GWAS of serum ALT and AST reveals an association of *SLC30A10* Thr95Ile with hypermanganesemia symptoms

Lucas D. Ward[1], Ho-Chou Tu[1], Chelsea B. Quenneville[1], Shira Tsour[1], Alexander O. Flynn-Carroll[1], Margaret M. Parker[1], Aimee M. Deaton[1], Patrick A. J. Haslett[1], Luca A. Lotta[2], Niek Verweij[2], Manuel A. R. Ferreira[2], Regeneron Genetics Center, Geisinger-Regeneron DiscovEHR Collaboration, Aris Baras[2], Gregory Hinkle[1], Paul Nioi[1]

1. Alnylam Pharmaceuticals, Cambridge, MA 02142
2. Regeneron Genetics Center, Tarrytown, NY 10591

Abstract

To investigate mechanisms of hepatocellular damage, we performed genome-wide association studies (GWAS) on alanine aminotransferase (ALT) and aspartate aminotransferase (AST) serum activities across 411,048 subjects from four ancestry groups in the UK Biobank, and found 100 loci associating with both enzymes. The rare missense variant *SLC30A10* Thr95Ile (rs188273166) associates with a larger elevation in ALT and AST than any other variant tested and this association also replicates in the DiscovEHR study. SLC30A10 excretes manganese from the liver to the bile duct, and rare homozygous loss of function causes the syndrome hypermanganesemia with dystonia-1 (HMNDYT1) which involves cirrhosis. Consistent with hematological symptoms of hypermanganesemia, *SLC30A10* Thr95Ile carriers have increased hematocrit and risk of iron deficiency anemia. Carriers also have increased risk of extrahepatic bile duct cancer. These associations suggest that genetic variation in *SLC30A10* adversely affects more individuals than patients with diagnosed HMNDYT1.

Introduction

Liver disease remains an area of high unmet medical need, causing 3.5% of deaths worldwide, and the burden of liver disease is rising rapidly, driven mainly by increasing rates of nonalcoholic fatty liver disease (NAFLD)[1,2]. Better characterizing the genetic determinants of liver disease may lead to new therapies[3]. 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 preclinical drug evaluation to anticipate and avoid off-target effects[4,5].

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[6,7]. The activities of circulating liver enzymes have been found to be highly heritable[6,8-12], and variation even within the normal reference range is predictive of disease[13]. Accordingly, genome-wide association studies (GWAS) of activities of circulating liver enzymes across large population samples have proven powerful for understanding the molecular basis of liver disease[6,14-24]. Combined GWAS of ALT and AST have previously revealed genetic associations providing potential therapeutic targets for liver disease such as *PNPLA3*[25] and *HSD17B13*[26]. To further study the genetics of hepatocellular damage, we performed GWAS on serum activities of ALT and AST in 411,048 subjects, meta-analyzed across four ancestry groups in the UK Biobank (UKBB).

Results

Discovery of ALT- and AST-associated loci by GWAS

We performed a GWAS of ALT and AST in four sub-populations in the UKBB (demographic properties, **Supplementary Table 1;** sample sizes, number of variants tested, and λ_{GC} values, **Supplementary Table 2;** genome-wide significant associations, **Supplementary Table 3**; 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 single nucleotide polymorphisms (SNPs) or indels separated by at least 500 kilobases and pairwise linkage disequilibrium (LD) r^2 less than 0.2. Enzyme activities 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, the aim of this study was to detect genetic determinants of liver health that have downstream effects on both ALT and AST due to hepatocellular damage; therefore we focused the remainder of our analyses only on the variants associated with serum activity of both enzymes (labeled with black text on **Figure 1**).

