Supplementary Material

Genetic risk factors in drug-induced liver injury due to isoniazid-containing anti-tuberculosis drug regimens.

Table of contents

Supplementary Methods

Selection of Indian descent individuals from biome data, 1KG and study cohort:

To assess the extent of population structure of the Indian cohort and matched cases with appropriate controls, and derive eigenvectors to account for confounding, we applied principal component analysis on Indian DILI cohort and BioMe cohort, using the smartPCA program from the EIGENSTRAT package (version 3.0)[1] on 58.000 overlapping SNPs (minor allele frequency, MAF>0.05) across the range of used genotyping arrays. First, each cohort were merged with 1000 Genome samples to find and select South Asian ethnic group by K-means clustering. Then, we repeated Eigenstrat's principal component analysis on 1,578 predicted South-Asian samples across cohorts to define population structure within the Indian group and to eliminate outliers. Outlier samples were defined as samples with principal component (PC) values deviating by 3 standard deviations from the mean of that main PCs axes.

1 Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. . *Nature Genetics* **38**, 904-909 (2006).

Supplementary Tables

Table S1: Origins of cases in the European group.

Study	No. of cases	Initial Reference
DILIGEN	12	Ng C.S. *et al. Eur J Clin Pharmacol* 70, 1079-86 (2014).
DILIN	16	Urban T.J. *et al. Pharmacogenet Genomics* 22, 784-95 (2012).
DILIN	6	Nicoletti P. *et al. Gastroenterology* 152, 1078-89 (2017).
iDILIC	28	Nicoletti P. *et al. Gastroenterology* 152, 1078-89 (2017).
DILIN	8	Cirulli E.T. *et al. Gastroenterology* 156, 1707-16 (2019).
Total	**70**	

Once described, each sample was included in all the subsequent studies listed.

Table S2: Haplotype analysis for HLA-B*52:01 and HLA-C*12:02 in Indian and European cohorts

	Indian cohort				Caucasia cohort			
Marker	**AF cases**	**AF controls**	**OR**	**P**	**AF cases**	**AF controls**	**OR**	**P**
HLA-B*52:01/HLA-C*12:02	0.13	0.07	1.83	0.046	0.04	0.01	6.58	0.00008
HLA-C*12:02	0.01	0.01	0.550	0.55	0	0	0	1.00
HLA-B*52:01	0.01	0.005	2.790	0.34	0	0	0	1.00
None	0.85	0.91	0.61	0.08	0.96	0.99	0.160	0.0001

OR = Odds ratio of a multivariate regression model correcting for population stratification; AF Cases = allele frequency in cases; AF Controls = allele frequency in controls; P = multinomial p value.

Table S3: The most significant HLA allele associations in Europeans stratifying the cohort by two groups: 1) Isoniazid alone (n=43) and in combination (n=27).

	cases treated with Isoniazid in combination			cases treated with Isoniazid alone		
MARKER	**OR**	**95%CI**	**P**	**OR**	**95%CI**	**P**
HLA-DRB1*15:02	7.176	1.67-30.88	0.008	5.789	1.73-19.41	0.004
HLA-B*52:01	11.09	3.27-37.6	0.0001	3.82	0.9-16.26	0.07
HLA-C*12:02	11.31	3.35-38.25	9,7x10-7	3.797	0.89-16.15	0.07

OR = Odds ratio of a multivariate regression model correcting for population stratification; 95%CI = confidence interval of the Odds Ratio; AF Cases = allele frequency in cases; AF Controls = allele frequency in controls; P = multinomial p value.

Table S4: Frequency of individual NAT2 genotypes in European and Indian case/control cohorts and their association in a multivariate regression model

