Supplementary Appendix

Supplement to:

Association of Liver Injury From Specific Drugs, or Groups of Drugs, With Polymorphisms in HLA and Other Genes in a Genome-wide Association Study

Paola Nicoletti, Guruprasad P. Aithal, Einar S. Bjornsson, Raul J. Andrade, Ashley Sawle, Marco Arrese, Huiman X. Barnhart, Emmanuelle Bondon-Guitton, Paul H. Hayashi, Fernando Bessone, Alfonso Carvajal, Ingolf Cascorbi, Elizabeth T. Cirulli, Naga Chalasani, Anita Conforti, Sally A. Coulthard, Mark J. Daly, Christopher P. Day, John F. Dillon, Robert J. Fontana, Jane I. Grove, Pär Hallberg, Nelia Hernández, Luisa Ibáñez, Gerd A. Kullak-Ublick, Tarja Laitinen, Dominique Larrey, M. Isabel Lucena, Anke H. Maitland-van der Zee, Jennifer H. Martin, Mariam Molokhia, Munir Pirmohamed, Elizabeth E. Powell, Shengying Qin, Jose Serrano, Camilla Stephens, Andrew Stolz, Mia Wadelius, Paul B. Watkins, Aris Floratos, Yufeng Shen, Matthew R. Nelson, Thomas J. Urban and Ann K. Daly

Table of Contents

Collaborators and Contributors to case recruitment

Members of the SAEC scientific management committee
Arthur L Holden [SAEC], Matt Nelson [GlaxoSmithKline], Steve Lewitzky [Novartis], John Sullivan [Amgen], Fredrik Nyberg [AstraZeneca], Peter Shaw [Merck], Alexander Vandell [Daichi-Saiko], Justin Davis [Abbvie], Michael Dunn [Wellcome Trust]

DILIGEN collaborators and investigators:
A.K. Daly (PI), C.P. Day, (Newcastle University); G.P. Aithal (Nottingham Digestive Diseases Centre); M. Pirmohamed (University of Liverpool). Research nurses: J. Henderson, R. Wake (Newcastle University); C Davies, S. Henry (Nottingham Digestive Diseases Centre); K. Hawkins, A. Hanson, J. Evely (University of Liverpool), S. Cleary (Dundee). Other contributors to case recruitment: H. Hussaini (Truro), W. Griffiths (Addenbrooks Hospital, Cambridge), J. Collier (John Radcliffe Infirmary, Oxford), A. Brind (North Staffordshire), N. Fisher (Dudley), J. Shearman (South Warwick), A. Grant (Leicester Royal Infirmary), A. Austin (Derby), F. Gordon (Bristol), M. Cramp (Plymouth), S. Saksena (North Durham), H J McMurtry (Chorley), S. Verma (Brighton), H. Mitchison (Sunderland), A. M. Elsharkawy (Birmingham), H. Dallal (Middlesbrough), C McDonald (Carlisle), J. Metcalf (Hartlepool)

DILIN collaborators
For complete listing see http://dilin.org/publications/

EUDRAGENE collaborators and investigators
Mariam Molokhia (Joint PI), KCL, UK, Paul McKeigue (Joint PI), University of Edinburgh, Scotland, Data Analysis Committee: Bruno Stricker, Erasmus MC, NL, Qun-Ying Yue, Medical Products Agency, Uppsala, Sweden. Centre investigators: Alfonso Carvajal, Universidad de Valladolid, Spain, Luisa Ibáñez, Fundació Institut Català de Farmacologia, Hospital Universitari Vall d' Hebron, Universitat Autònoma, Barcelona, Spain, Maryse Lapeyre-Mestre, Université de Toulouse, France, Jean-Louis Montastruc, Université de Toulouse, France, Dr. Anita Conforti, University Hospital, Verona, Italy, Giampaolo Velo, University Hospital, Verona, Italy, Michel Eichelbaum, Dr. Margarete Fischer-Bosch Institute Clinical Pharmacology, Stuttgart. Contributors to case recruitment/other: Emmanuelle Bondon-Guitton, Université de Toulouse, France, Inés Salado, Universidad de Valladolid, Spain, Lourdes Vendrell, Fundació Institut Català de Farmacologia, Hospital Universitari Vall d' Hebron, Barcelona, Spain, Francesca Succurro, University Hospital, Verona, Italy, Marco Smerghetto, University Hospital, Verona, Italy, Mark Eijgelsheim, Erasmus MC, NL, Ramazan Buyukcelik, Erasmus MC, NL, Pascal Arp, Erasmus MC, NL, André Uitterlinden Erasmus MC, Richard Jackson, UK, Ferran Orsola, UK, Justin Matthews, UK. Pharmacovigilance centres involved in case ascertainment (France, Spain, Italy): L'Association Française des Centres Régionaux de Pharmacovigilance (CRPV), Sistema Español de Farmacovigilancia. Agencia Española de Medicamentos y Productos Sanitarios, Italian Pharmacovigilance Centres of Veneto Region, Dr William Bernal, Kings College London, UK Clinical Research Network.

Spanish DILI Registry collaborators
Hospital Universitario Virgen de la Victoria, Málaga (coordinating center): RJ Andrade, MI Lucena, C Stephens, M García-Cortés, M Robles-Díaz, I Medina-Cáliz,

