Genome-wide association study of circulating liver enzymes reveals an expanded role for manganese transporter SLC30A10 in liver health

Lucas D. Ward[1], Ho-Chou Tu[1], Chelsea Quenneville[1], Alexander O. Flynn-Carroll[1], Margaret M. Parker[1], Aimee M. Deaton[1], Patrick A. J. Haslett[1], Gregory Hinkle[1], Paul Nioi[1]

1. Alnylam Pharmaceuticals, Cambridge, MA 02142

Abstract

To better understand molecular pathways underlying liver health and disease, we performed genome-wide association studies (GWAS) on circulating levels of alanine aminotransferase (ALT) and aspartate aminotransferase (AST) across 408,300 subjects from four ethnic groups in the UK Biobank, focusing on variants associating with both enzymes. Of these variants, the strongest effect is a rare (MAF in White British = 0.12%) missense variant in the gene encoding manganese efflux transporter SLC30A10, Thr95Ile (rs188273166), associating with a 5.9% increase in ALT and a 4.2% increase in AST. Carriers have higher prevalence of all-cause liver disease (OR = 1.70; 95% CI = 1.24 to 2.34) and higher prevalence of extrahepatic bile duct cancer (OR = 23.8; 95% CI = 9.1 to 62.1) compared to non-carriers. Over 4% of the cases of extrahepatic cholangiocarcinoma in the UK Biobank carry *SLC30A10* Thr95Ile. Unlike variants in *SLC30A10* known to cause the recessive syndrome hypermanganesemia with dystonia-1 (HMNDYT1), the Thr95Ile variant has a detectable effect even in the heterozygous state. Also unlike HMNDYT1-causing variants, Thr95Ile results in a protein that is properly trafficked to the plasma membrane when expressed in HeLa cells. These results suggest that coding variation in *SLC30A10* impacts liver health in more individuals than the small population of HMNDYT1 patients.

Introduction

Liver disease remains an area of high unmet medical need, and better characterizing the environmental and genetic determinants of liver disease is key to developing new therapeutic strategies. In addition, liver injury is a common side effect of drugs, and is a frequent reason that drugs fail to progress through the development pipeline; understanding the molecular mechanisms of liver injury can aid in rational drug design to avoid off-target effects. Circulating liver enzymes

are sensitive biomarkers of liver injury; in particular, alanine aminotransferase (ALT) and aspartate aminotransferase (AST) are released into the circulation during damage to hepatocyte membranes[1,2]. One powerful approach for understanding the molecular basis of liver disease has been to perform genome-wide association studies (GWAS) of levels of circulating liver enzymes across large population samples[1,3-13]. Combined GWAS of ALT and AST have previously revealed genetic associations providing potential therapeutic targets for liver disease such as *PNPLA3*[14] and *HSD17B13*[15]. To further study the genetics of hepatocellular damage, we performed GWAS on circulating levels of ALT and AST in 408,300 subjects, meta-analyzed across four ethnic groups in the UK Biobank.

Results

GWAS summary

We performed a GWAS of ALT and AST in four sub-populations in the UK Biobank (Asian or Asian British, Black or Black British, Chinese, and Black or Black British; sample sizes, number of variants tested, and λ_{GC} values, **Supplementary Table 1;** genome-wide significant associations, **Supplementary Table 2**; Manhattan and QQ plots for each enzyme and sub-population, **Supplementary Figures 1 and 2**). After meta-analyzing across sub-populations to obtain a single set of genome-wide p-values for each enzyme (Manhattan plots, **Figure 1**), we found 244 and 277 independent loci associating at $p < 5 \times 10^{-8}$ with ALT and AST, respectively, defined by lead SNPs separated by at least 500 kilobases and pairwise linkage disequilibrium (LD) r^2 less than 0.2. Enzyme levels were strongly associated with coding variants in the genes encoding the enzymes, representing strong protein quantitative trait loci in *cis* (cis-pQTLs). For example rs147998249, a missense variant Val452Leu in *GPT* (glutamic-pyruvic transaminase) encoding ALT, strongly

associates with ALT ($p < 10^{-300}$) and rs11076256, a missense variant Gly188Ser in *GOT2* (glutamic-oxaloacetic transaminase 2) encoding the mitochondrial isoform of AST, strongly associates with AST ($p = 6.3 \times 10^{-62}$). While these strong *cis*-pQTL effects validated our ability to detect direct genetic influences on ALT and AST levels, the aim of this study was to detect genetic determinants of liver health that in turn had downstream effects on both ALT and AST due to hepatocellular damage; therefore we focused the remainder of our analyses on the variants only associated with levels of both enzymes (labeled with black text on **Figure 1**).

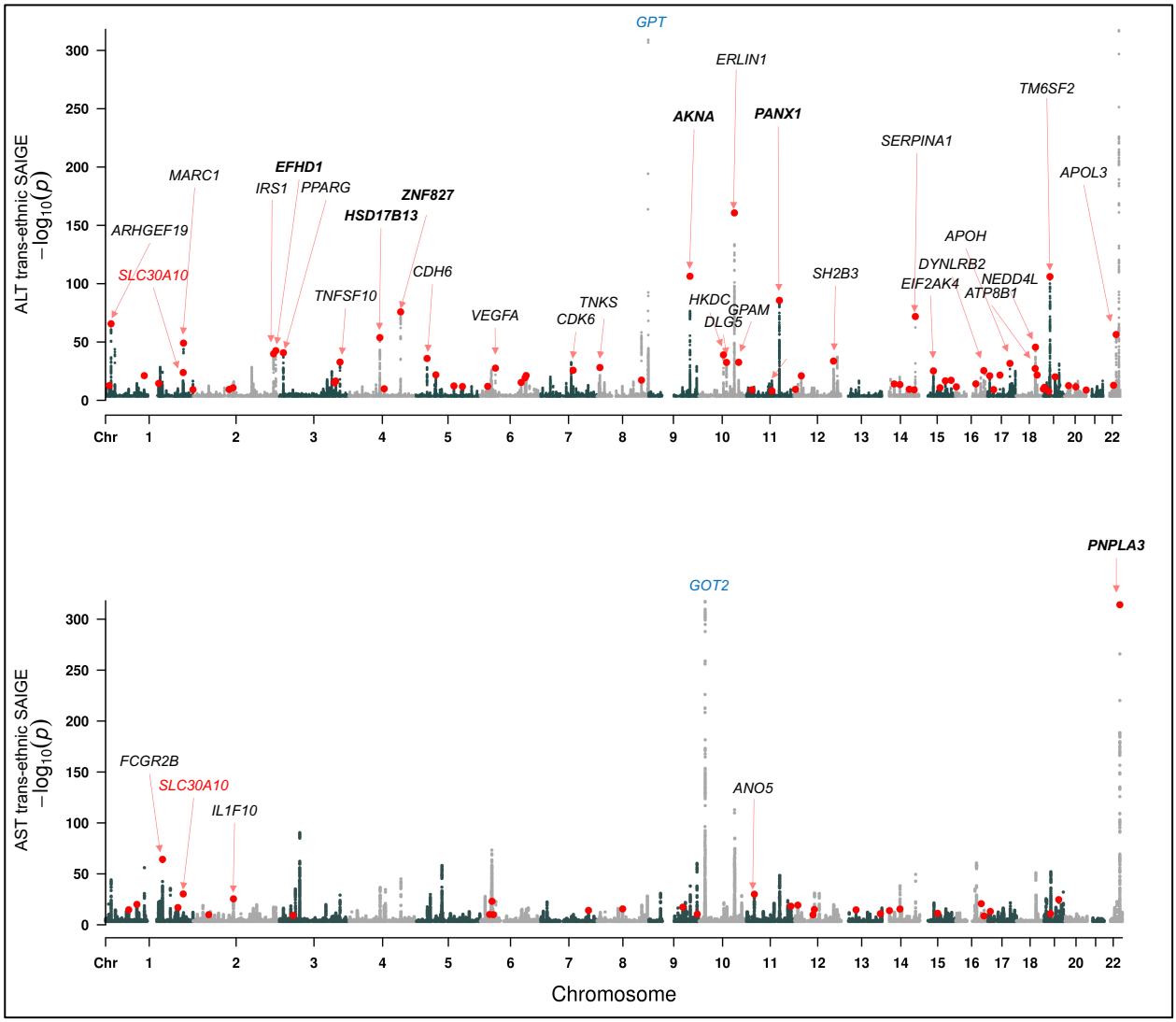


Figure 1: Manhattan plots showing trans-ethnic GWAS results for ALT and AST. Red dots indicate lead SNPs for shared signals between the two GWAS; for clarity, the shared signals are marked only once, on the plot for the GWAS in which the more significant association is detected. Cis-pQTLs (at *GPT* and *GOT2*) are labeled in blue. Loci with shared signals are labeled (for clarity, only when $p < 10\text{-}25$ and only on the GWAS for which the association is most significant). Loci previously reported to associate with both ALT and AST are named in bold. *SLC30A10*, the main topic of this report, is labeled in red on both plots.

Focusing only on loci with both ALT and AST GWAS signals (lead SNPs from either GWAS were identical or shared proxies with $r^2 \geq 0.8$), we found a total of 100 independent loci associated with both enzymes (**Figure 2, Supplementary Table 3**). As expected, effect sizes on ALT and AST at these loci were highly correlated ($r = 0.98$), and at all 100 loci the direction of effect on ALT and AST was concordant. Of these 100 loci, six were coincident or in strong LD with a published ALT or AST SNP in the EBI-NHGRI GWAS Catalog, and 15 were within 500kb of a published ALT or AST SNP; 33 of the loci harbored a moderate or high-impact variant as defined by the VEP or LOFTEE algorithms; and of the remaining 67 entirely noncoding loci, 19 were coincident or in strong LD with the strongest eQTL for a gene in liver, muscle, or kidney suggesting that effects on gene expression may drive their associations with ALT and AST. A majority (70 of the 100 loci) were shared with a distinct published association in the GWAS Catalog, suggesting pleiotropy with other traits.

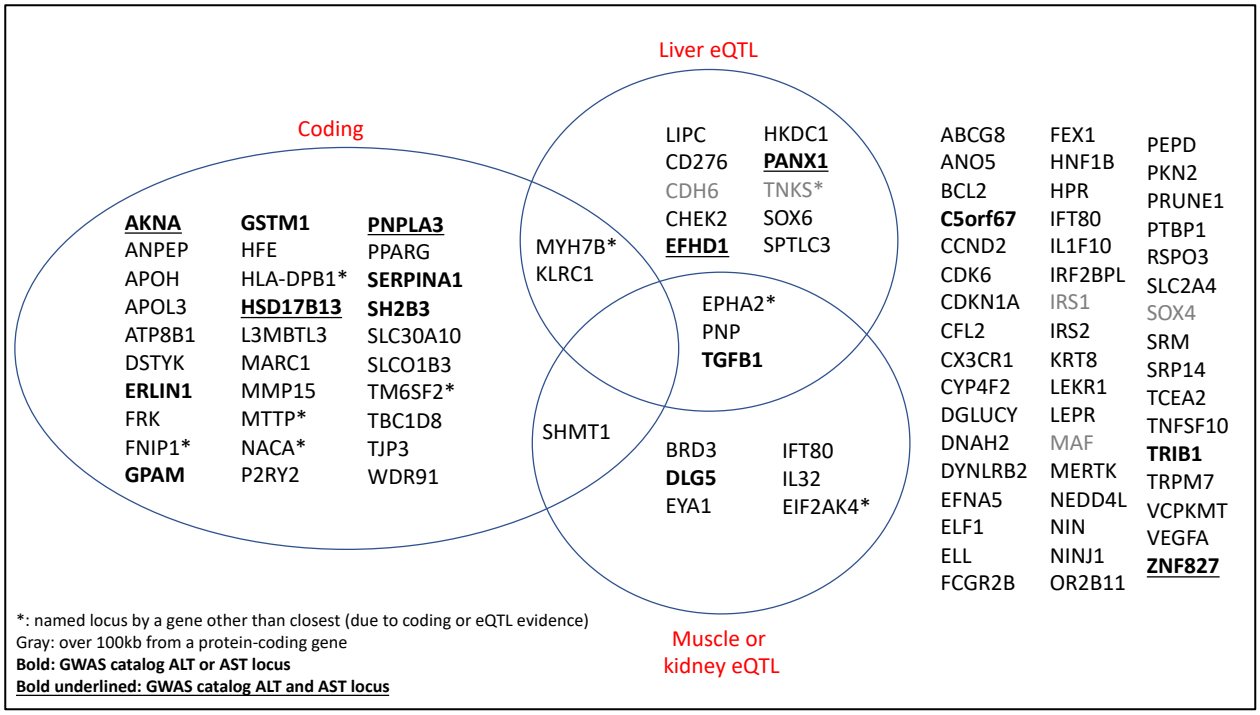


Figure 2: Classification of the top ALT- and AST- associated loci based on annotations. Unless otherwise noted with an asterisk, loci are named by the closest protein-coding gene. "Coding" indicates that one of the variants linked to the lead SNP is predicted to have a moderate or high impact on a protein-coding gene. "Liver eQTL" and "Muscle or kidney eQTL" indicate that one of the variants linked to the lead SNP is the strongest eQTL for a gene in those tissues by GTEx.