		Europeans (70 cases)					Indians (55 cases)					Meta-analysis				
Genotypes	**Acetylator status**	**AF controls**	**AF cases**	**OR**	**95%CI**	**PV**	**AF controls**	**AF cases**	**OR**	**95%CI**	**PV**	**Direction of effec**	**OR**	**95%CI**	**PV_m**	**HetPV**
NAT2*4/*4	rapid	0.05	0.06	1.18	0.43-3.28	0.73	0.06	0.02	0.29	0.03-2.20	0.23	Discordant				
NAT2*4/*5	intermidiate	0.20	0.19	0.88	0.40-1.62	0.70	0.15	0.07	0.49	0.17-1.39	0.18	Concordant	0.65	0.45-0.95	0.42	0.25
NAT2*4/*6	intermidiate	0.14	0.16	1.16	0.60-2.22	0.64	0.16	0.18	1.06	3.93-7.31	0.86	Concordant	1.11	0.65-1.88	0.62	0.99
NAT2*4/*7	intermidiate	0.01	0.01	1.42	0.19-10.40	0.72	0.03	0.04	1.18	0.27-5.10	0.81	Concordant	1.24	0.28-5.58	0.68	0.91
NAT2*5/*5	slow	0.21	0.16	0.68	0.33-1.29	0.24	0.11	0.04	0.31	0.07-1.30	**0.11**	Concordant	0.46	0.32-0.62	0.10	0.26
NAT2*5/*6	slow	0.26	0.23	0.83	0.47-1.45	0.52	0.26	0.35	1.49	0.83-2.68	0.17	Discordant	0.94	0.68-1.40	0.29	0.28
NAT2*5/*7	slow	0.02	0.01	0.69	0.09-5.02	0.71	0.05	0.02	0.38	0.05-2.84	0.34	Concordant	0.44	0.22-0.85	0.51	0.44
NAT2*6/*6	Ultra Slow	0.08	0.13	1.60	0.79-3.24	0.19	0.13	0.18	1.53	0.74-3.14	0.24	Concordant	1.56	0.71-3.42	0.10	0.50
NAT2*6/*7	Ultra Slow	0.02	0.06	3.59	1.28-10.00	**0.01**	0.06	0.11	2.38	0.96-5.88	**0.06**	Concordant	2.64	0.42-16.73	0.002	0.35
NAT2*7/*7	Ultra Slow	0.0002	0.00			-	0.01	0			-					

AF = allele frequency; OR = Odds ratio of a multivariate regression model correcting for population stratification; 95%CI = confidence interval of the Odds Ratio; PV = multinomial p value; PV_m = meta-analysis p value; HetPV= Heterogeneity p value.

Table S5: Summary of findings for selected candidate genes

European associations

SNP	**CHR**	**POS**	**GENE**	**OR**	**LCI**	**UCI**	**P**
rs6413420	10	135340829	CYP2E1	0.60	0.25	1.48	0.27
rs72862138	10	135341973	CYP2E1	0.81	0.20	3.32	0.77
rs9919386	10	135345603	CYP2E1	0.65	0.16	2.64	0.55
rs6413419	10	135345675	CYP2E1	0.65	0.16	2.64	0.55
rs6413421	10	135345811	CYP2E1	0.61	0.25	1.48	0.27
rs2515641	10	135351362	CYP2E1	0.51	0.26	1.02	0.06
rs7081484	10	135352490	CYP2E1	0.71	0.17	2.92	0.64
rs2480257	10	135352509	CYP2E1	0.67	0.42	1.08	0.10
rs2480256	10	135352514	CYP2E1	0.67	0.42	1.08	0.10
rs12721607	3	119526203	PXR/NR1I2	0.32	0.04	2.27	0.25
rs3732356	3	119529113	PXR/NR1I2	1.05	0.56	1.99	0.88
rs6797879	3	119533773	PXR/NR1I2	1.26	0.65	2.46	0.50
rs2276707	3	119534153	PXR/NR1I2	0.94	0.60	1.45	0.77
rs1232663050	16	66975195	CES2	0.43	0.06	3.11	0.41
rs11568309	16	66977997	CES2	1.45	0.75	2.78	0.27