J Sanabria, B García-Muñoz, R Alcántara, I Moreno, A Gonzalez-Jimenez, A Papineau, A Ortega-Alonso.
Hospital Regional Universitario de Málaga: M Jiménez-Pérez, R González-Grande.
Hospital Torrecárdenas, Almería: MC Fernández, G Peláez, M Casado, M González-Sánchez.
Hospital Universitario Virgen de Valme, Sevilla: M Romero-Gómez, R Calle-Sanz, R Millan-Dominguez, B Fombuena, R Gallego, L Rojas, A Rojas, J Ampuero, JA del Campo Antonio Gil Gómez.
Hospital de Mendaro, Guipúzcoa: A Castiella, EM Zapata, L Zubiaurre.
Hospital Central de Asturias, Oviedo: R Pérez-Álvarez, L Rodrigo-Sáez
Hospital Costa del Sol, Marbella (Málaga): JM Navarro, IM Mendez-Sánchez, Ana Chaves.
Hospital Sant Pau, Barcelona: G Soriano, C Guarner, EM Román.
Hospital Morales Meseguer, Murcia: H Hallal, E García-Oltra, JC Titos-Arcos, A Pérez-Martínez, C Sánchez-Cobarro, JM Egea-Caparrós.
Hospital de Donosti, San Sebastián: J Arenas, MI Gomez-Osua, A Gómez-García, FJ Esandi.
Hospital de Basurto, Bilbao: S Blanco, P Martínez-Odriozola.
Hospital Alto Deba Mondragón, Guipúzcoa: P Otazua.
Hospital Universitario San Cecilio, Granada: J Salmerón, A Gila.
Hospital Clínico Valladolid: JM González, S Lorenzo.
Hospital La Fe, Valencia: M Prieto, I Conde Amiel, M Garcia-Eliz, M Berenguer.
Hospital de Sagunto, Valencia: J Primo, JR Molés, A Garayoa.
Hospital de Laredo, Cantabria: M Carrascosa.
Hospital 12 de Octubre, Madrid: E Gómez-Domínguez.
Hospital Germans Trias i Pujol, Badalona, Barcelona: E Montané, AL Arellano, M Farré, AM Barriocanal, Y Sanz.
Hospital Clínic, Barcelona: M Bruguera, P Gines, S Lens, JC García.
Hospital Universitario de Canarias, La Laguna, Tenerife: A Aldea-Perona, M Hernández-Guerra, M Moreno-San Fiel, C Boada-Fernández del Campo.
Hospital Infanta Elena, Valdemoro, Madrid: M Tejedor.
Hospital de Alcorcón, Alcorcón, Madrid: C Fernández, M Fernández-Gil.
Hospital Reina Sofía, Córdoba: JL Montero, M de la Mata.
Hospital Miguel Servet, Zaragoza: J Fuentes-Olmo, EM Fernández-Bonilla.
Complejo Hospitalario Universitario de Albacete, Albacete: JM Moreno, P Martínez-Rodenas, M Garrido.
Hospital Xeral Calde, Lugo: S Ávila.
Hospital Virgen de las Nieves, Granada: F Nogueras.
Hospital Puerta de Hierro, Madrid: J de la Revilla, M Trapero, M Gómez.
Hospital Quirón de Marbella, Málaga: VM Aguilar, M De Sola.
Hospital Puerta del Mar, Cádiz: P Rendón.
Hospital Parc Taulí de Sabadell, Barcelona: M Vergara, J Sánchez Delgado.
Hospital Arnau de Villanova, Valencia: O González-López.
Hospital Carlos III, Madrid: J García-Samaniego, A Madejón.
Hospital de Galdakao: JL Cabriada.
Hospital Marqués de Valdecilla, Santander: J Crespo.

Other iDILIC collaborators

Eric Eliasson and Patrik K. Magnusson (Karolinska Institutet, Stockholm, Sweden); Ulrica Ramqvist, Elisabet Stjernberg, Sofie Collin, Eva Prado, Agnes Wadelius, Martha Wadelius, Agnes Kataja Knight and Hugo Kohnke at SWEDEGENE (Department of Medical Sciences, Uppsala University, Sweden), Dick Menzies

(Montreal), Renate Udo and Marie L. De Bruin (Utrecht Insitute for Pharmaceutical Sciences, Utrecht University, the Netherlands), L.C. Baak (Onze Lieve Vrouwe Gasthuis, Amsterdam, The Netherlands), J.T. Brouwer (Reinier de Graaf Gasthuis, Delft, The Netherlands), J.P.H. Drenth (Radboud University Medical Center, Nijmegen, The Netherlands), M. Klemt-Kropp (Medisch Centrum Alkmaar, The Netherlands), T.C.M.A. Schreuder (Slingeland Ziekenhuis, Doetichem, The Netherlands), L.A. van der Waaij (Martini Ziekenhuis, Groningen, The Netherlands), F.H.J. Wolfhagen (Tweesteden Ziekenhuis, Tilburg, The Netherlands), A Mäkilä and T. Vasankari (Turku University Hospital, Turku, Finland), I. Rajalahti (Tampere University Hospital, Tampere, Finland), J. Koskela, K. Kainu and A. Lindqvist (Helsinki University Central Hospital, Helsinki, Finland), Eulàlia Pérez and Mònica Sabaté (FICF, Barcelona, Spain), Anneke N. Werk (University of Kiel, Germany), Thomas Stammschulte and Ursula Gundert-Remy (Drug Commission of the German Medical Association (DCGMA), Berlin, Germany), Alex Ruiz (Departmento de Gastroenterología, Escuela de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile).

Supplementary Methods

Imputation

The imputation was performed in batches dividing the cohorts according to genotyping platforms. For each batch, we first phased the data by SHAPEIT (version v2.r727),[1-2]. Then, imputation was carried out using IMPUTE2 (version 3) with 1000 Genomes Project (release v3[3]) ethnically mixed dataset as the reference panel.[4 5] We retained imputed genotypes with: (a) posterior probability > 0.9 in each genotyping batch, (b) no significant difference in missingness between cases and controls ($\chi 2$ test, p-value > 0.0001), (c) no significant deviation from Hardy-Weinberg expectations (p-value > 0.0001), (d) no variants missing greater than 5% of genotypes in any single genotyping batch and (e) info score greater than 0.8 in each genotyping batch, (f) MAF in the 1000 Genomes Project $\geq$ 0.01. Batch effects for imputed SNPs were corrected by testing for association between ethnically-matched controls typed by different platforms (using logistic regression). SNPs with association p-values less than 0.005 were excluded from the analysis. For each cohort, four digit HLA alleles were also inferred using HIBAG[6] with the reference predictor panels specific for each genotyped chip.[7]

Genome-wide association study QC for each cohort

QC was conducted at both single marker and subject levels before performing the SNP imputation. Any marker that did not pass the following criteria was excluded from analysis: (i) genotype call rate in the batch of subjects greater than 95%, (ii) missing genotype rate greater than 5%, (ii) p-value for Hardy-Weinberg equilibrium greater than 10^{-7} in controls (if applicable). Any subject that did not pass the following criteria was excluded from analysis: (i) missing genotype rate < 0.05 among

the SNPs that passed QC; (ii) not a sample duplicate or closely related based on estimated identity-by-descent (IBD) using PLINK v 1.07

Quality controls on the A*33:01 association

To assess whether the A*33:01 signal was an artefact of population structure, we tested population-specific association. Although the allele is rarer in Northern Europeans, with no difference between Sweden and the UK (Allele Frequency $(AF)_{uk}$ = 0.004; AF_{sw} = 0.003), the OR was comparable across the three major clusters (Table S3). The heterogeneity test cannot reject the null hypothesis ($\chi 2$ method p-value of 0.06). We also confirmed that the association was not due to synthetic differences in imputation performance relating to the genotype chips. The AF was comparable across the three control groups genotyped by different platforms (AF_{1M} = 0.01, AF_{HEC} = 0.018, AF_{OE} = 0.013). Logistic regression to test for differences between genotype platforms among control samples showed no difference in the Spanish (p-value=0.44) or Italian (p-value=0.9) subsets.