The strongest estimated effect of the lead SNPs on either enzyme was a novel association: rs188273166, a rare (MAF in White British = 0.12%) missense variant (Thr95Ile) in *SLC30A10*, associated with a 4.2% increase in ALT (95% CI, 4.6% to 7.1%; $p = 1.6 \times 10^{-24}$) and 5.9% increase in AST (95% CI: 3.4% to 5.0%; $p = 4.9 \times 10^{-31}$). Because Thr95Ile is coding and not strongly linked to any other SNPs, we considered it likely to be the causal variant driving the association at the *SLC30A10* locus. *SLC30A10* encodes a manganese efflux transporter (solute carrier family 30 member 10, also known as zinc transporter 10 or ZnT10)[16,17]. Loss-of-function mutations in *SLC30A10* have been reported to cause a rare recessive syndrome, hypermanganesemia with

dystonia 1 (HMNDYT1), characterized by cirrhosis, dystonia, parkinsonism, polycythemia, and hypermanganesemia[17,19-24]. The next strongest effect on either enzyme was rs28929474, a missense variant (the Pi-Z allele) in *SERPINA1* (serpin family A member 1) which causes alpha-1 antitrypsin deficiency (AATD) in its homozygous state[18], associated with a 2.5% increase in ALT (95% CI, 2.1% to 2.8%; $p = 1.4 \times 10^{-72}$) and 1.3% increase in AST (95% CI, 1.1% to 1.5%; $p = 3.2 \times 10^{-50}$); AATD manifests with both lung and liver damage. The most statistically significant association with either enzyme was rs738409, a common missense variant (Ile148Met) in *PNPLA3* (patatin like phospholipase domain containing 3) known to strongly increase risk of liver disease[14]; it is associated with a 2.2% increase in ALT (95% CI, 2.1% to 2.3%; $p < 10^{-300}$) and a 1.3% increase in AST (95% CI, 1.3% to 1.4%; $p < 10^{-300}$).

Association of ALT- and AST-associated variants with liver disease risk

We tested the 100 lead SNPs from the ALT and AST GWAS analysis for association with a combined liver disease phenotype (ICD10 codes K70-77, I85, and C22; 14,648 cases and 487,968 controls across the entire biobank), meta-analyzing liver disease association results across all four sub-populations (**Supplementary Table 3)**. Of the 100 lead SNPs, 26 SNPs associate with liver disease with $p < 0.05$. As expected, variants associated with an increase in ALT and AST tend to be associated with a proportional increase in liver disease risk (across all lead SNPs, effect size $r = 0.73$ for ALT, $r = 0.71$ for AST; **Figure 3**). Consistent with *SLC30A10* Thr95Ile (rs188273166) having the strongest effect on ALT and AST of all lead SNPs, it also has the strongest estimated effect on liver disease risk (OR = 1.70; 95% CI, 1.24 to 2.34; $p = 7.4 \times 10^{-3}$).

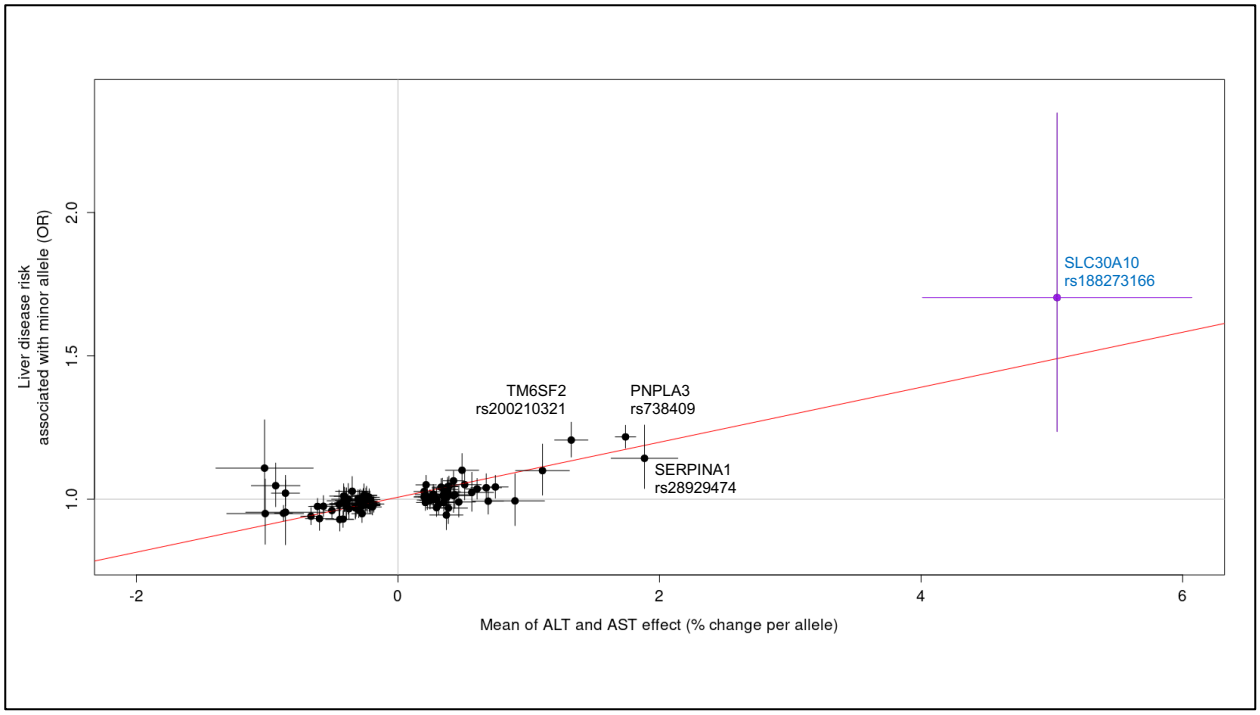


Figure 3: Comparison of effect of lead SNPs on ALT and AST with effect on liver disease. Effect sizes (beta in units of $\log_{10}$ ALT and AST and odds ratios) are from PLINK analysis; error bars are 95% confidence intervals on these effect sizes.

Because *SLC30A10* Thr95Ile had the strongest effect of all of our lead SNPs and also has not been reported as being associated with any phenotypes in the literature, we centered the following analyses on better understanding its function.

Validation of *SLC30A10* Thr95Ile genotype and replication of ALT and AST associations

Because rare variants are especially prone to errors in array genotyping[25], we sought to validate the array genotype calls for *SLC30A10* Thr95Ile in a subset of 301,473 individuals who had also been exome sequenced (**Supplementary Table 4**). The one individual homozygous for the minor (alternate) allele by array was confirmed by exome sequencing; no further homozygotes were

identified. Of 702 individuals called as heterozygous for Thr95Ile by array data who had available exome data, 699 (99.6%) were confirmed heterozygous by exome sequencing.

To replicate the rs188273166 association with ALT and AST in an independent cohort, we identified ethnic groups besides the White British subpopulation harboring the variant. The only two other populations with a substantial number of *SLC30A10* Thr95Ile carriers were individuals identifying as Other White (33/16,299 are carriers) and as White Irish (35/12,715 are carriers) (**Supplementary Table 5**). Meta-analyzing the association results across these two outgroups revealed higher mean ALT and AST in the small sample of additional Thr95Ile carriers, although the effect was not statistically significant for ALT (ALT increase = 3.7%; 95% CI, -1.1% to 8.4%, p = 0.084; AST increase = 4.1%; 95% CI, 1.1% to 7.2%; $p = 1.7 \times 10^{-3}$) (**Supplementary Figure 3**).

Magnitude of ALT and AST elevation in *SLC30A10* Thr95Ile carriers

After establishing the association between *SLC30A10* Thr95Ile and ALT and AST, we sought to further explore the relationship between genotype and enzyme levels to understand clinical relevance. Carriers of Thr95Ile had a mean ALT of 27.37 U/L vs 23.54 U/L for noncarriers, and a mean AST of 28.85 U/L vs 26.22 U/L for noncarriers. Counting individuals with both ALT and AST elevated above 40 U/L, a commonly-used value for the upper limit of normal (ULN)[2], 5.6% of carriers vs 3.6% of noncarriers had both enzymes elevated at the time of their UK Biobank sample collection, a relative increased risk of 58% (**Supplementary Table 6**).

Independence of *SLC30A10* Thr95Ile from neighboring ALT and AST associations

Because we applied distance and LD pruning to the results of the genome-wide scan, it was unclear how many independent association signals existed at the *SLC30A10* locus. Revisiting trans-ethnic association results in a window including 1 Mb flanking sequence upstream and downstream of *SLC30A10* revealed 76 SNPs with genome-wide significant associations with both ALT and AST (**Figure 4**). These 76 SNPs clustered into three loci: *SLC30A10* (only Thr95Ile, rs188273166); *MARC1* (mitochondrial amidoxime reducing component 1, lead SNP rs2642438 encoding missense Ala165Thr, reported in a recent preprint[26], and six additional SNPs together spanning 68 kilobases); and *LYPLAL1-ZC3H11B* (intergenic region between lyophospholipase like 1 and zinc finger CCCH-type containing 11B, with array-genotyped SNP rs6541227 and 67 imputed SNPs spanning 46 kilobases), a locus previously reported to associate with non-alcoholic fatty liver disease (NAFLD)[27].

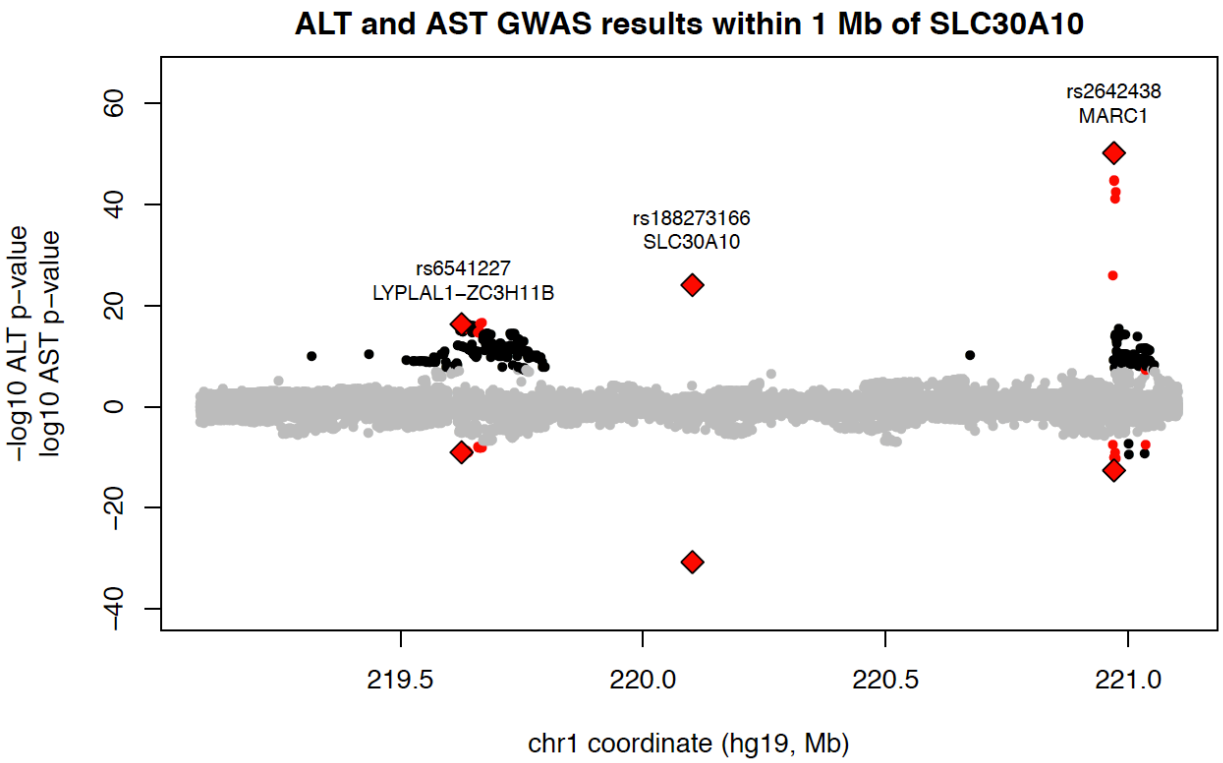


Figure 4: Miami plot of trans-ethnic SAIGE GWAS results within 1 Mb of the gene body of *SLC30A10*. ALT associations are shown in the positive direction and AST associations in the negative direction. SNP associations reaching genome-wide significance for one enzyme are colored black; for both enzymes, colored red; directly-genotyped SNPs significant for both enzymes, red diamonds.