Indian associations

SNP	**CHR**	**POS**	**GENE**	**OR**	**LCI**	**UCI**	**P**
rs9919386	10	135345603	CYP2E1	0.62	0.08	4.75	0.65
rs6413419	10	135345675	CYP2E1	0.62	0.08	4.75	0.65
rs2515641	10	135351362	CYP2E1	1.42	0.88	2.29	0.15
rs2480257	10	135352509	CYP2E1	0.93	0.61	1.42	0.74
rs2480256	10	135352514	CYP2E1	0.93	0.61	1.42	0.75
rs3841391	3	119501580	PXR/NR1I2	1.07	0.70	1.63	0.76
rs116598923	3	119501748	PXR/NR1I2	0.39	0.09	1.65	0.20
rs79023775	16	55846306	CES1	0.54	0.13	2.21	0.39
rs115629050	16	55853545	CES1	0.90	0.33	2.50	0.85
rs2307234	16	55855248	CES1	0.91	0.33	2.50	0.85
rs2307227	16	55855361	CES1	0.91	0.33	2.50	0.85
rs750392264	16	66976576	CES2	0.39	0.05	2.88	0.35
rs1232663050	16	66977895	CES2	0.73	0.39	1.38	0.34
rs11568309	16	66977997	CES2	1.25	0.49	3.18	0.64

CHR, chromosome; POS, position; OR, odds ratio; LCI, lower limit of confidence interval; UCI, upper limit of confidence interval.

Supplementary Figures:

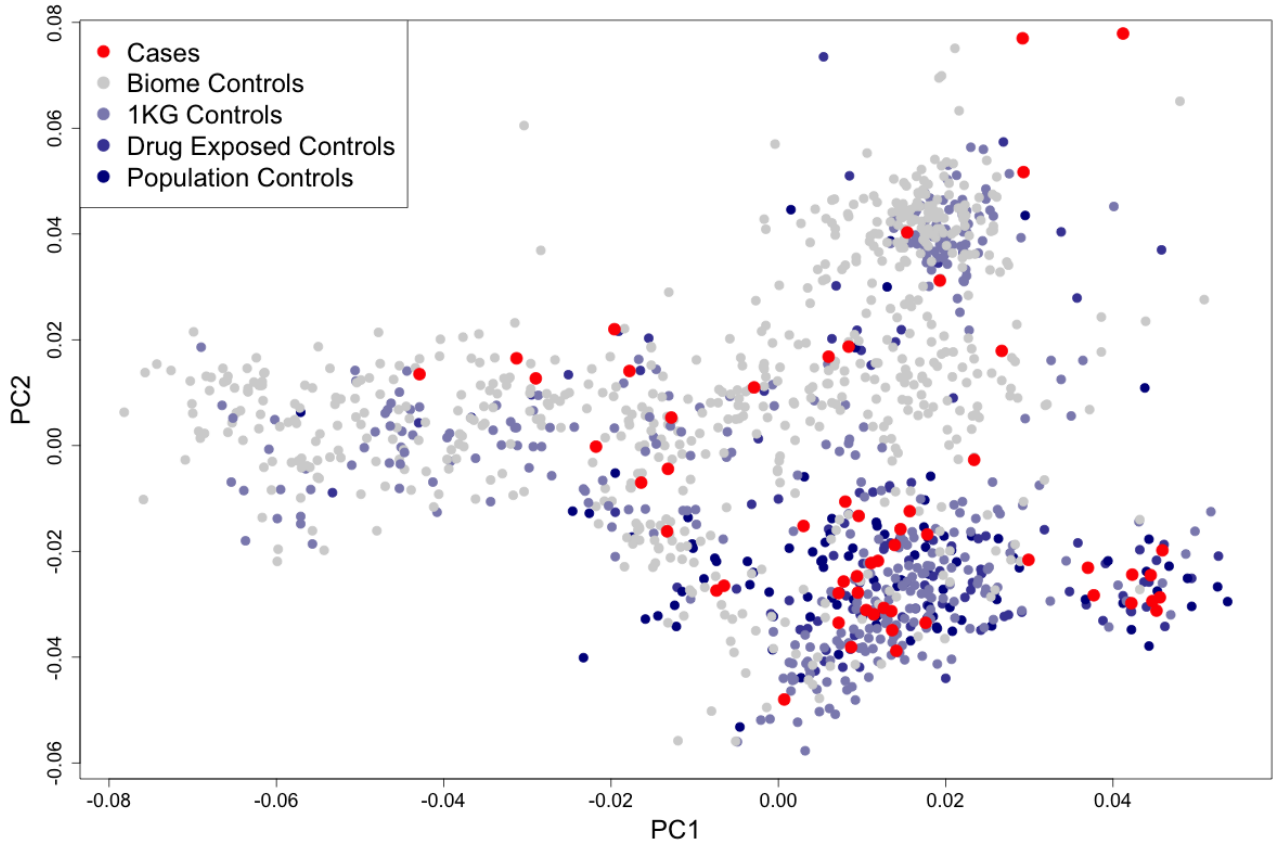

Figure S1: Scatterplot representing the first two principal components of the Indian study cohort. Red dots are cases and the coloured dots are controls. Controls are divided based on being treated with ATD, healthy adults recruited for this study, 1000G project samples and BioMe samples.