Validation of the predicted genotypes

We validated the predictions of the most associated SNPs within the discovery samples by matching the predicted and the typed genotypes. We calculated the concordance rate as the percentage of accurately predicted genotypes over the total number of samples typed in the validation based on specimen's availability. In particular, the rs72631567 genotypes were validated in 564 discovery cases (386 DILIN cases and 178 iDILIC) with respectively 99.7% and 100% concordance. Both the rs114577328 and rs28458792 genotypes were fully validated in 386 DILIN cases with 100% concordance. The two most associated SNPs in the statin comparison, rs116561224 and rs28458792, were typed in 25 iDILIC statin cases and rs116561224 only was typed in 378 DILIN samples across multiple causal drugs. The concordance

was 100% for iDILIC cases and 97.6% for DILIN cases. In both cohorts this genotyping was performed by TaqMan® predesigned and custom SNP genotyping assays (ThermoFisher Scientific, Waltham, MA) in accordance with the manufacturer's recommendations.

References

1. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nature Methods 2013;10:5-6.

2. Howie B, Fuchsberger C, Stephens M, et al. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet 2012;44:955-9.

3. Delaneau O, Marchini J. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nat Commun 2014;5:3934.

4. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nature Reviews Genetics 2010;11:499-511.

5. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genetics 2009;5:e1000529.

6. Zheng X, Shen J, Cox C, et al. HIBAG--HLA genotype imputation with attribute bagging. Pharmacogenomics J 2014;14:192-200.

7. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet 2007;81:1084-97.

Supplementary Figures

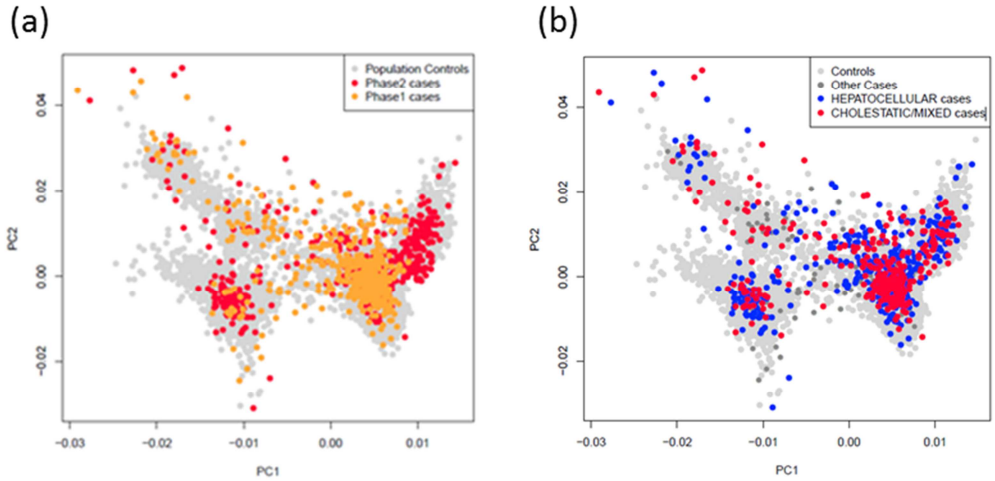

Figure S1. Scatterplots representing the first two principal components of the current study cohort. The homogenous distribution between cases and controls across the three major European clusters is shown. In panel (a) cases from phase II are highlighted in red and the cases from phase I in orange. In panel (b) the cholestatic/mixed cases are highlighted in red and the hepatocellular cases in blue.

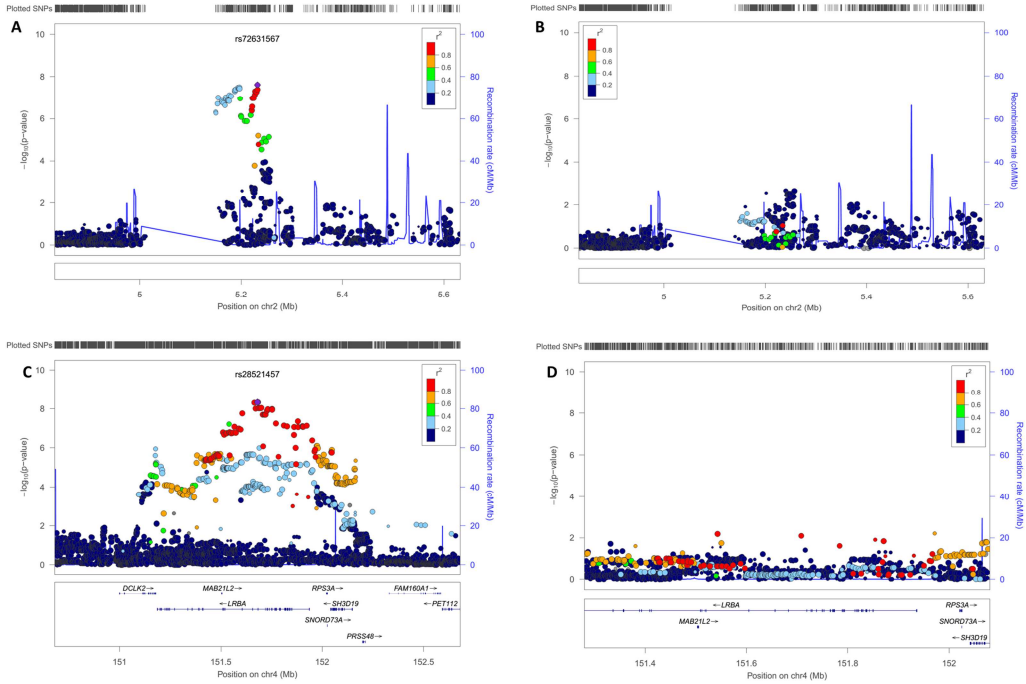

Figure S2. Regional Manhattan plots for chromosome 2 and 4 in the region of the rs72631567 and rs28521457 signals. (A) Chromosome 2 for the overall cohort, (B) Chromosome 2 for the overall cohort conditioned on rs72631567. (C) Chromosome 4 for the HC only cohort, (D) Chromosome 4 for the HC only cohort conditioned on rs28521457.

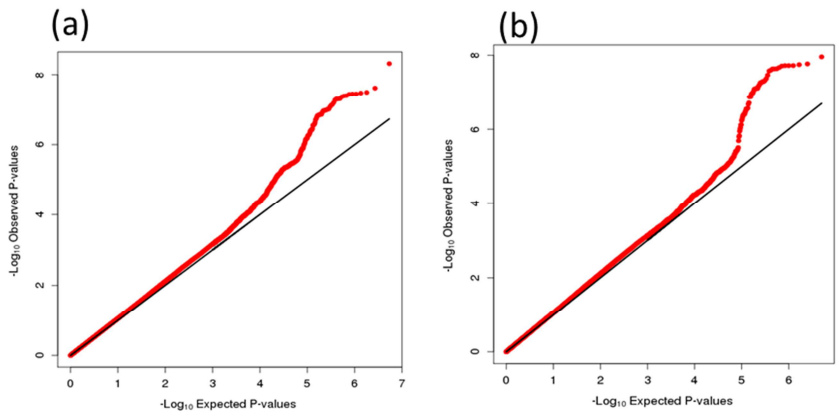

Figure S3. QQ plots. (a) for the overall original analysis (b) after eliminating variants in MHC region. The QQ plot in the (b) panel highlights the signal on chromosome 2. The inflation factor is 1.05 after correction and 4.65 before correction.