To test for independence between these three loci, we performed ALT and AST association tests for each of the three array-typed SNPs while including the genotype of either one or both of the others as covariates. Associations were similar in these conditional analyses, suggesting that each of these three associations are not confounded by linkage disequilibrium with the other regional association signals (**Supplementary Table 7**.) Therefore *SLC30A10* Thr195Ile has a novel independent association with ALT and AST levels.

Linkage of Thr95Ile to GWAS SNPs at *SLC30A10*

A GWAS of circulating toxic metals[28] discovered an association between a common intronic SNP in *SLC30A10* (rs1776029; MAF in White British, 19.5%) and blood manganese levels, where the reference allele – which is the minor allele – is associated with increased circulating manganese. We calculated linkage disequilibrium statistics between rs1776029 and Thr95Ile and found that the minor allele of Thr95Ile (A) was in almost perfect linkage with the minor allele of rs1776029 (A) (r2 = 0.005, D' = 0.98); Thr95Ile (rs188273166) is 154 times more frequent among carriers of at least one copy of the minor allele of common SNP rs1776029 (95% CI = 84 – 325; Fisher's $p < 2.2 \times 10^{-16}$). These results suggest that the previously reported association of rs1776029 with circulating manganese may be partially or completely explained by linkage with Thr95Ile (**Supplementary Table 8**); however genotypes of Thr95Ile in the manganese GWAS or manganese measurements in the UK Biobank would be needed in order to perform conditional analysis. We then systematically tested nearby SNPs reported in the GWAS Catalog for any phenotype for linkage to Thr95Ile, measured by high |D'|. Combining GWAS Catalog information and |D'| calculations, we find nearly perfect linkage (|D'| > 0.90) between rs188273166-A (rare missense Thr95Ile) with rs1776029-A (intronic), rs2275707-C (3'UTR), and rs884127-G (intronic), all within the gene body of *SLC30A10* (**Supplementary Table 9**). In addition to increased blood Mn[28], these three common alleles have been associated with decreased magnesium/calcium ratio in urine[29], decreased mean corpuscular hemoglobin (MCH)[30-32], increased red blood cell distribution width[30-32], and increased heel bone mineral density (BMD)[32-35].

A recent study not yet in the GWAS catalog reported an association between another common intronic SNP in *SLC30A10* (rs759359281; MAF in White British, 5.6%) and liver MRI-derived cT1 measures, a proxy for liver fibrosis and steatohepatitis[36]. However, the reported cT1-increasing allele (liver disease risk allele) of rs759359281, which is the minor allele, is in complete linkage (D' = 1) with the major allele of Thr95Ile (rs188273166); in other words, the cT1-increasing allele and Thr95Ile liver disease risk allele occur on different haplotypes, suggesting that the mechanism of this reported cT1 association is independent of Thr95Ile.

Phenome-wide associations of *SLC30A10* Thr95Ile

To explore other phenotypes associated with *SLC30A10* Thr95Ile, we tested for association with 135 quantitative traits and 4,533 ICD10 diagnosis codes within the White British population (**Supplementary Tables 10 and 11**). Besides ALT and AST elevation, rs188273166 was associated with other indicators of hepatobiliary damage such as decreased HDL cholesterol and apolipoprotein A (ApoA)[37], decreased albumin, and increased GGT. Other phenome-wide significant quantitative trait associations were increases in hemoglobin concentration and hematocrit (**Table 1**).

Quantitative trait	Association with *SLC30A10* Thr95Ile					
	P	signif.	Effect (SD)	95% CI lower	95% CI upper	N
Aspartate aminotransferase	2.90E-32	***	0.38	0.31	0.44	387770
Alanine aminotransferase	1.95E-25	***	0.32	0.25	0.38	389063
HDL cholesterol	1.18E-15	***	-0.22	-0.28	-0.15	356214
Apolipoprotein A	7.25E-11	***	-0.19	-0.25	-0.12	354247
Albumin	2.12E-08	***	-0.17	-0.24	-0.10	356374
Hemoglobin concentration	8.11E-06	**	0.09	0.04	0.15	396065
Hematocrit (percentage)	1.33E-05	**	0.10	0.04	0.15	396065
Gamma glutamyltransferase	9.86E-05	**	0.12	0.05	0.18	388997
Heel BMD	2.05E-03	*	0.13	0.05	0.22	237053

MCHC	7.73E-02	N.S.	0.05	-0.01	0.12	327167
Erythrocyte dist. width	3.46E-01	N.S.	0.02	-0.04	0.09	327168

Table 1: Association results of *SLC30A10* Thr95Ile with selected quantitative traits. P values are from SAIGE analysis; effect size estimates are from PLINK analysis. (***) indicates genome-wide significance $p < 5 \times 10^{-8}$; (**) indicates phenome-wide significance of $p < 3.7 \times 10^{-4}$; (*) indicates nominal significance of $p < 0.05$. All traits are rank-based inverse-normal transformed and effect size is in units of standard deviations of the transformed values.

The only phenome-wide significant association of a diagnosis with *SLC30A10* Thr95Ile was C24.0, extrahepatic bile duct carcinoma (OR = 23.8; 95% CI = 9.1 to 62.1; $p = 1.2 \times 10^{-6}$). The overall prevalence of extrahepatic bile duct in the UK Biobank is 119 in 502,616 individuals (1 in 4,224) while the prevalence among carriers is 5 in 1,171 (1 in 234); 4% of individuals with extrahepatic bile duct carcinoma carry Thr95Ile. As there is an overlap of extrahepatic bile duct carcinoma cases with intrahepatic bile duct carcinoma cases, which are included in the all-cause liver disease phenotype, we examined whether the association of Thr95Ile with all-cause liver disease (noted earlier) was driven solely by individuals with bile duct carcinoma. However, we found in this phenome-wide scan that the ICD10 code block K70-K77, encompassing non-cancer liver disease, was also associated with *SLC30A10* Thr95Ile (OR = 1.6; 95% CI = 1.2 to 2.3; p = 0.048). This suggests that the presence of Thr95Ile affects both bile duct carcinoma and liver disease risk. Searching for neurological manifestations similar to HMNDYT1, we find no association with Parkinson's disease or dystonia but note that we are powered to exclude only strong effects because of the small case number for these traits; our confidence interval for association with Parkinson's and dystonia both include an effect of the same magnitude as on liver disease (**Table 2**).

	Association with *SLC30A10* Thr95Ile							
Diagnosis	P	sig.	N cases (biobank)	Prevalence (biobank)	N cases (carriers)	Prevalence (carriers)	OR	95% CI

C24.0: Malignant neoplasm of extrahepatic bile duct	1.17E-06	**	119	0.02%	5	0.45%	23.8	9.1	62.1
K70-K77: Diseases of liver	4.78E-02	*	13884	2.76%	40	3.58%	1.6	1.2	2.3
D50: Iron deficiency anemia	1.35E-02	*	22371	4.45%	62	5.55%	1.4	1.0	1.8
G20-G26: Extrapyramidal and movement disorders	7.64E-01	N.S.	10160	2.02%	19	1.70%	0.8	0.5	1.3
G20: Parkinson's disease	3.95E-01	N.S.	2426	0.48%	4	0.36%	0.5	0.1	1.9
G21: Secondary parkinsonism	2.77E-01	N.S.	812	0.16%	0	NA	1.6	0.3	10.6
G24: Dystonia	5.69E-01	N.S.	2354	0.47%	4	0.36%	1.2	0.5	3.0

Table 2: association results of *SLC30A10* Thr95Ile with selected ICD10 diagnosis codes. P values are from SAIGE analysis; effect size estimates are from PLINK analysis (exponentiated betas from logistic regression). (**) indicates phenome-wide significance of $p < 1.1 \times 10^{-5}$; (*) indicates nominal significance of $p < 0.05$.

To test whether the association with extrahepatic bile duct cancer was driven by a nearby association, we performed an association study of the diagnosis with variants within a window including 1 Mb flanking sequence upstream and downstream of *SLC30A10*; Thr95Ile was the strongest association (**Supplementary Figure 4**).

Bioinformatic characterization of *SLC30A10* Thr95Ile

To understand potential functional mechanisms of the Thr95Ile variant, we examined bioinformatic annotations of SLC30A10 Thr85Ile. The UNIPROT database shows that Thr95Ile occurs in the third of six transmembrane domains and shares a domain with a variant known to cause HMNDYT1 (**Figure 5**). In silico analysis predicts that Thr95Ile is a damaging mutation; it has a CADD PHRED score of 23.9, placing it in the top 1% of deleteriousness scores for genome wide potential variants; the algorithm SIFT predicts it as deleterious; and the algorithm PolyPhen-2 gives it a HumDiv score of 0.996 (probably damaging) and a HumVar score of 0.900 (possibly damaging.) Cross-species protein sequence alignment in PolyPhen-2 shows only threonine or

serine at position 95 across animals. These properties suggest that Thr95Ile substitution ought to have an effect the function of the SLC30A10 protein.

```
        10         20         30         40         50
MGRYSGKTCR LLFMLVLTVA FFVAELVSGY LGNSIALLSD SFNMLSDLIS
        60         70         80         90        100
LCVGLSAGYI ARRPTRGFSA TYGYARAEVV GALSNAVFLT ALCFTIFVEA
       110        120        130        140        150
VLRLARPERI DDPELVLIVG VLGLLVNVVG LLIFQDCAAW FACCLRGRSR
       160        170        180        190        200
RLQQRQQLAE GCVPGAFGGP QGAEDPRRAA DPTAPGSDSA VTLRGTSVER
       210        220        230        240        250
KREKGATVFA NVAGDSFNTQ NEPEDMMKKE KKSEALNIRG VLLHVMGDAL
       260        270        280        290        300
GSVVVVITAI IFYVLPLKSE DPCNWQCYID PSLTVLMVII ILSSAFPLIK
       310        320        330        340        350
ETAAILLQMV PKGVNMEELM SKLSAVPGIS SVHEVHIWEL VSGKIIATLH
       360        370        380        390        400
IKYPKDRGYQ DASTKIREIF HHAGIHNVTI QFENVDLKEP LEQKDLLLLC
       410        420        430        440        450
NSPCISKGCA KQLCCPPGAL PLAHVNGCAE HNGGPSLDTY GSDGLSRRDA
       460        470        480
REVAIEVSLD SCLSDHGQSL NKTQEDQCYV NRTHF
```

Figure 5: Protein sequence of SLC30A10 along with experimental evidence cited by UNIPROT. Underlined are the six transmembrane domains. Highlighted red is Thr95Ile. Highlighted in yellow are variants that demonstrated abolished Mn transport or membrane localization in vitro. Highlighted in blue are variants that demonstrated lesser or no effect on Mn transport or membrane localization in vitro. In bold are variants known to cause SLC30A10 deficiency (HMNDYT1).