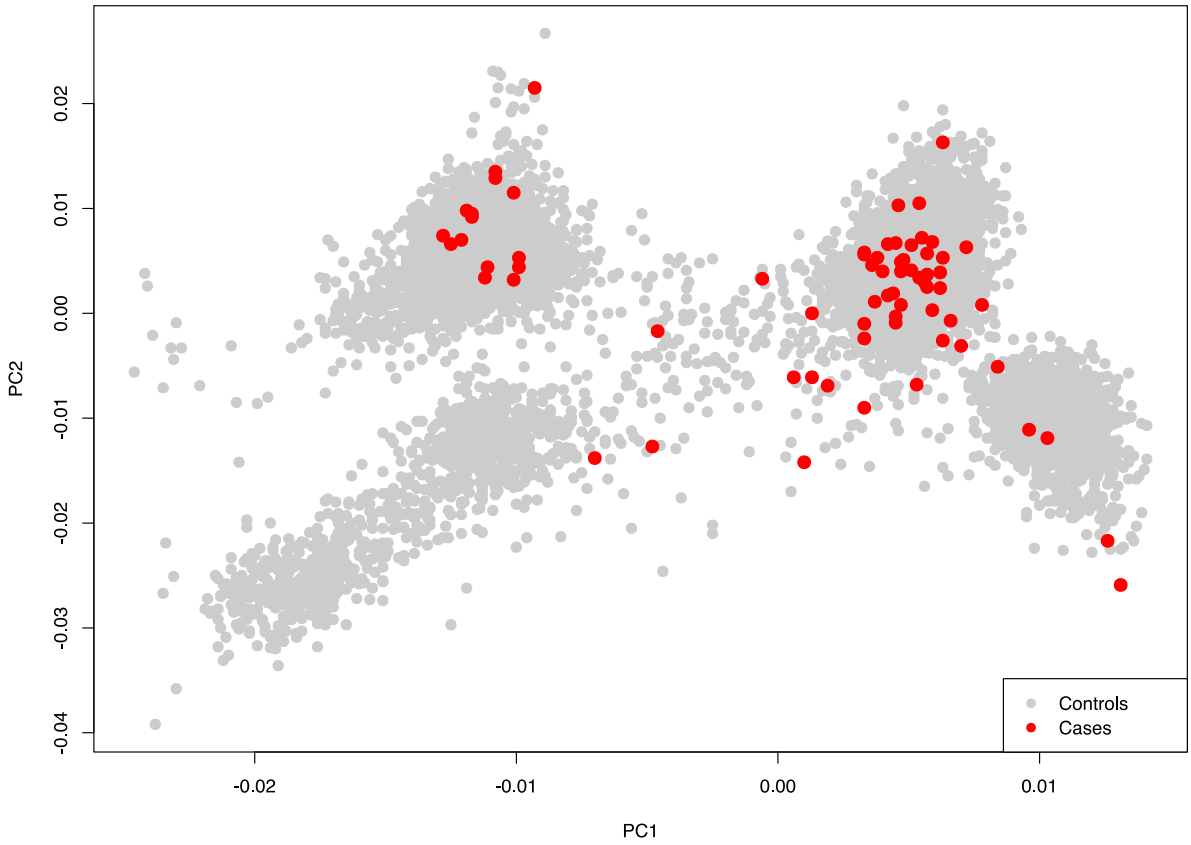

Figure S2: **Scatterplot representing the first two principal components of the Caucasian study cohort**. Red dots are cases and the grey dots are controls. The controls cluster in four groups representing the Italian, Spanish, UK and Swedish major control populations.