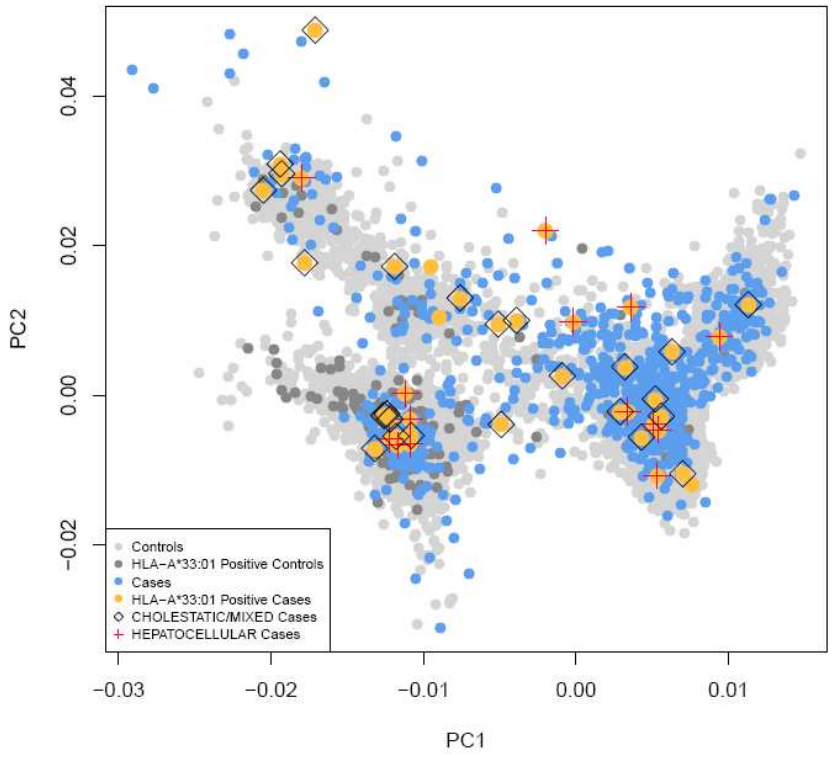

Figure S4. Scatterplot representing the first two principal components of the current study cohort. The A*33:01-positive cases and their injury type are highlighted. A*33:01-positive cases are homogenously distributed in all the major population clusters.

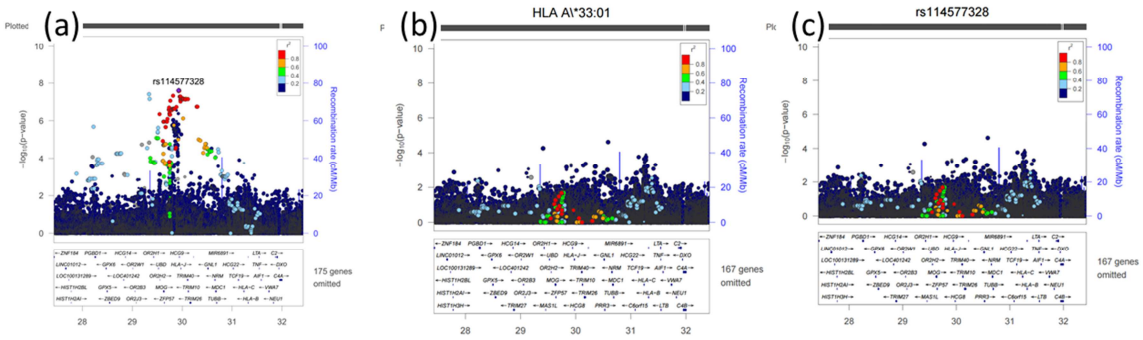

Figure S5. MHC Regional Manhattan plots for (a) the overall cohort (b) the same cohort conditioned on A*33:01 (purple dot) and (c) the same cohort conditioned on rs114577328 (purple dot).

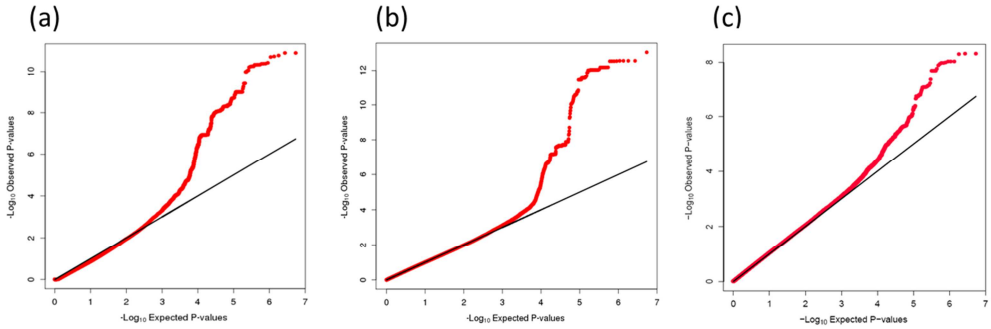

Figure S6. QQ plot for (a) Terbinafine cases only, (b) Cholestatic/Mixed cases only, (c) Hepatocellular cases only.

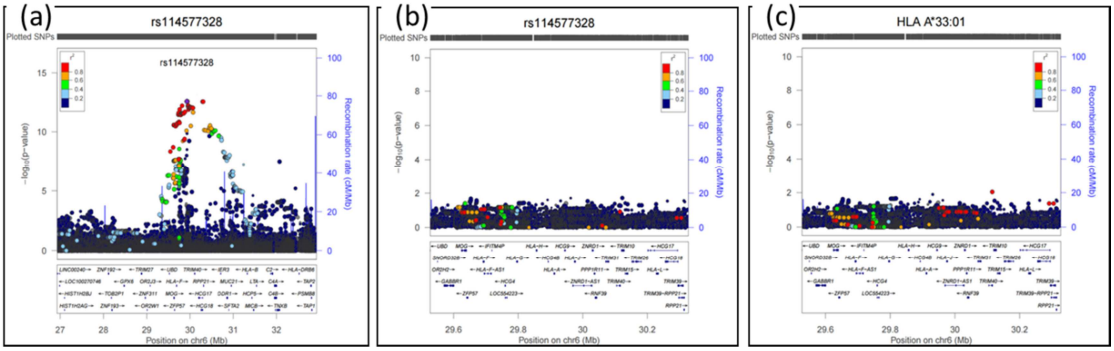

Figure S7. MHC Region Manhattan plots for (a) all cholestatic/mixed cases (b) the same cases conditioned on rs114577328 (purple dot), the top SNP (c) the same cases conditioned on A*33:01 (purple dot), the top HLA allele