Characterization of *SLC30A10* variants *in vitro*

To test the protein localization of SLC30A10 harboring Thr95Ile as well as other variants, we created constructs with Thr95Ile (rs188273166) and the HMNDYT1-causing variants Leu89Pro (rs281860284) and del105-107 and transfected these constructs into HeLa cells.

Immunofluorescence staining revealed membrane localization for wild-type (WT) SLC30A10 which was abolished by the two HMNDYT1 variants, consistent with previous reports which showed that the HMNDYT1 variants proteins are mislocalized in the endoplasmic reticulum (ER)[38]. In contrast, Thr95Ile showed membrane localization similar to WT, suggesting that Thr95Ile does not cause a deficit in protein trafficking to the membrane (**Figure 6**). Thr95Ile may affect SLC30A10 function in other ways, for example altering Mn transport efficiency.

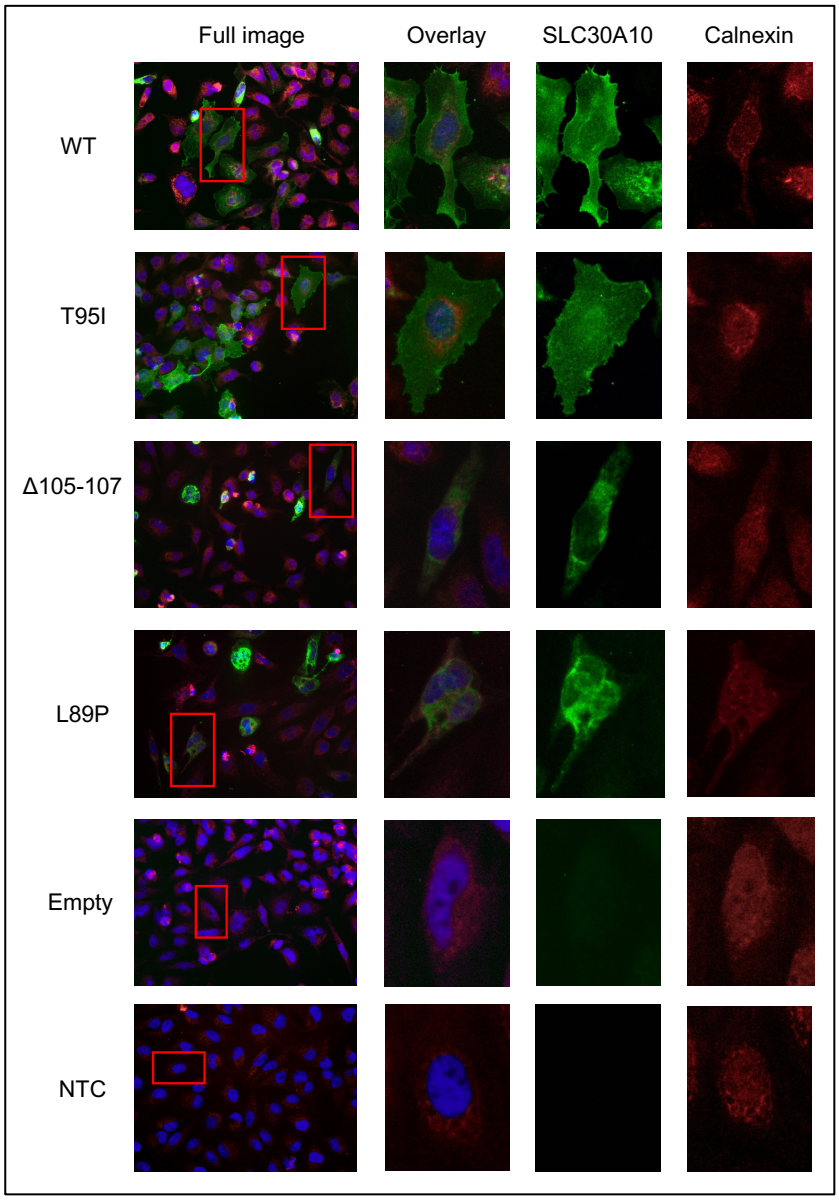


Figure 6: Immunofluorescence imaging of SLC30A10 protein constructs expressed in cultured HeLa cells. WT = wild type; T95I = Thr95Ile; Δ105-107 and L89P, HMNDYT1-causing variants reported previously [38]; empty = transfected with empty vector; NTC = non transfected control. Calnexin staining (in red) indicates the endoplasmic reticulum (ER).

Discussion

Expanded genetic landscape of risk for hepatocellular damage

Our trans-ethnic GWAS of ALT and AST reveals a broad genetic landscape of loci that modulate risk of hepatocellular damage or other diseases that cause increases in circulating ALT and AST, bringing the number of loci known to associate with levels for both enzymes from 10 (currently in the GWAS Catalog) to 100. Two loci had been previously reported in majority-European ancestry GWAS as associating with both ALT and AST: *PNPLA3*[3,4,7-9,39-41] and *HSD17B13*[3,7,15]; we detect colocalized ALT and AST signals at both of these loci. Broadening beyond majority-European ancestry GWAS, an additional eight loci have been previously identified: *PANX1*[7], *ALDH2*[6,8], *CYP2A6*[7,39], *ABCB11*[7], *ZNF827*[7], *EFHD1*[7], *AGER-NOTCH4*[7,42], and *AKNA*[7]; we replicate four of these in our trans-ethnic GWAS (*PANX1, ZNF827*, *EFHD1*, and *AKNA*.) We are limited by the lack of diversity in the UK Biobank and expect that studies in more diverse populations will result in the discovery of new loci and alleles.

Among the novel loci are many that had been previously identified as risk loci for liver disease, but had never been explicitly associated through ALT or AST GWAS, such as *SERPINA1* (associated with alpha-1 antitrypsin deficiency[18]), *HFE* (homeostatic iron regulator, associated with hemochromatosis[43]), and *TM6SF2* (transmembrane 6 superfamily member 2, associated with NAFLD[44-46]). Others are known to associate with biliary disease, such as *ABCG8* (ATP-binding

cassette subfamily G member 8, associated with gallstones[47]), *ANPEP* (alanyl aminopeptidase, associated with gallstones[48]), *HNF1B* (HNF1 homeobox B, associated with gallstones[48]), and *SLCO1B3* (solute carrier organic anion transporter family member 1B3, implicated in digenic rotor type hyperbilirubinemia[49]), along with loci previously associated with increased GGT (*EPHA2*, *CDH6*, *DLG5*, *CD276*, *DYNLRB2,* and *NEDD4L*)[3,7]. Consistent with the fact that ALT and AST elevation can be caused by kidney or muscle damage, we detect an association with *ANO5* (anoctamin 5), which has been implicated in several autosomal recessive muscular dystrophy syndromes[50,51], and several loci associated with expression of genes in muscle or kidney but not liver (*SHMT1, BRD3, DLG5, EYA1, IFT80, IL32*, and *EIF2AK4.*) We expect only a subset of the loci from this screen to be directly causally implicated in hepatocellular damage; many may predispose to a condition where liver damage is secondary or affect kidney or muscle, an important limitation of this approach.

Properties of *SLC30A10* Thr95Ile

The variant with the strongest predicted effect on ALT and AST, *SLC30A10* Thr95Ile (rs188273166), is a rare variant carried by 1,117 of the 487,327 array-genotyped participants in the UK Biobank. While Thr95Ile is found in some individuals of non-European ancestry, it is at much higher frequency in European-ancestry populations, with carrier frequency in our sample by UK country of birth ranging from a minimum of 1 in 479 people born in Wales to a maximum of 1 in 276 people born in Scotland (**Supplementary Table 5**). The increased frequency we see in European-ancestry populations is not merely due to those populations' overrepresentation in the UK Biobank, but is also consistent with global allele frequency data catalogued in dbSNP[52].

The Thr95Ile variant occurs in the third of six transmembrane domains of the SLC30A10 protein[53], the same domain affected by a previously-reported loss-of-function variant causing HMNDY1 (hypermanganesemia with dystonia 1), Leu89Pro (rs281860284)[38] (**Figure 6**). In vitro, Leu89Pro abolishes trafficking of SLC30A10 to the membrane[38], and another study pointed to a functional role of polar or charged residues in the transmembrane domains of SLC30A10 for manganese transport function[54]. Bioinformatic analysis suggest that Thr95Ile should impact protein function. Our site-directed mutagenesis experiment of SLC30A10 shows that Thr95Ile, unlike reported HMNDYT1-causing variants, results in a protein that is properly trafficked to the cell membrane. Further biochemical studies will be required to investigate whether the Thr95Ile variant of SLC30A10 has reduced manganese efflux activity, which would be consistent with our genetic observations.

Comparison of *SLC30A10* Thr95Ile phenotypes to HMNDYT1 phenotypes

SLC30A10 (also known as *ZNT10*, and initially identified through sequence homology to zinc transporters[16]) encodes a cation diffusion facilitator expressed in hepatocytes, the bile duct epithelium, enterocytes, and neurons[22] that is essential for excretion of manganese from the liver into the bile and intestine[17,22]. Homozygous loss-of-function of *SLC30A10* was recently identified as the cause of the rare disease HMNDYT1, which in addition to hypermanganesemia and dystonia is characterized by liver cirrhosis, polycythemia, and Mn deposition in the brain[19-24,38]. Other hallmarks include iron depletion and hyperbilirubinemia. Mendelian disorders of *SLC30A10* and the other hepatic Mn transporter genes *SLC39A8* (solute carrier family 39 member 8, causing congenital disorder of glycosylation type IIn)[55] and *SLC39A14* (solute carrier family 39 member 14, implicated in hypermanganesemia with dystonia 2)[56], along with experiments in transgenic

mice [57,58], have confirmed the critical role of each of these genes in maintaining whole-body manganese homeostasis[59]. Notably, while all three of the genes have Mendelian syndromes with neurological manifestations, only *SLC30A10* deficiency (HMNDYT1) is known to be associated with liver disease[59].

We detect two key aspects of HMNDYT1 – increased circulating liver enzymes and increased hematocrit – exceeding phenome-wide significance in heterozygous carriers of *SLC30A10* Thre95Ile (rs188273166). Among other hepatic phenotypes that have been reported in HMNDYT1 cases, we also detect an association with overall liver disease and anemia, but no evidence of hyperbilirubinemia. The neurological aspect of HMNDYT1, parkinsonism and dystonia, is not detectably enriched among Thr95Ile carriers; however, we have limited power and cannot exclude an enrichment. It is therefore intriguing to consider that carrier status of Thr95Ile may represent a very mild manifestation of HMNDYT1.

Comparison of Thr95Ile phenotypes to *SLC30A10* common variant phenotypes

Apart from rare variants in *SLC30A10* causing HMNDYT1, Thr95Ile can also be compared to common SNPs in *SLC30A10* that have been associated with phenotypes by GWAS. We find that the minor allele of Thr95Ile is in almost complete linkage with a common intronic SNP associated with increased blood manganese. Other GWAS SNPs in almost perfect linkage with Thr95Ile associate with decreased MCH, increased RBC distribution width, decreased magnesium/calcium ratio, and increased heel bone mineral density (BMD). Decreased MCH could reflect the anemia experienced by HMNDYT1 patients, caused by the closely linked homeostatic regulation of manganese and iron[17]. Increased BMD may reflect the protective role of manganese in bone

maintenance[60,61]. Looking for the subset of these phenotypes available in our scan of Thr95Ile, we do find a nominally significant increase in BMD and a suggestive increase in MCH but no detectable increase in erythrocyte distribution width. By contrast, we find that the a common intronic SNP in *SLC30A10* recently reported to associate with liver MRI cT1, a sign of steatohepatitis and fibrosis[36], is in complete linkage with the major allele of Thr95Ile, suggesting an independent genetic mechanism but also providing independent evidence of the role of *SLC30A10* SNPs in liver health.

