310-317.
7. Mosholder AD, Mathew J, Alexander JJ, Smith H, Nambiar S. Cardiovascular Risks with Azithromycin and Other Antibacterial Drugs. N Engl J Med 2013; 368:1665-1668.
8. Chalasani N, Bonkovsky HL, Fontana R, Lee W, Stolz A, Talwalkar J, Reddy KR, Watkins PB, Navarro V, Barnhart H, Gu J, Serrano J; Features and Outcomes of 899 Patients with Drug-Induced Liver Injury: The DILIN Prospective Study. Gastroenterology 2015 Jun;148(7):1340-52.e7