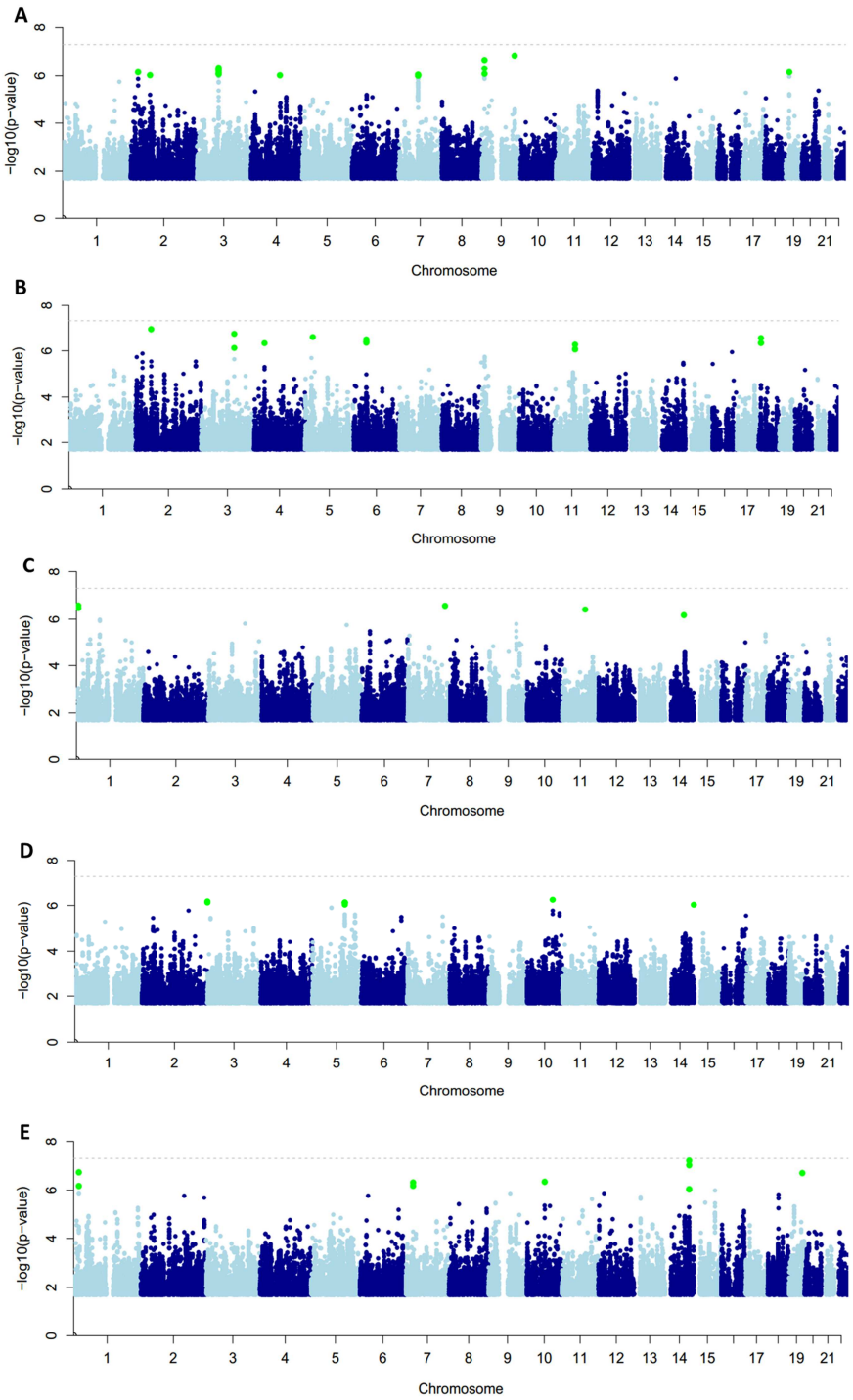

Figure S8. Manhattan plots for DILI due to (A) Anti-TB drugs, (B) Fluoroquinolones, (C) NSAIDs, (D) Diclofenac (E) Nitrofurantoin.

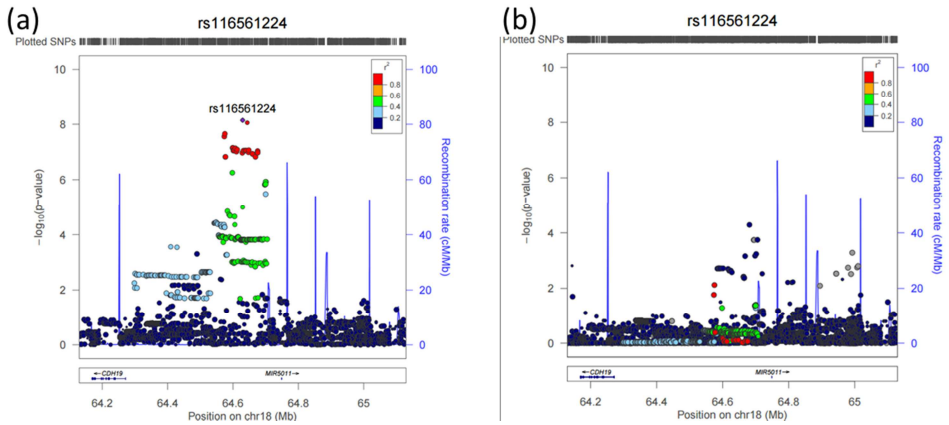

Figure S9. Regional Manhattan plots for chromosome 18 in the area of the rs116561224 signal for (a) statin-induced DILI cases and (b) the same cases with conditioning on rs116561224

Table S1. Causal drugs in the replication cohort.

A

DRUG	PD	SNP typing	HLA typing	TOT
Sulfamethoxazol	yes	17		17
Nitrofurantoin	yes	15		15
Isoniazid	yes	11		11
Ciprofloxacin	yes	9		9
Azathioprine	yes	8		8
Atorvastatin	yes	8		8
Minocycline	yes	7		7
Infliximab	yes	7		7
Cefazolin	yes	6		6
Terbinafine	yes	6	2	8
Levofloxacin	yes	6		6
Azithromycin	yes	5		5
Carbamazepine	yes	5		5
Mercaptopurine	yes	4		4
Fenofibrate	yes	4	3	7
Oxaliplatin		4		4
Lisinopril		4		4
Methylprednisolone		3		3
Exemestane		3		3
Flavocoxid		3		3
Metformin		3		3
Pravastatin	yes	3		3
Sulfasalazine		3		3
Vancomycin		3		3
Sertraline	yes	1	3	4
Methyldopa	yes	1	2	3
Erytromycin	yes	-	1	1
Others (98 drugs)		123		123
TOTAL		272	11	283

B

DRUG	PD	DILIN	iDILIC	TOT
Atorvastatin	yes	8	8	16
Pravastatin	yes	3	3	6
Fluvastatin	yes		1	1
Rosuvastatin	yes	1		
Lovastatin	yes	1		
Simvastatin	yes	1	3	4
Total		**14**	**15**	**27**

DRUG = causal drug; PD=presence of the drug in the discovery cohort; TOT = total number of cases. The list in A panel includes only drugs for which at least 3 cases were available. The category "other" includes a total of 98 different drugs. The drugs highlighted in bold are the drugs found to be associated with A*33:01.