The linked GWAS SNPs may be interpreted through two mechanistic hypotheses: first, the associations may all be causally driven by Thr95Ile carriers in the studies, which the GWAS SNPs tag; alternatively, the associations may be driven by effects of the common SNPs themselves, which are noncoding but may influence *SLC30A10* (or another gene in cis) by modulating expression or post-transcriptional regulation; or some combination of both. To distinguish between these, measurements of Mn would need to be available to perform conditional analyses. If the GWAS SNPs have an effect independent of Thr95Ile, *SLC30A10* seems likely (although not certain) to nevertheless be the causal gene at the locus, due to the similarity in phenotypes to HMNDYT1 and Thr95Ile. A putative regulatory mechanism could be through transcriptional or post-transcriptional regulatory elements, as the haplotype includes a SNP (rs2275707) overlapping both the 3'-UTR of *SLC30A10* and H3K4me1 histone modifications characteristic of enhancers active only in brain and liver[62].

Clinical relevance: manganese homeostasis in health and disease

Manganese (Mn) is a trace element required in the diet for normal development and function, serving as a cofactor and regulator for many enzymes. However, it can be toxic in high doses; because Mn(II) and Mn(III) can mimic other cations such as iron, it can interfere with systemic iron homeostasis and disrupt in other biochemical processes[63,64]; at the cellular level, it is cytotoxic and poisons the mitochondria by interfering with the electron transport chain enzymes that rely on Fe-S clusters[65]. The hallmark of occupational exposure through inhalation is neurotoxicity manifesting as parkinsonism and dystonia (manganism, or Mn intoxication)[63,64]. Neurotoxicity is an aspect of the Mendelian syndromes caused by loss of function of all three of the hepatic manganese transporters; interestingly, GWAS has also identified a common missense variant in *SLC39A8* as a risk factor for schizophrenia and many other diseases[66,67]; altered function of glycosyltransferases due to manganese overload in the neurons is a proposed mechanism for neurological manifestations of this variant[68]. Because manganese is excreted through the liver into the bile, increased circulating manganese secondary to liver damage may be a contributing factor to the neurological manifestations of chronic hepatic encephalopathy (CHE), known as chronic acquired hepatocerebral degeneration (CAHD)[69-71]. However, liver toxicity is not a hallmark of environmental or occupational exposure. Importantly, of the Mendelian syndromes of genes encoding manganese transporters, only *SLC30A10* (causing HMNDYT1) involves hepatic symptoms[59,72]. Hepatotoxicity in HMNDYT1 is thought to be due to cytotoxic manganese overload within hepatocytes; polycythemia is thought to be caused by upregulation of erythropoietin by manganese; and iron anemia through systemic dysregulation of iron homeostasis by excess manganese[72,73]. Our results indicate that polymorphism in *SLC30A10* is a risk factor for

manganese-induced hepatocellular damage, polycythemia, and iron anemia in a much broader population beyond the rare recessive syndrome HMNDYT1.

The association of *SLC30A10* Thr95Ile with extrahepatic bile duct cancer was unexpected, as it has not been described in conjunction with HMNDYT1. Bile duct cancer (cholangiocarcinoma) is a rare disease (age-adjusted incidence of 1 – 3 per 100,000 per year); cirrhosis, viral hepatitis, primary sclerosing cholangitis, and parasitic fluke infection have been identified as risk factors[74,75]. It is unclear whether low levels of manganese in the bile, or high levels of manganese in the hepatocytes and bile duct epithelial cholangiocytes, could be directly carcinogenic; manganese-dependent superoxide dismutase (MnSOD, or SOD2) is a tumor suppressor[76]. A simpler possibility is that cytotoxic manganese overload in hepatocytes and the bile duct epithelium causes localized inflammation that predisposes to cancer through similar mechanisms as other hepatobiliary risk factors. To our knowledge, *SLC30A10* Thr95Ile would be the strongest genetic cholangiocarcinoma risk factor identified to date, being carried by 4% of the extrahepatic bile duct cancer cases in our sample. Because both *SLC30A10* Thr95Ile and extrahepatic bile duct cancer are exceedingly rare, validation of this association in either another very large biobank or in a cohort of cholangiocarcinoma patients will be necessary.

Clinical relevance: genome interpretation

Currently, *SLC30A10* Thr95Ile (rs188273166) is listed as a variant of uncertain significance in the ClinVar database[77]. Individuals carrying *SLC30A10* Thr95Ile may benefit from increased surveillance of liver function and blood manganese levels and subsequent treatment if these conditions arise, because evidence from HMNDYT1 patients has demonstrated that chelation

therapy combined with iron supplementation is effective at reversing the symptoms of SLC30A10 insufficiency[78]. Further studies will be needed to define whether other damaging missense variants or protein-truncating variants in *SLC30A10*, including the variants known to cause HMNDYT1, also predispose to liver disease in their heterozygous state. Because we only observe one homozygous carrier of *SLC30A10* Thr95Ile in our data (who has no diagnosed liver or neurological disease, and unfortunately does not have ALT or AST measurements), further study will also be needed to understand the inheritance model of this association; we cannot determine whether risk in homozygotes is stronger than risk in heterozygotes, unlike cases of HMNDYT1 where identified cases have all been homozygous loss-of-function.

More broadly, the case of *SLC30A10* fits a pattern of discoveries of recent discoveries showing that recessive Mendelian disease symptoms can manifest in heterozygous carriers of deleterious variants, blurring the distinction between recessive and dominant disease genes and bridging the gap between common and rare disease genetics[79,80]. These discoveries are possible only by combining massive, biobank-scale genotype and phenotype datasets such as the UK Biobank.

Methods

Sub-population definition and PC calculation

Sub-populations for analysis were obtained through a combination of self-reported ethnicity and genetic principal components. First, the White British population was defined using the categorization performed previously by the UK Biobank[81] (Field 22006 value "Caucasian"); briefly, this analysis selected the individuals who identify as White British (Field 21000), performed a series of subject-level QC steps (to remove subjects with high heterozygosity or missing rate over 5%, removing subjects with genetic and self-reported sex discrepancies and

putative sex chromosome aneuploidies, and removing subjects with second or first degree relatives and an excess of third degree relatives), performed Bayesian outlier detection using the R package aberrant [82] to remove ancestry outliers using principal components (PCs) 1+2, 3+4, and 5+6 (calculated from the global UK Biobank PCs stored in Field 22009), selected a subset of SNPs in preparation for PCA by limiting to directly-genotyped (info = 1), missingness across individuals < 2%, MAF > 1%, regions of known long range LD, and pruning to independent markers with pairwise LD < 0.1. Based on this procedure used by the UK Biobank to define the "White British" subset, we defined three additional populations, using other self-reported ancestry groups as starting points (Field 21000 values "Asian or Asian British", "Black or Black British", and "Chinese"). Principal components were estimated in PLINK using the unrelated subjects in each subgroup. We then projected all subjects onto the PCs. For calculation of per-variant allele frequency and missingness thresholds, calculation of LD, and for association analyses performed in PLINK, the unrelated sets were used. For association analyses performed in SAIGE, the related sets were used.

For validation in an independent population, two other self-reported ethnicity groups with a sufficient number of *SLC30A10* Thr95Ile carriers were assembled, who were not included in "White British" (Field 21000 values "White" subgroup "Irish", and "White" subgroup "Any other white background" or no reported subgroup).

Array genotype data for association analysis

Data were obtained from the UK Biobank through application 26041. Genotypes were obtained through array typing and imputation as described previously[81]. For genome-wide association

analysis, variants were filtered so that imputation quality score (INFO) was greater than 0.8. Genotype missingness, Hardy-Weinberg equilibrium (HWE), and minor allele frequency (MAF) were then each calculated across the unrelated subset of individuals in each of the four sub-populations. For each sub-population a set of variants for GWAS was then defined by filtering missingness across individuals less than 2%, HWE p-value > 10^{-12}, and MAF > 0.1%.

Phenotype data

For genome-wide analysis, blood biochemistry values were obtained for ALT (Field 30620) and AST (Field 30650) and $\log_{10}$ transformed, consistent with previous genetic studies[3,83], resulting in an approximately normal distribution.

For phenome-wide analysis, ICD10 codes were obtained from inpatient hospital diagnoses (Field 41270), causes of death (Field 40001 and 40002), the cancer registry (Field 40006), and GP clinical event records (Field 42040). A selection of 135 quantitative traits was obtained from other fields, encompassing anthropomorphic measurements, blood and urine biochemistry, smoking, exercise, sleep behavior, and liver MRI; all were inverse rank normalized using the RNOmni R package[84].

Genome-wide association studies of ALT and AST

Because of the high level of relatedness in the UK Biobank participants[85], to maximize power by retaining related individuals we used SAIGE software package[86] to perform generalized mixed model analysis for GWAS. A genetic relatedness matrix (GRM) was calculated for each sub-population with a set of 100,000 LD-pruned variants selected from across the allele frequency spectrum. SAIGE was run on the filtered imputed variant set in each sub-population using the

following covariates: age at recruitment, sex, BMI, and the first 12 principal components of genetic ancestry (learned within each sub-population as described above). Manhattan plots and Q-Q plots were created using the qqman R package[87]. The association results for each enzyme were meta-analyzed across the four populations using the METAL software package[88] using the default approach (using p-value and direction of effect weighted according to sample size) To report p-value results, the default approach was used. To report effect sizes and standard errors, because the authors of the SAIGE method advise that parameter estimation may be poor especially for rare variants[89], the PLINK software package v1.90[90] was run on lead variants on the unrelated subsets of each subpopulation, and then the classical approach (using effect size estimates and standard errors) was used in METAL to meta-analyze the resulting betas and standard errors.

Identifying independent, colocalized association signals between the two GWAS

Meta-analysis results for each enzyme were LD clumped using the PLINK software package, v1.90[90] with an r^2 threshold of 0.2 and a distance limit of 10 megabases, to group the results into approximately independent signals. LD calculations were made using genotypes of the White British sub-population because of their predominance in the overall sample. Lead SNPs (the SNPs with the most significant p-values) from these "$r^2 > 0.2$ LD blocks" were then searched for proxies using an r^2 threshold of 0.8 and a distance limit of 250 kilobases, resulting in "$r^2 > 0.8$ LD blocks" defining potentially causal SNPs at each locus. The "$r^2 > 0.8$ LD blocks" for the ALT results were then compared to the "$r^2 > 0.8$ LD blocks" for the AST results, and any cases where these blocks shared at least one SNP between the two GWAS were treated as potentially colocalized association signals between the two GWAS. In these cases, a representative index SNP was chosen to represent the results of both GWAS by choosing the index SNP of the GWAS with the more significant p-

value. Next, these putative colocalized association signals were then distance pruned by iteratively removing neighboring index SNPs within 500 kilobases of each index SNP with less significant p-values (the minimum p-value between the two GWAS was used for the distance pruning procedure.) The Manhattan plot of METAL results with labeled colocalization signals was created using the CMplot R package[91] .

Annotation of associated loci and variants

Index SNPs and their corresponding strongly-linked ($r^2 > 0.8$) variants were annotated using the following resources: distance to closest protein-coding genes as defined by ENSEMBL v98 using the BEDTools package[92], impact on protein-coding genes using the ENSEMBL Variant Effect Predictor software package[93]; eQTLs (only the most significant eQTL per gene-tissue combination) from GTEx v8 (obtained from the GTEx Portal) for liver, kidney cortex, and skeletal muscle[94]; a published meta-analysis of four liver eQTL studies[95]; the eQTLGen meta-analysis of blood eQTL studies[96]; and GWAS results from the NHGRI-EBI GWAS Catalog (r2020-01-27)[97], filtered to associations with $p < 5 \times 10^{-8}$.

Association ALT- and AST-associated loci with liver disease

Index SNPs were tested for association with any liver disease. A broad liver disease phenotype was constructed by combining all diagnoses mapping to the ICD10 codes K70-K77 (diseases of liver), I85 (esophageal varices), and C22 (malignant neoplasm of liver and intrahepatic bile ducts) in inpatient hospital diagnoses (Field 41270), causes of death (Field 40001 and 40002), cancer registry (Field 40006), and GP clinical event records (Field 42040). For each index SNP and sub-population, two association tests were performed using SAIGE: in the subset of individuals with

GP clinical event records available, and in the subset without GP clinical event records available. These pairs of association tests were then meta-analyzed in METAL using the default approach (using p-values) between the with-GP and without-GP subsets. Then, association results were meta-analyzed across the four sub-populations using the same method to obtain the final p-value. To obtain effect sizes and standard errors, the same procedure was performed but using PLINK (on the unrelated subset of each population) and using the classical method in METAL (using effects and standard errors.)