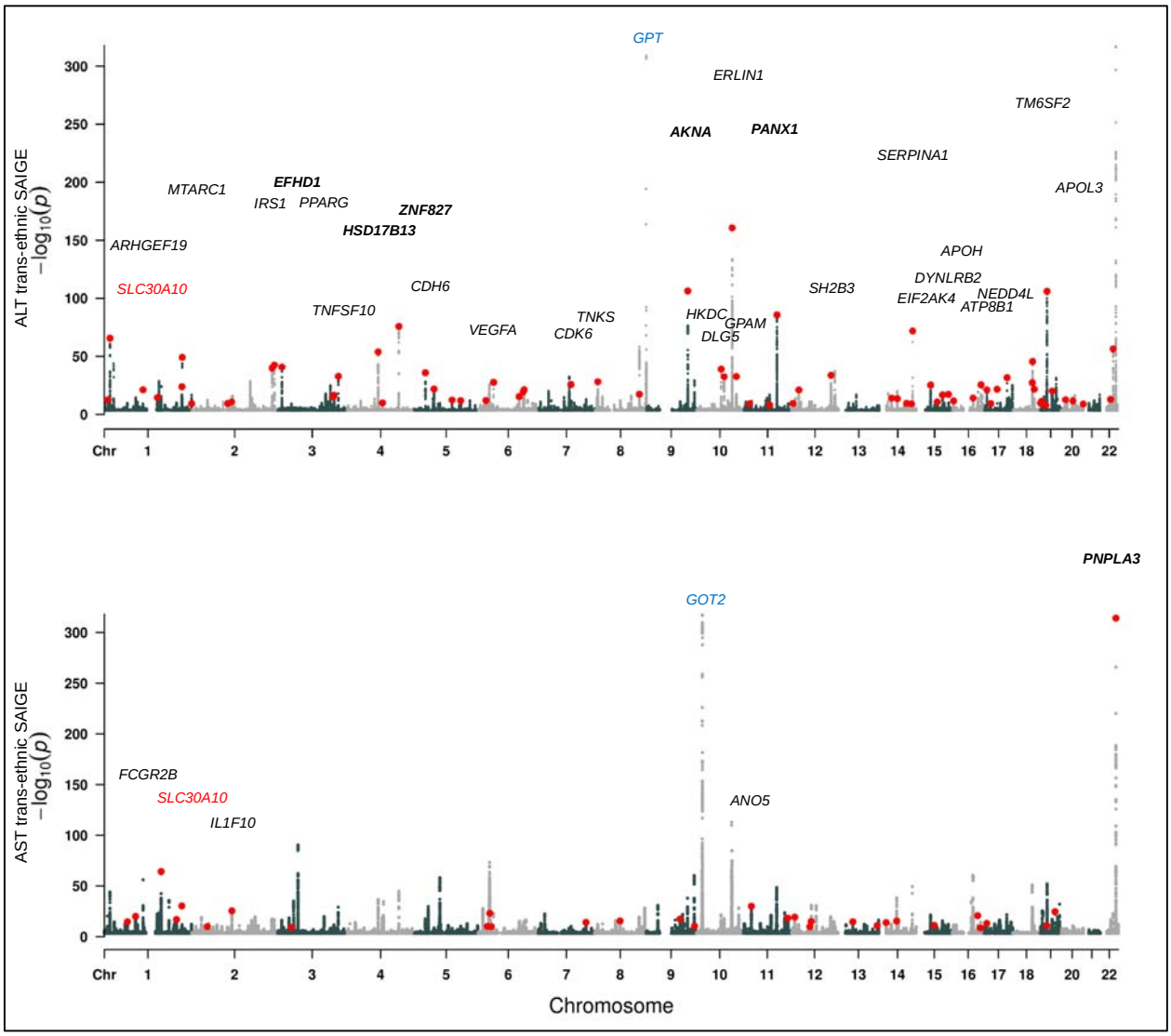


Figure 1: Manhattan plots showing trans-ancestry GWAS results for ALT and AST. Red dots indicate lead variants for shared signals between the two GWAS; for clarity, the shared signals are marked only once, on the plot for the GWAS in which the more significant association is detected. Cis-pQTLs (at *GPT* and *GOT2*) are labeled in blue. Loci with shared signals are labeled (for clarity, only when p < 10-25 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 variants 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 4**). 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 variant in the EBI-NHGRI GWAS Catalog, and 15 were within 500kb of a published ALT or AST variant; 33 of the loci harbored a missense or predicted protein-truncating variant; 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.