Panel A shows the breakdown of causal drugs for the replication cohort and the methods utilized to replicate the main results. 272 DILIN samples underwent to direct SNPs typing while 12 samples whose DILI was due to A*33:01-associated drugs underwent HLA typing. Panel B shows the breakdown of causal drugs for the statin-specific replication cohort: besides the 14 samples previously collected within 272 DILIN samples we add extra 15 iDILIC statin cases.

Table S2. Genotyping details for the DILI control cohorts

COHORT	#SAMPLEs	CHIP
Welcome Trust Case Control Consortium (WTCCC)	4824	Illumina 1M BeadChip
Spanish Cohort (phs000346.v1)	2077	Illumina 1M BeadChip
Hypergenes	901	Illumina 1M BeadChip
Swedish Twin Registry	1499	Illumina HumanOmniExpress BeadChip
National Spanish DNA Bank	209	Illumina 1M Duo/Illumina Infinium HumanCoreExome BeadChip
	173	Illumina Infinium HumanCoreExome BeadChip
iSAEC Italian Penicillin Tolerant Controls	147	Illumina HumanOmniExpress BeadChip
PGX40001	103	Illumina 1M Duo BeadChip
POPulation REference Sample (POPRES)	655	Illumina 1M Duo BeadChip

Table S3. Causative drugs across the overall DILI cohort

DRUGs	Phase1	Phase2	Combined	P
DICLOFENAC	29	38	67	0.37
NITROFURANTOIN	28	36	64	0.44
ISONIAZID	16	20	36	0.61
AZATHIOPRINE	6	21	27	<0.01
MINOCYCLINE	10	17	27	0.24
SULFAMETHOXAZOLE/TRIMETHOPRIM	12	14	26	0.84
ATORVASTATIN	4	19	23	<0.01
CIPROFLOXACIN	5	16	21	0.03
NIMESULIDE	12	8	20	0.37
VALPROICACID	14	4	18	0.02
SIMVASTATIN	5	12	17	0.14
ISONIAZID/PYRAZINAMIDE/RIFAMPIN	6	9	15	0.80
IBUPROFEN	4	10	14	0.18
TERBINAFINE	4	10	14	0.18
AZITHROMYCIN	3	10	13	0.09
CEFAZOLIN	7	6	13	0.79
ISONIAZID/PYRAZINAMIDE/RIFAMPIN/ETHAMBUTOL	6	6	12	1.00
LEVOFLOXACIN	5	6	11	1.00
PHENYTOIN	9	2	11	<0.01
ERYTHROMYCIN	2	8	10	0.11
IMATINIB	2	8	10	0.11
CELECOXIB	9		9	<0.01
MERCAPTOPURINE	5	4	9	0.75
NAPROXEN	2	7	9	0.18
METHIMAZOLE	2	6	8	0.29
ANABOLIC STEROID	7		7	<0.01
CARBAMAZEPINE	2	5	7	0.28
DULOXETINE	7		7	<0.01
ESTRADIOL/LEVONORGESTREL	2	5	7	0.45
FENOFIBRATE	4	3	7	0.72
FLUVASTATIN	2	5	7	0.45
METHOTREXATE	3	4	7	1.00
MOXIFLOXACIN	2	5	7	0.45
ROFECOXIB	4	3	7	0.72
TELITHROMYCIN	7		7	<0.01
DISULFIRAM	1	5	6	0.22
FLUPIRTIN		6	6	0.03
INFLIXIMAB	1	5	6	0.22
LAMOTRIGINE	5	1	6	0.12
DOXYCYCLINE	2	3	5	1.00
OMEPRAZOLE	2	3	5	1.00
PIROXICAM	3	2	5	0.68
ROSUVASTATIN	1	4	5	0.37
SERTRALINE		5	5	0.06
TICLOPIDINE	1	4	5	0.37
ENALAPRIL	1	3	4	0.62
METHYLDOPA	4		4	0.06
MONTELUKAST	1	3	4	0.62
NICOTINICACID	4		4	0.06
PRAVASTATIN	2	2	4	1.00
SEVOFLURANE	1	3	4	0.62
VENLAFAXINE		4	4	0.12
OTHER	135	71	204	<0.01
TOTAL	**411**	**451**	**862**	-

Phase 1 = number of cases extracted from previous DILI study (Urban TJ et al. Pharmacogenet Genomics 2012;22:784-95 (reference 12 in main text); Phase 2 = number of new cases; Combined = total number of cases; P = Fisher's Exact test p-value to test the disproportions between the two cohorts.

The list includes only drugs for which at least 4 cases were available. The category "other" includes a total of 140 different drugs.

Table S4. Effect of the A*33:01, rs72631567 and rs28521457 signals across populations and recruitment phases

Marker	COHORTs	PHASE	OR	95% CI	P	AF Cases	AF Controls
A*33:01	European DILI cohort	I	2.29	1.31-4	0.003	0.02	0.01
		II	2.96	1.94-4.52	$4.7*10^{-7}$	0.03	-
	Cholestatic and Mixed DILI cohort	I	4.73	2.38-9.38	$8.4*10^{-6}$	0.03	-
		II	5.76	3.36-9.88	$1.8*10^{-10}$	0.04	-
	North European Cohort	I +II	2.64	1.48-4.69	0.001	0.01	0.00
	Spanish Cohort	I +II	1.96	1.02-3.75	0.04	0.05	0.03
	Italian Cohort	I +II	5.69	2.43-13.33	$6.1*10^{-5}$	0.07	0.01
rs72631567	European DILI cohort	I	2.02	1.45-2.82	$3.6*10^{-5}$	0.05	0.03
		II	1.99	1.45-2.71	$1.1*10^{-5}$	0.05	-
	North European Cohort	I +II	1.65	1.215-2.236	0.001	0.04	0.03
	Spanish Cohort	I +II	2.12	1.21-3.75	0.009	0.06	0.03
	Italian Cohort	I +II	3.21	1.57-6.52	0.001	0.09	0.03
rs28521457	European DILI cohort	I +II	1.56	1.26-1.94	$5.0*10^{-5}$	0.06	0.04
	Hepatocellular DILI cohort	I	2.48	1.74-3.54	$4.6*10^{-7}$	0.09	-
		II	1.93	1.32-2.65	0.0003	0.07	-
	North European Hepatocellular DILI	I +II	1.91	1.41-2.56	$2.0*10^{-5}$	0.07	0.04
	Spanish Hepatocellular DILI	I +II	2.25	1.2-4.23	0.01	0.08	0.02
	Italian Hepatocellular DILI	I +II	1.93	0.44-8.46	0.38	0.08	0.04