Sequencing-based validation of rs188273166 array genotyping

Whole exome sequencing was available for 301,473 of the 487,327 array-genotyped samples. DNA was extracted from whole blood and was prepared and sequenced by the Regeneron Genetics Center (RGC). A complete protocol has been described elsewhere[98]. Briefly, the xGen exome capture was used and reads were sequenced using the Illumina NovaSeq 6000 platform. Reads were aligned to the GRCh38 reference genome using BWA-mem[99]. Duplicate reads were identified and excluded using the Picard MarkDuplicates tool (Broad Institute)[100]. Variant calling of SNVs and indels was done using the WeCall variant caller (Genomics Plc.)[101] to produce a GVCF for each subject. GVCFs were combined to using the GLnexus joint calling tool[102]. Post-variant calling filtering was applied as described previously[98].

Replication of *SLC30A10* Thr95Ile associations

ALT and AST association tests were repeated as described for the genome-wide scans, using SAIGE and PLINK, in the “Other White” and “White Irish” populations, for the *SLC30A10*

Thr95Ile (rs188273166) variant. Results were meta-analyzed across the two populations. A forest plot was created using the forestplot package in R[103].

Testing linkage of *SLC30A10* Thr95Ile to common GWAS SNPs

To test linkage of *SLC30A10* Thr95Ile (rs188273166) to common GWAS SNPs, the GWAS Catalog was searched for all results where "Mapped Gene" was assigned to *SLC30A10*; because of the very relevant phenotype, blood Mn-associated SNP rs1776029, an association that is not in the GWAS Catalog, was also included in the analysis, as well at cT1-associated SNP rs759359281. LD calculations were performed in PLINK, using the White British unrelated subpopulation, between rs188273166 and the GWAS SNPs with the options --r2 dprime-signed in-phase with-freqs --ld-window 1000000 --ld-window-r2 0. For rs1776029, an additional Fisher's exact test was performed to determine the confidence interval of the enrichment of rs188273166 on the rs1776029 haplotype. The linked alleles from PLINK were then used in conjunction with the effect allele from the reported papers to determine the direction of effect. The GWAS Atlas website[32] was used (the PheWAS tool) to determine the direction of effect for the linked alleles from the original paper; in cases where the original paper from the GWAS Catalog did not report a direction of effect, other papers for the same phenotype and SNP from GWAS Atlas were used to determine the direction of effect and cited accordingly. Reference epigenome information for the GWAS SNPs was obtained by searching for rs1776029 in HaploReg v4.1[104].

Phenome-wide association study of *SLC30A10* Thr95Ile

A phenome-wide association study of *SLC30A10* Thr95Ile (rs188273166) was performed by running SAIGE and PLINK against a set of ICD10 diagnoses and quantitative traits, obtained as

described above, and meta-analyzed using METAL as described above for the test of association with liver disease. ICD10 diagnoses were filtered to include only those at a three-character (category), four-character (category plus one additional numeral), or "block" level that were frequent enough to test in both subpopulations and without significant collinearity with the sex covariate: at least 100 diagnoses overall, at least 10 diagnoses in both the with- and without-GP data subgroups, and at least 10 diagnoses in both men and women. This resulted in 4,533 ICD10 codes to test, serving as the multiple hypothesis burden. For 5 of these ICD10 code phenotypes, the first step of SAIGE (fitting the null logistic mixed model to estimate the variance component and other model parameters) failed before the association test; for these phenotypes, only PLINK was run.

Bioinformatic analysis of *SLC30A10* Thr95Ile

To visualize Thr95Ile on the protein sequence of SLC30A10, UNIPROT entry Q6XR72 (ZN10_HUMAN) was accessed[53]. In UNIPROT, natural variants causing HMNDYT1[22,24,38] and mutagenesis results[38,54,105] were collated from the literature and highlighted. CADD score v1.5[106] was downloaded from the authors' website. SIFT score was obtained from the authors' website using the "dbSNP rsIDs (SIFT4G predictions)" tool[107]. PolyPhen score and multiple species alignment was obtained from the authors' website using the PolyPhen-2 tool[108].

Immunofluorescence of SLC30A10 localization in cultured cells

HeLa cells (ATCC®, Manassas, VA) were grown in Eagle's Minimum Essential Medium (ATCC®, Manassas, VA) containing 10% fetal bovine serum (Gibco, Carlsbad, CA) at 37°C and 5% CO_2. All plasmid transfections were performed using Lipofectamine™ 2000 (Invitrogen,

Carlsbad, CA) and Opti-MEM (Gibco, Grand Island, NY) according to manufacturer's specifications. FLAG-tagged SLC30A10 plasmid constructs designed with a linker sequence in pCMV6-AN-3DDK (Blue Heron Biotech, Bothell, WA) included wild type, del105-107, L89P, T95I, and used an empty vector for one of the negative controls.

HeLa cells were grown on 8-chambered slides for 48 hours post-transfection. IF procedures were performed at room temperature unless otherwise noted. HeLa cells were rinsed in PBS, fixed with 4% paraformaldehyde (in water) (Electron Microscopy Sciences, Hatfield, PA) for 10 min, rinsed in 4°C PBS, and permeabilized for 5 minutes with 0.1% Triton X-100 (Sigma-Aldrich, St. Louis, MO). After rinsing in PBS and blocking in 2% BSA (in PBS) (Jackson ImmunoResearch, West Grove, PA) for 30 minutes, the cells were stained with 2% BSA blocking solution containing monoclonal ANTI-FLAG® M2-FITC, Clone M2 (dilution 1:100; Sigma-Aldrich, St. Louis, MO) and Calnexin Monoclonal Antibody, Clone AF18 (dilution 1:100; Invitrogen, Carlsbad, CA). After three final washes in PBS, mounting medium with DAPI (Vector Laboratories, Burlingame, CA) was added and sealed under a coverslip with nail polish. Images were captured with the REVOLVE Echo microscope at 20X magnification.

Acknowledgements

This research has been conducted using the UK Biobank resource, application number 26041. We thank the UK Biobank participants for their donations to this resource. We thank the Regeneron Genetics Center for performing exome sequencing and analysis.

Author Contributions

A.D., A.F.C., L.W., M.P., and P.N. performed computational analyses; C.Q. and H.C.T. performed experiments; L.W. wrote the manuscript; all authors interpreted results and edited the manuscript.

Competing Interests

The authors are all employees of Alnylam Pharmaceuticals, Inc.

References

1 van Beek, J. H. *et al.* The genetic architecture of liver enzyme levels: GGT, ALT and AST. *Behav Genet* **43**, 329-339, doi:10.1007/s10519-013-9593-y (2013).

2 Pratt, D. S. & Kaplan, M. M. Evaluation of abnormal liver-enzyme results in asymptomatic patients. *The New England journal of medicine* **342**, 1266-1271, doi:10.1056/NEJM200004273421707 (2000).

3 Chambers, J. C. *et al.* Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nature genetics* **43**, 1131-1138, doi:10.1038/ng.970 (2011).

4 Young, K. A. *et al.* Genome-Wide Association Study Identifies Loci for Liver Enzyme Concentrations in Mexican Americans: The GUARDIAN Consortium. *Obesity (Silver Spring)* **27**, 1331-1337, doi:10.1002/oby.22527 (2019).

5 Park, T. J. *et al.* Genome-wide association study of liver enzymes in korean children. *Genomics Inform* **11**, 149-154, doi:10.5808/GI.2013.11.3.149 (2013).

6 Moon, S. *et al.* The Korea Biobank Array: Design and Identification of Coding Variants Associated with Blood Biochemical Traits. *Sci Rep* **9**, 1382, doi:10.1038/s41598-018-37832-9 (2019).

7 Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nature genetics* **50**, 390-400, doi:10.1038/s41588-018-0047-6 (2018).
8 Kamatani, Y. *et al.* Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nature genetics* **42**, 210-215, doi:10.1038/ng.531 (2010).
9 Kim, Y. J. *et al.* Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. *Nature genetics* **43**, 990-995, doi:10.1038/ng.939 (2011).
10 Prins, B. P. *et al.* Genome-wide analysis of health-related biomarkers in the UK Household Longitudinal Study reveals novel associations. *Sci Rep* **7**, 11008, doi:10.1038/s41598-017-10812-1 (2017).
11 Namjou, B. *et al.* GWAS and enrichment analyses of non-alcoholic fatty liver disease identify new trait-associated genes and pathways across eMERGE Network. *BMC Med* **17**, 135, doi:10.1186/s12916-019-1364-z (2019).
12 Gurdasani, D. *et al.* Uganda Genome Resource Enables Insights into Population History and Genomic Discovery in Africa. *Cell* **179**, 984-1002 e1036, doi:10.1016/j.cell.2019.10.004 (2019).
13 Gilly, A. *et al.* Very low-depth whole-genome sequencing in complex trait association studies. *Bioinformatics* **35**, 2555-2561, doi:10.1093/bioinformatics/bty1032 (2019).
14 Romeo, S. *et al.* Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet* **40**, 1461-1465, doi:10.1038/ng.257 (2008).
15 Abul-Husn, N. S. *et al.* A Protein-Truncating HSD17B13 Variant and Protection from Chronic Liver Disease. *The New England journal of medicine* **378**, 1096-1106, doi:10.1056/NEJMoa1712191 (2018).
16 Seve, M., Chimienti, F., Devergnas, S. & Favier, A. In silico identification and expression of SLC30 family genes: an expressed sequence tag data mining strategy for the characterization of zinc transporters' tissue expression. *BMC Genomics* **5**, 32, doi:10.1186/1471-2164-5-32 (2004).
17 Tuschl, K. *et al.* Syndrome of hepatic cirrhosis, dystonia, polycythemia, and hypermanganesemia caused by mutations in SLC30A10, a manganese transporter in man. *American journal of human genetics* **90**, 457-466, doi:10.1016/j.ajhg.2012.01.018 (2012).
18 Brantly, M., Nukiwa, T. & Crystal, R. G. Molecular basis of alpha-1-antitrypsin deficiency. *Am J Med* **84**, 13-31, doi:10.1016/0002-9343(88)90154-4 (1988).
19 Brna, P., Gordon, K., Dooley, J. M. & Price, V. Manganese toxicity in a child with iron deficiency and polycythemia. *J Child Neurol* **26**, 891-894, doi:10.1177/0883073810393962 (2011).
20 Gospe, S. M., Jr. *et al.* Paraparesis, hypermanganesaemia, and polycythaemia: a novel presentation of cirrhosis. *Arch Dis Child* **83**, 439-442, doi:10.1136/adc.83.5.439 (2000).
21 Lechpammer, M. *et al.* Pathology of inherited manganese transporter deficiency. *Ann Neurol* **75**, 608-612, doi:10.1002/ana.24131 (2014).
22 Quadri, M. *et al.* Mutations in SLC30A10 cause parkinsonism and dystonia with hypermanganesemia, polycythemia, and chronic liver disease. *American journal of human genetics* **90**, 467-477, doi:10.1016/j.ajhg.2012.01.017 (2012).