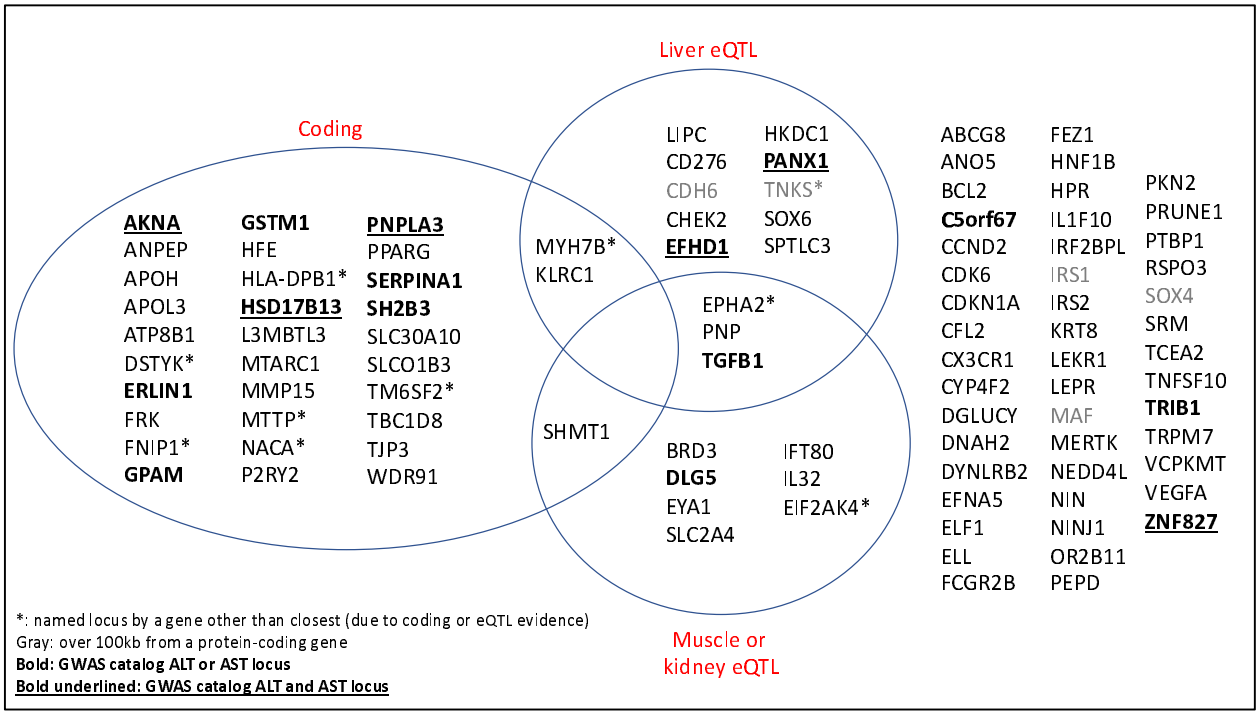


Figure 2: Classification of the top ALT- and AST- associated loci based on annotations. Unless otherwise noted with an asterisk, loci are named by the closest protein-coding gene. "Coding" indicates that one of the variants linked to the lead variant is predicted to have a moderate or high impact on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

Comparing the effect sizes of all lead variants (**Supplementary Table 4**), the strongest estimated effect 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 variants, 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)[27,28]. 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[28-34]. 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[35], 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[25]; it is