PHASE=recruitment phase; OR=Odds ratio; 95%CI=95% Confidence Interval; P=logistic p-value; AF=Allele Frequency

Table S5. Missing genotypes rate for the most associated SNPs within case and control groups in the comparisons where SNPs were significant

SNP	**CHR**	***missing rate in cases***	***missing rate in controls***	***P***
rs114577328	6	0	0.0002	1.0
rs72631567	2	0.007	0.005	0.3
rs28521457	4	0	0.0009	1.0

P= X^2 p-value of missing genotype rate between cases and controls

Table S6. Causative drugs across the rs72631567 signal on chromosome 2

DRUGS	No. Cases Tested	OR	95% CI	P	AF
FLUVASTATIN	7	10.15	2.58-39.85	0.0009	0.21
FENOFIBRATE	7	8.85	1.80-43.63	0.0074	0.17
LAMOTRIGINE	6	9.40	1.53-57.61	0.02	0.17
CIPROFLOXACIN	21	7.41	3.16-17.36	$4.0x10^{-6}$	0.14
ISONIAZID/PYRAZIN	12	5.80	1.70-19.75	0.005	0.14
AZITHROMYCIN	13	4.92	1.39-17.39	0.01	0.12
MERCAPTOPURINE	9	4.16	0.51-34.08	0.2	0.11
ATORVASTATIN	23	3.33	1.16-9.55	0.02	0.09
ISONIAZID	36	3.42	1.45-8.07	0.005	0.08
NITROFURANTOIN	64	2.33	1.08-5.0	0.03	0.06
CONTROL	10588	-	-	-	0.03

OR=Odds Ratio; 95%CI=95% confidence interval; P=logistic p-value; AF=Minor Allele Frequency in cases;

Table S7. Summary of A*33:01-B*14:02-C*08:02 haplotype specific analysis across study cohorts

DRUGs	*OR*	*P*	*HF CAs*	*HF CTLs*
OVERALL ANALYSIS	2.7	$1.8x10^{-7}$	0.02	0.009
CHOLESTATIC-MIXED ANALYSIS	5.7	$3.9x10^{-13}$	0.04	
TERBINAFINE	49.2	$9.5x10^{-11}$	0.21	
TICLOPIDINE	201.0	$7.2x10^{-6}$	0.40	
FENOFIBRATE	68.5	$1.1x10^{-7}$	0.29	
ERITHROMYCIN	13.1	0.002	0.10	
ENALAPRIL	11.4	0.1	0.13	
METHILDOPA	41.5	0.002	0.13	
SERTRALINE	11.6	0.04	0.10	

HF CAs=Haplotype Frequency in cases; HF CTLs=Haplotype Frequency in controls; OR=Odds Ratio; P=logistic p-value

Table S8. Summary of the validation by direct HLA typing

	Predictions	Validations		
COHORT	**#subjects (#CARRIERs)**	**# subjects**	**Not-CARRIERs (FN)**	**CARRIERs (FP)**
TERBINAFINE	14 (6)	13	6 (1)	6 (0)
FENOFIBRATE	7 (3)	6	3 (0)	3 (0)
METHYLDOPA	4 (2)	3	2 (0)	0 (1)
SERTRALINE	5 (2)	3	3 (0)	-
ENALAPRIL	4 (2)	2	1 (0)	1 (0)
ERYTHROMYCIN	10 (2)	6	6 (0)	-
TICLOPIDINE	5 (4)	2	-	2 (0)

FN=False Negative; FP=False Positive

Table S9. The A*33:01 signal across the main six A*33:01-associated drugs by type of injury

DRUG	TI	No. Cases Tested	OR	95% CI	P
TICLOPIDINE	CM*	3	93.5	18.78-465.9	2.37E-05
	HC	2	36.5	1.77-750.9	0.02
METHYLDOPA	CM	-	-	-	-
	HC	3	54.4	4.7-635.7	0.001
FENOFIBRATE	CM	7	58.7	12.3-279.8	$3.2*10^{-7}$
	HC	0	-	-	-
TERBINAFINE	CM	9	88.1	19.28-402.4	7.57E-09
	HC	5	-	-	-
ENALAPRIL	CM*	3	46.8	8.52-256.6	1.65E-03
	HP	1	-	-	-
SERTRALINE	CM	1	-	-	-
	HC	4	40.1	4.8-335.9	0.0006
ERYTHROMYCIN	CM	4	24.1	2.2-264	0.009
	HC	5	9.2	0.9-91.1	0.06
OTHERS	ALL	815	1.4	0.9-2.2	0.17
	CM	279	2.6	1.4-4.9	0.003
	HC	438	1.0	0.5-2	0.9

#DRUG=causal drug; TI = type of injury; OR=Odds Ratio; 95% CI = 95% confidence interval; P=logistic p-value

*For groups with 3 subjects the association has been tested by Fisher's Exact test.

Table S10. List of all causal drugs where at least one case carries a A*33:01 allele

DRUGs	# Carriers
DRONEDARONE	1
PIPERACILLINSODIUM/TAZOBACTAM	1
LAMOTRIGINE	1
MOXIFLOXACIN	1
METHIMAZOLE	1
CELECOXIB	1
IBUPROFEN	1
CEFAZOLIN	1
ISONIAZID	2
GENERIC COMBINATIONS OF NUTRIENTS	1
VALPROIC ACID	1
DICLOFENAC	3
ATORVASTATIN	1
AZATHIOPRINE	1
SULFAMETHOXAZOLE/TRIMETHOPRIM	1
NITROFURANTOIN	2

#Carriers = total number of A*33:01 positive carriers

Table S11. The most represented causative drugs across the rs28521457 signal on chromosome 4

DRUGS	No. Cases Tested	OR	95% CI	P	AF
CELECOXIB	2	50.56	3.378-756.8	0.004	0.50
EBROTIDINE	3	11.42	2.098-62.18	0.005	0.33
NIMESULIDE	4	8.51	1.73-41.82	0.008	0.25
ISONIAZID/PYRAZINAMIDE/RIFAMPIN/ETHAMBUTOL	9	7.39	2.406-22.7	0.0005	0.22
MERCAPTOPURINE	7	6.65	1.835-24.1	0.004	0.21
TELITHROMYCIN	5	6.22	1.36-28.43	0.02	0.20
IMATINIB	10	4.59	1.308-16.08	0.02	0.15
DICLOFENAC	35	2.40	1.024-5.604	0.04	0.09
SIMVASTATIN	12	2.19	0.5164-9.264	0.29	0.08
MINOCYCLINE	19	2.17	0.6713-7.028	0.20	0.08
ISONIAZID	32	2.15	0.8605-5.351	0.10	0.08
OTHERS	119	-	-	-	0.16
CONTROLS	10588	-	-	-	0.04

OR=Odds Ratio; 95%CI=95% confidence interval; P=logistic p-value; AF=Allele Frequency in cases

The table shows the causal drugs with more than two positive carriers only.