23 Sahni, V. *et al.* Case report: a metabolic disorder presenting as pediatric manganism. *Environ Health Perspect* **115**, 1776-1779, doi:10.1289/ehp.10421 (2007).
24 Tuschl, K. *et al.* Hepatic cirrhosis, dystonia, polycythaemia and hypermanganesaemia--a new metabolic disorder. *J Inherit Metab Dis* **31**, 151-163, doi:10.1007/s10545-008-0813-1 (2008).
25 Weedon, M. N. *et al.* Assessing the analytical validity of SNP-chips for detecting very rare pathogenic variants: implications for direct-to-consumer genetic testing. *bioRxiv*, 696799, doi:10.1101/696799 (2019).
26 Emdin, C. A. *et al.* A missense variant in Mitochondrial Amidoxime Reducing Component 1 gene and protection against liver disease. *bioRxiv*, 594523, doi:10.1101/594523 (2019).
27 Speliotes, E. K. *et al.* Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. *PLoS Genet* **7**, e1001324, doi:10.1371/journal.pgen.1001324 (2011).
28 Ng, E. *et al.* Genome-wide association study of toxic metals and trace elements reveals novel associations. *Hum Mol Genet* **24**, 4739-4745, doi:10.1093/hmg/ddv190 (2015).
29 Corre, T. *et al.* Common variants in CLDN14 are associated with differential excretion of magnesium over calcium in urine. *Pflugers Arch* **469**, 91-103, doi:10.1007/s00424-016-1913-7 (2017).
30 Kichaev, G. *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *American journal of human genetics* **104**, 65-75, doi:10.1016/j.ajhg.2018.11.008 (2019).
31 Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415-1429 e1419, doi:10.1016/j.cell.2016.10.042 (2016).
32 Tian, D. *et al.* GWAS Atlas: a curated resource of genome-wide variant-trait associations in plants and animals. *Nucleic Acids Res* **48**, D927-D932, doi:10.1093/nar/gkz828 (2020).
33 Kim, S. K. Identification of 613 new loci associated with heel bone mineral density and a polygenic risk score for bone mineral density, osteoporosis and fracture. *PloS one* **13**, e0200785, doi:10.1371/journal.pone.0200785 (2018).
34 Morris, J. A. *et al.* An atlas of genetic influences on osteoporosis in humans and mice. *Nature genetics* **51**, 258-266, doi:10.1038/s41588-018-0302-x (2019).
35 Kemp, J. P. *et al.* Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. *Nature genetics* **49**, 1468-1475, doi:10.1038/ng.3949 (2017).
36 Parisinos, C. A. *et al.* Genome-wide and Mendelian randomisation studies of liver MRI yield insights into the pathogenesis of steatohepatitis. *J Hepatol*, doi:10.1016/j.jhep.2020.03.032 (2020).
37 Trieb, M. *et al.* Liver disease alters high-density lipoprotein composition, metabolism and function. *Biochim Biophys Acta* **1861**, 630-638, doi:10.1016/j.bbalip.2016.04.013 (2016).
38 Leyva-Illades, D. *et al.* SLC30A10 is a cell surface-localized manganese efflux transporter, and parkinsonism-causing mutations block its intracellular trafficking and efflux activity. *J Neurosci* **34**, 14079-14095, doi:10.1523/JNEUROSCI.2329-14.2014 (2014).
39 Yuan, X. *et al.* Population-based genome-wide association studies reveal six loci influencing plasma levels of liver enzymes. *Am J Hum Genet* **83**, 520-528, doi:10.1016/j.ajhg.2008.09.012 (2008).

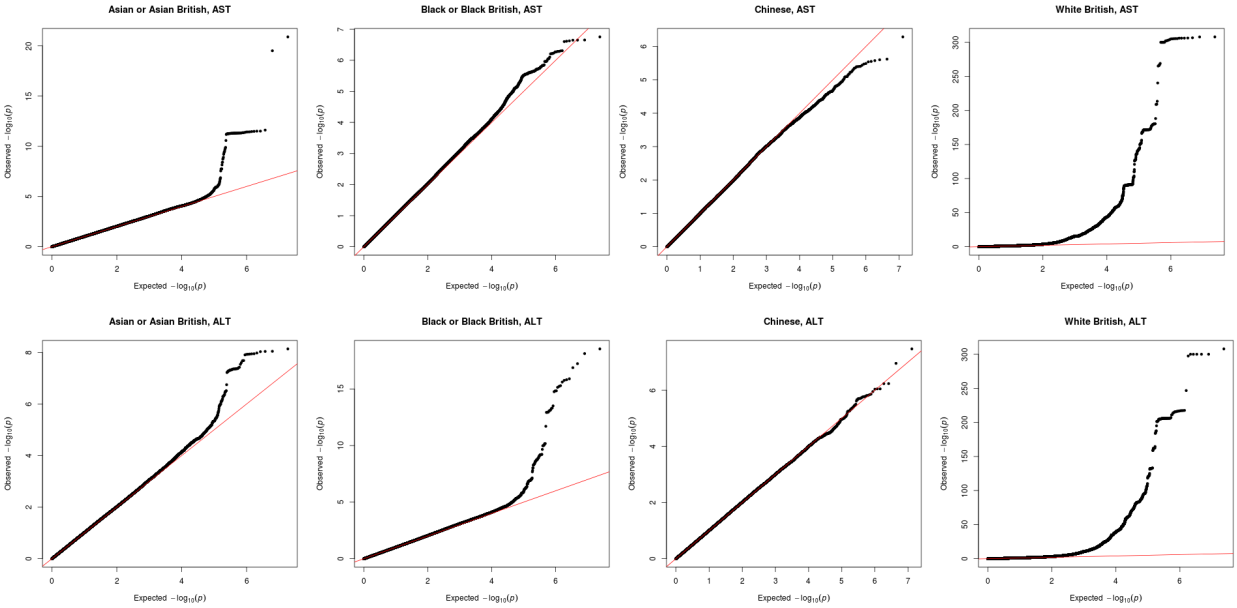
40 Liu, Y. *et al.* Genome-Wide Study Links PNPLA3 Variant With Elevated Hepatic Transaminase After Acute Lymphoblastic Leukemia Therapy. *Clin Pharmacol Ther* **102**, 131-140, doi:10.1002/cpt.629 (2017).
41 Whitfield, J. B. *et al.* Biomarker and Genomic Risk Factors for Liver Function Test Abnormality in Hazardous Drinkers. *Alcohol Clin Exp Res* **43**, 473-482, doi:10.1111/acer.13949 (2019).
42 Xu, C. F. *et al.* HLA-B*57:01 Confers Susceptibility to Pazopanib-Associated Liver Injury in Patients with Cancer. *Clin Cancer Res* **22**, 1371-1377, doi:10.1158/1078-0432.CCR-15-2044 (2016).
43 Feder, J. N. *et al.* A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nature genetics* **13**, 399-408, doi:10.1038/ng0896-399 (1996).
44 Kozlitina, J. *et al.* Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. *Nature genetics* **46**, 352-356, doi:10.1038/ng.2901 (2014).
45 Holmen, O. L. *et al.* Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. *Nature genetics* **46**, 345-351, doi:10.1038/ng.2926 (2014).
46 Liu, Y. L. *et al.* TM6SF2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease. *Nat Commun* **5**, 4309, doi:10.1038/ncomms5309 (2014).
47 Buch, S. *et al.* A genome-wide association scan identifies the hepatic cholesterol transporter ABCG8 as a susceptibility factor for human gallstone disease. *Nature genetics* **39**, 995-999, doi:10.1038/ng2101 (2007).
48 Ferkingstad, E. *et al.* Genome-wide association meta-analysis yields 20 loci associated with gallstone disease. *Nat Commun* **9**, 5101, doi:10.1038/s41467-018-07460-y (2018).
49 van de Steeg, E. *et al.* Complete OATP1B1 and OATP1B3 deficiency causes human Rotor syndrome by interrupting conjugated bilirubin reuptake into the liver. *J Clin Invest* **122**, 519-528, doi:10.1172/JCI59526 (2012).
50 Tsutsumi, S. *et al.* The novel gene encoding a putative transmembrane protein is mutated in gnathodiaphyseal dysplasia (GDD). *American journal of human genetics* **74**, 1255-1261, doi:10.1086/421527 (2004).
51 Penttila, S. *et al.* Eight new mutations and the expanding phenotype variability in muscular dystrophy caused by ANO5. *Neurology* **78**, 897-903, doi:10.1212/WNL.0b013e31824c4682 (2012).
52 Sherry, S. T., Ward, M. & Sirotkin, K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res* **9**, 677-679 (1999).
53 Arnold, L. M., Hirsch, I., Sanders, P., Ellis, A. & Hughes, B. Safety and efficacy of esreboxetine in patients with fibromyalgia: a fourteen-week, randomized, double-blind, placebo-controlled, multicenter clinical trial. *Arthritis and rheumatism* **64**, 2387-2397, doi:10.1002/art.34390 (2012).
54 Zogzas, C. E., Aschner, M. & Mukhopadhyay, S. Structural Elements in the Transmembrane and Cytoplasmic Domains of the Metal Transporter SLC30A10 Are Required for Its Manganese Efflux Activity. *J Biol Chem* **291**, 15940-15957, doi:10.1074/jbc.M116.726935 (2016).

55 Park, J. H. *et al.* SLC39A8 Deficiency: A Disorder of Manganese Transport and Glycosylation. *American journal of human genetics* **97**, 894-903, doi:10.1016/j.ajhg.2015.11.003 (2015).
56 Tuschl, K. *et al.* Mutations in SLC39A14 disrupt manganese homeostasis and cause childhood-onset parkinsonism-dystonia. *Nat Commun* **7**, 11601, doi:10.1038/ncomms11601 (2016).
57 Scheiber, I. F., Wu, Y., Morgan, S. E. & Zhao, N. The intestinal metal transporter ZIP14 maintains systemic manganese homeostasis. *J Biol Chem* **294**, 9147-9160, doi:10.1074/jbc.RA119.008762 (2019).
58 Mercadante, C. J. *et al.* Manganese transporter Slc30a10 controls physiological manganese excretion and toxicity. *J Clin Invest* **129**, 5442-5461, doi:10.1172/JCI129710 (2019).
59 Katz, N. & Rader, D. J. Manganese homeostasis: from rare single-gene disorders to complex phenotypes and diseases. *J Clin Invest* **129**, 5082-5085, doi:10.1172/JCI133120 (2019).
60 Bae, Y. J. & Kim, M. H. Manganese supplementation improves mineral density of the spine and femur and serum osteocalcin in rats. *Biol Trace Elem Res* **124**, 28-34, doi:10.1007/s12011-008-8119-6 (2008).
61 Strause, L. G., Hegenauer, J., Saltman, P., Cone, R. & Resnick, D. Effects of long-term dietary manganese and copper deficiency on rat skeleton. *J Nutr* **116**, 135-141, doi:10.1093/jn/116.1.135 (1986).
62 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).
63 Crossgrove, J. & Zheng, W. Manganese toxicity upon overexposure. *NMR Biomed* **17**, 544-553, doi:10.1002/nbm.931 (2004).
64 O'Neal, S. L. & Zheng, W. Manganese Toxicity Upon Overexposure: a Decade in Review. *Curr Environ Health Rep* **2**, 315-328, doi:10.1007/s40572-015-0056-x (2015).
65 Chen, J. Y., Tsao, G. C., Zhao, Q. & Zheng, W. Differential cytotoxicity of Mn(II) and Mn(III): special reference to mitochondrial [Fe-S] containing enzymes. *Toxicol Appl Pharmacol* **175**, 160-168, doi:10.1006/taap.2001.9245 (2001).
66 Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427, doi:10.1038/nature13595 (2014).
67 Costas, J. The highly pleiotropic gene SLC39A8 as an opportunity to gain insight into the molecular pathogenesis of schizophrenia. *Am J Med Genet B Neuropsychiatr Genet* **177**, 274-283, doi:10.1002/ajmg.b.32545 (2018).
68 Mealer, R. G. *et al.* A schizophrenia risk locus alters brain metal transport and plasma glycosylation. *bioRxiv*, 757088, doi:10.1101/757088 (2019).
69 Krieger, D. *et al.* Manganese and chronic hepatic encephalopathy. *Lancet* **346**, 270-274, doi:10.1016/s0140-6736(95)92164-8 (1995).
70 Rajoriya, N., Brahmania, M. & J, J. F. Implications of Manganese in Chronic Acquired Hepatocerebral Degeneration. *Ann Hepatol* **18**, 274-278, doi:10.5604/01.3001.0012.7938 (2019).

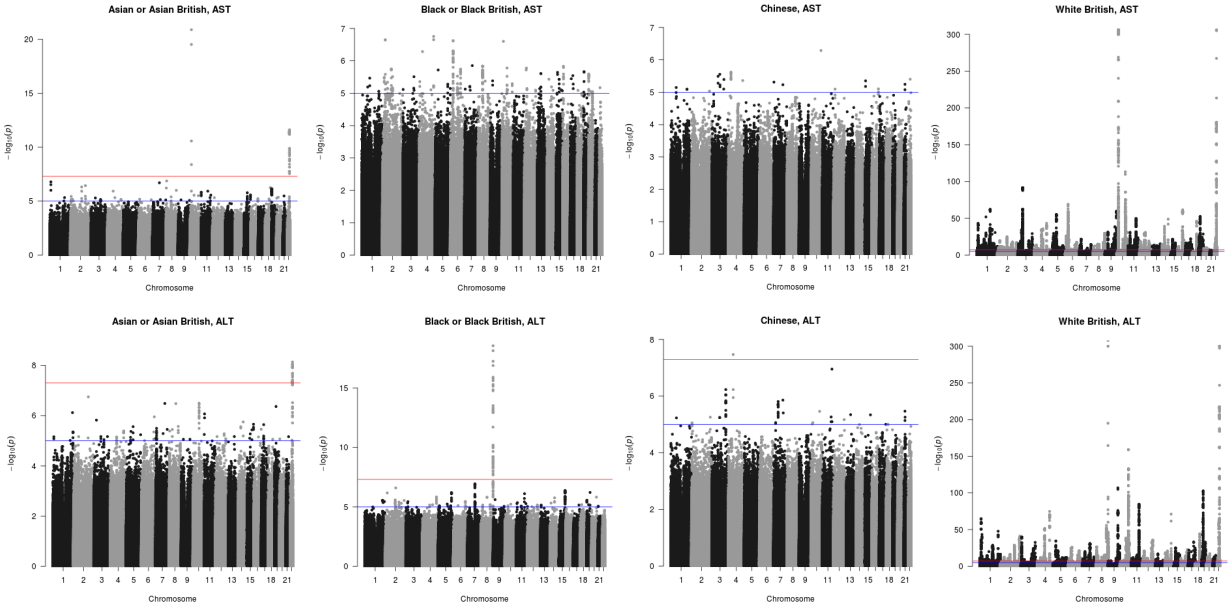
71 Burkhard, P. R., Delavelle, J., Du Pasquier, R. & Spahr, L. Chronic parkinsonism associated with cirrhosis: a distinct subset of acquired hepatocerebral degeneration. *Arch Neurol* **60**, 521-528, doi:10.1001/archneur.60.4.521 (2003).
72 Anagianni, S. & Tuschl, K. Genetic Disorders of Manganese Metabolism. *Curr Neurol Neurosci Rep* **19**, 33, doi:10.1007/s11910-019-0942-y (2019).
73 Ebert, B. L. & Bunn, H. F. Regulation of the erythropoietin gene. *Blood* **94**, 1864-1877 (1999).
74 Tyson, G. L. & El-Serag, H. B. Risk factors for cholangiocarcinoma. *Hepatology* **54**, 173-184, doi:10.1002/hep.24351 (2011).
75 Razumilava, N. & Gores, G. J. Cholangiocarcinoma. *Lancet* **383**, 2168-2179, doi:10.1016/S0140-6736(13)61903-0 (2014).
76 Kim, A. Modulation of MnSOD in Cancer:Epidemiological and Experimental Evidence. *Toxicol Res* **26**, 83-93, doi:10.5487/TR.2010.26.2.083 (2010).
77 Landrum, M. J. *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* **42**, D980-985, doi:10.1093/nar/gkt1113 (2014).
78 Stamelou, M. & Bhatia, K. P. A new treatable genetic disorder of manganese metabolism causing dystonia-parkinsonism and cirrhosis: the "new" Wilson's disease? *Mov Disord* **27**, 962, doi:10.1002/mds.25031 (2012).
79 Bastarache, L. *et al.* Phenotype risk scores identify patients with unrecognized Mendelian disease patterns. *Science* **359**, 1233-1239, doi:10.1126/science.aal4043 (2018).
80 Hou, Y.-C. C. *et al.* Precision Medicine Advancements Using Whole Genome Sequencing, Noninvasive Whole Body Imaging, and Functional Diagnostics. *bioRxiv*, 497560, doi:10.1101/497560 (2018).
81 Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv*, 166298, doi:10.1101/166298 (2017).
82 Bellenguez, C. *et al.* A robust clustering algorithm for identifying problematic samples in genome-wide association studies. *Bioinformatics* **28**, 134-135, doi:10.1093/bioinformatics/btr599 (2012).
83 Nioi, P. *et al.* Variant ASGR1 Associated with a Reduced Risk of Coronary Artery Disease. *The New England journal of medicine* **374**, 2131-2141, doi:10.1056/NEJMoa1508419 (2016).
84 McCaw, Z. R., Lane, J. M., Saxena, R., Redline, S. & Lin, X. Operating Characteristics of the Rank-Based Inverse Normal Transformation for Quantitative Trait Analysis in Genome-Wide Association Studies. *bioRxiv*, 635706, doi:10.1101/635706 (2019).
85 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209, doi:10.1038/s41586-018-0579-z (2018).
86 Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature genetics* **50**, 1335-1341, doi:10.1038/s41588-018-0184-y (2018).
87 Turner, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv*, 005165, doi:10.1101/005165 (2014).
88 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191, doi:10.1093/bioinformatics/btq340 (2010).

89 Zhou, W. (2018).
90 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7, doi:10.1186/s13742-015-0047-8 (2015).
91 Yin, L. *A high-quality drawing tool designed for Manhattan plot of genomic analysis*, <https://github.com/YinLiLin/R-CMplot> (2018).
92 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
93 McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122, doi:10.1186/s13059-016-0974-4 (2016).
94 Consortium, G. T. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213, doi:10.1038/nature24277 (2017).
95 Strunz, T. *et al.* A mega-analysis of expression quantitative trait loci (eQTL) provides insight into the regulatory architecture of gene expression variation in liver. *Sci Rep* **8**, 5865, doi:10.1038/s41598-018-24219-z (2018).
96 Võsa, U. *et al.* Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv*, 447367, doi:10.1101/447367 (2018).
97 Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005-D1012, doi:10.1093/nar/gky1120 (2019).
98 Van Hout, C. V. *et al.* Whole exome sequencing and characterization of coding variation in 49,960 individuals in the UK Biobank. *bioRxiv*, 572347, doi:10.1101/572347 (2019).
99 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
100 *Picard*, <http://broadinstitute.github.io/picard/> (
101 *WeCall*, <https://github.com/Genomicsplc/wecall> (
102 Lin, M. F. *et al.* GLnexus: joint variant calling for large cohort sequencing. *bioRxiv*, 343970, doi:10.1101/343970 (2018).
103 Gordon, M. & Lumley, T. forestplot: Advanced Forest Plot Using'grid'Graphics. *R package version* **1** (2015).
104 Ward, L. D. & Kellis, M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res* **44**, D877-881, doi:10.1093/nar/gkv1340 (2016).
105 Zhao, Y., Feresin, R. G., Falcon-Perez, J. M. & Salazar, G. Differential Targeting of SLC30A10/ZnT10 Heterodimers to Endolysosomal Compartments Modulates EGF-Induced MEK/ERK1/2 Activity. *Traffic* **17**, 267-288, doi:10.1111/tra.12371 (2016).
106 Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886-D894, doi:10.1093/nar/gky1016 (2019).
107 Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nat Protoc* **11**, 1-9, doi:10.1038/nprot.2015.123 (2016).
108 Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248-249, doi:10.1038/nmeth0410-248 (2010).

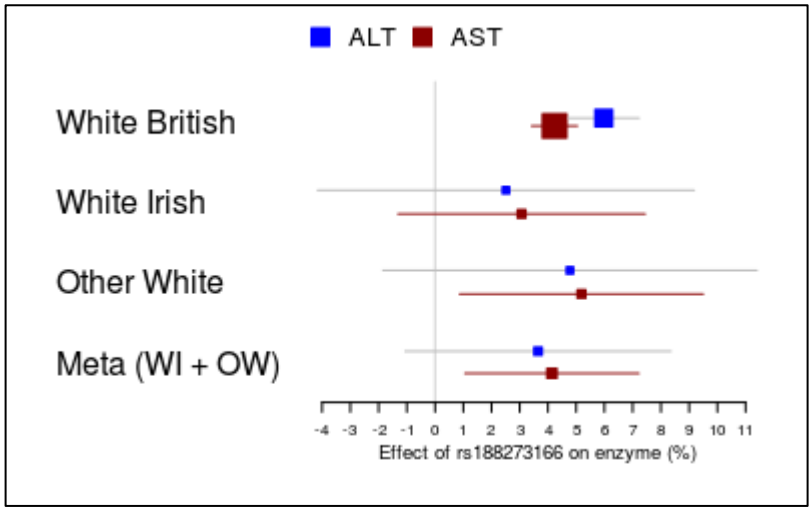
Supplementary Figures



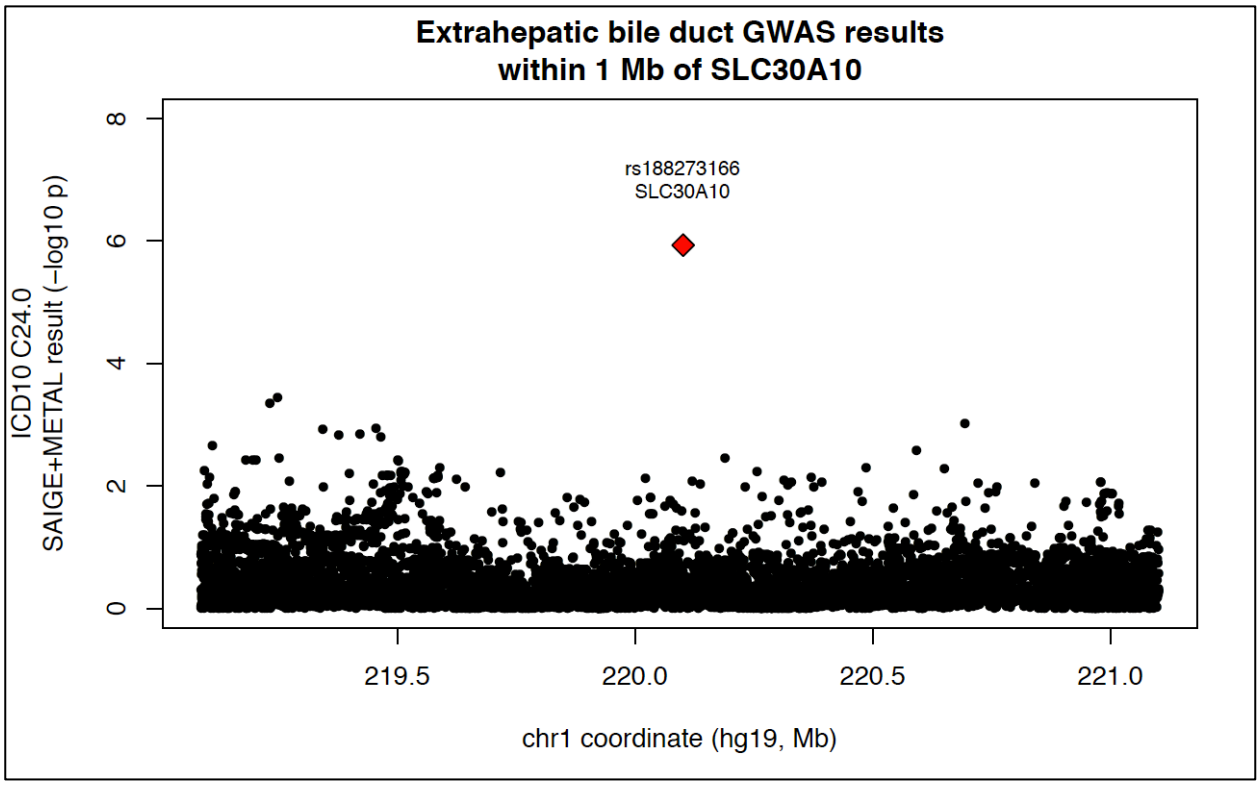
Supplementary Figure 1: Q-Q plots of GWAS p values (from SAIGE) for each sub-population and enzyme.



Supplementary Figure 2: Manhattan plots of GWAS p values (from SAIGE) for each sub-population and enzyme.



Supplementary Figure 3: Forest plot of *SLC30A10* Thr95Ile (rs188273166) association with ALT and AST in White British, White Irish, and Other White populations, and meta-analysis results (in METAL) of White Irish and Other White association results, combined. Boxes show effect size estimates from PLINK, and error bars show 95% confidence intervals.



Supplementary Figure 4: Association results (p value from SAIGE followed by METAL) for ICD10 diagnosis C24.0, extrahepatic bile duct cancer, in the *SLC30A10* gene and 1 megabase flanking sequence.