Table S12. Summary of drug and class comparisons with more than 40 samples

DRUG/CLASS	#CASES
NSAIDs	144
ANTI TUBERCULOSIS DRUGs	67
DICLOFENAC	67
NITROFURANTOIN	64
STATINs	59
FLUOROQUINOLONEs	43

Table S13. The most associated variants for each drug in the class-specific analysis

COMPARISON	SNP	CHR	BP	OR	95% CI	P
NITROFURANTOIN	rs72696020	14	88571907	7.18	3.52-14.66	6.16E-08
NITROFURANTOIN	rs72696089	14	88588493	6.98	3.42-14.24	9.48E-08
NITROFURANTOIN	rs6694270	1	19120377	2.59	1.81-3.71	1.85E-07
NITROFURANTOIN	rs10404821	19	51161088	2.56	1.80-3.65	1.98E-07
NITROFURANTOIN	rs61858823	10	66855253	5.63	2.88-11.02	4.52E-07
DICLOFENAC	rs114811931	5	160684731	6.59	3.54-12.28	2.76E-09
DICLOFENAC	rs113206698	10	94577904	3.79	2.25-6.39	5.61E-07
DICLOFENAC	rs115266745	3	821213	6.80	3.19-14.48	6.59E-07
DICLOFENAC	rs149014830	5	120876337	6.04	2.96-12.3	7.24E-07
DICLOFENAC	rs116316305	5	120853119	6.03	2.96-12.28	7.4E-07
ANTI TUBERCULOSIS DRUGs	rs117491755	9	119643656	3.92	2.35-6.51	1.43E-07
ANTI TUBERCULOSIS DRUGs	rs143575776	9	9593742	4.96	2.70-9.09	2.17E-07
ANTI TUBERCULOSIS DRUGs	rs73122578	3	79359633	2.42	1.71-3.40	4.48E-07
ANTI TUBERCULOSIS DRUGs	rs78671883	9	9558149	4.65	2.55-8.47	4.84E-07
ANTI TUBERCULOSIS DRUGs	rs73124503	3	79368237	2.42	1.71-3.41	5.12E-07
FLUOROQUINOLONEs	rs186920977	2	56649930	7.22	3.47-15.01	1.17E-07
FLUOROQUINOLONEs	rs144941777	2	56672204	7.22	3.47-15.01	1.17E-07
FLUOROQUINOLONEs	rs191153876	3	123637182	7.33	3.46-15.5	1.85E-07
FLUOROQUINOLONEs	rs116606120	5	28665952	8.60	3.79-19.5	2.54E-07
FLUOROQUINOLONEs	rs112655218	18	9841515	4.54	2.55-8.09	2.8E-07
NSAIDs	rs185305928	1	6905711	4.55	2.55-8.11	2.66E-07
NSAIDs	rs2240395	7	139718147	1.84	1.45-2.31	2.72E-07
NSAIDs	rs113607154	1	7004613	4.42	2.47-7.82	3.37E-07
NSAIDs	rs597480	11	85436868	1.84	1.45-2.32	3.9E-07
NSAIDs	rs2025009	14	68843605	1.81	1.43-2.29	6.77E-07

COMPARISON = causal drug/class; CHR=chromosome; BP=base-pair position; OR=Odds Ratio; 95%CI=95% confidence interval; P=logistic p-value

Table S14. Causative drugs associated with rs116561224 signal in the statin cohort

DRUG	No. Cases Tested	AF
SIMVASTATIN	17	0.21
ROSUVASTATIN	5	0.20
PRAVASTATIN	4	0.13
ATORVASTATIN	22	0.09
FLUVASTATIN	7	0.07
LOVASTATIN	3	0.00

DRUG = Causal drug; AF = Allele frequency

Table S15. Causative drugs associated with the rs114577328/A*33:01 signal in the additional case set

DRUG	PD	DC	SNP #TOT (#CARRIERs)	HLA #TOT (#CARRIERs)	TOT	CF
Terbinafine	yes	yes	6 (5)	2 (0)	8	0.63
Azathioprine	yes	yes	8 (2)	-	8	0.25
Exemestane			3 (2)	-	3	0.67
Amiodarone	yes		2 (1)	-	2	0.50
Daptomycin			2 (1)	-	2	0.50
Fenofibrate	yes	yes	4 (0)	3 (0)	7	0
Erytromycin	yes	yes		1 (0)	1	0
Ticlopidine	-	-	-	-	-	-
Methildopa	yes	yes	1 (0)	2 (0)	3	0
Enalapril	-	-	-	-	-	-
Sertraline	yes	yes	1 (1)	3 (2)	4	0.75

DRUG = causal drug; PD=presence of the drug in the discovery cohort; DC = Drug with at least one positive carrier in the discovery cohort; SNP #TOT (#CARRIERs) = total number of cases (total number of rs114577328/A*33:01 carriers) among the samples who underwent SNP typing; HLA #TOT (#CARRIERs) = total number of cases (total number of rs114577328/A*33:01 carriers) among the samples who underwent HLA typing; TOT= total number of samples that underwent SNP and HLA typing; CF = carriage frequency.

Table S16. Causative drugs associated with the chromosome 2 rs72631567 signal in the additional case set

DRUG	PD	DC	No. Cases Tested	AF
Cefotetan			1	0.50
Amiodarone	yes		2	0.25
Escitalopram	yes		2	0.25
Ethinylestradiol/Norgestimate	yes	yes	2	0.25
Mercaptopurine	yes	yes	4	0.13
Cefazolin	yes	yes	6	0.08
Minocycline	yes	yes	7	0.07
Azathioprine	yes	yes	8	0.06
Atorvastatin	yes	yes	8	0.06
Ciprofloxacin	yes	yes	9	0.06
Isoniazid	yes	yes	11	0.05
Sulfamethoxazole/Trimethoprim	yes	yes	17	0.03

DRUG = causal drug; PD=presence of the drug in the discovery cohort; DC = Drug with at least one positive carrier in the discovery cohort; TOT = total number of cases; AF = Minor allele frequency. Drugs also associated with an increased AF in the discovery cohort are indicated in bold.

Table S17. Causative drugs associated with the chromosome 4 rs28521457 signal in the additional case set

DRUG	PD	DC	No. Cases Tested	AF
Minocycline	yes	yes	7	0.07
Clindamycin	yes		1	0.50
Dronedarone	yes	yes	2	0.25
Ketoconazole			1	0.50
Methylprednisolone			3	0.17
Nefazodone			1	0.50
Nitrofurantoin	yes	yes	11	0.05

DRUG = causal drug; PD=presence of the drug in the discovery cohort; DC = Drug with at least one positive carrier in the discovery cohort; TOT = total number of cases; AF = Minor allele frequency.