121 associated with a 2.2% increase in ALT (95% CI, 2.1% to 2.3%; $p < 10^{-300}$) and a 1.3% increase
122 in AST (95% CI, 1.3% to 1.4%; $p < 10^{-300}$).
123
124 We observed significant heterogeneity in effects between sexes for both enzymes (Cochran's Q
125 test $p < 0.05/100$ for both enzymes) for three of the lead variants: rs9663238 at *HKDC1* (stronger
126 effects in women), rs28929474 at *SERPINA1* (stronger effects in men), and rs1890426 at *FRK*
127 (stronger effects in men) (**Supplementary Table 5**).
128
129 We tested the 100 lead variants from the ALT and AST GWAS analysis for association with a
130 broad liver disease phenotype (ICD10 codes K70-77; 14,143 cases and 416,066 controls), meta-
131 analyzing liver disease association results across all four sub-populations (**Supplementary**
132 **Table 4)**. Of the 100 lead variants, 28 variants associate with liver disease with $p < 0.05$. As
133 expected, variants associated with an increase in ALT and AST tend to be associated with a
134 proportional increase in liver disease risk (across all lead variants, Pearson correlation of betas r
135 = 0.82 for both enzymes; **Figure 3**). Liver disease is found more frequently in our sample of
136 carriers of *SLC30A10* Thr95Ile (rs188273166), proportional with the observation of increased
137 ALT and AST (OR = 1.47); however, owing to the small sample size of carriers and liver disease
138 cases, we are underpowered to confidently determine whether this high point estimate is due to
139 chance (although the 95% CI from the PLINK analysis used to estimate effects does not include
140 OR = 0, the p value from the SAIGE analysis which more accurately controls for Type I error in
141 highly unbalanced case-control studies is 0.07; see Methods).

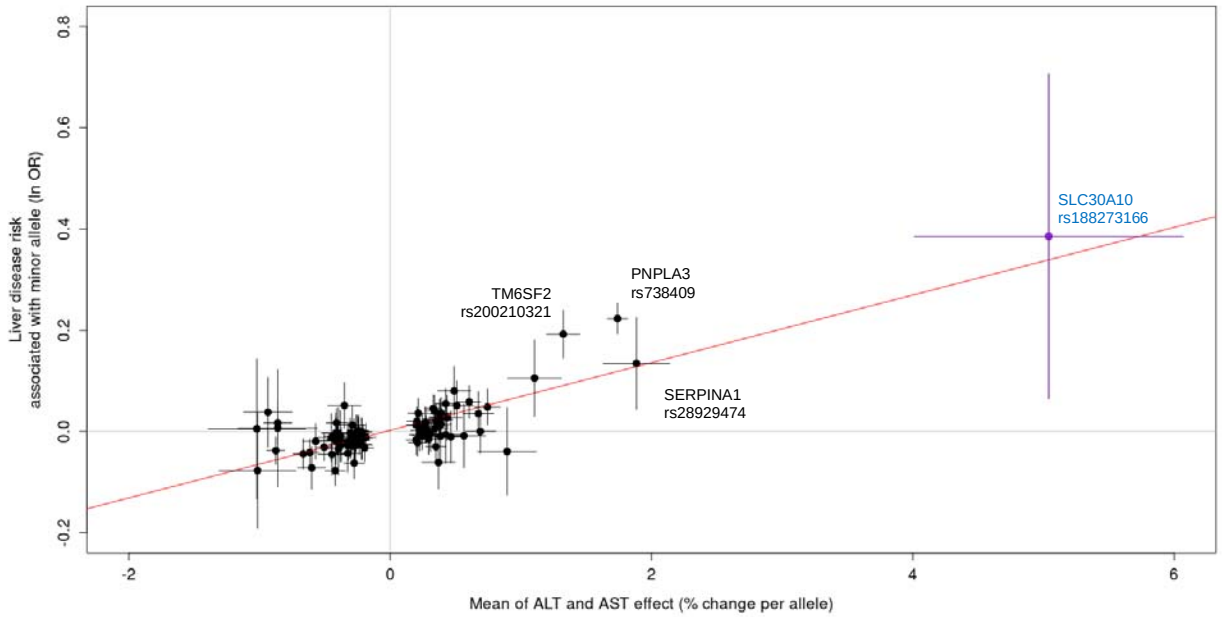


Figure 3: Comparison of effect of lead variants on ALT and AST with effect on liver disease. Effect sizes (beta from the regression of units of $\log_{10}$ ALT and AST, equivalent to percent change; and beta from the liver disease regression, equivalent to natural log of the odds ratio) are from PLINK analysis; error bars are 95% confidence intervals on these effect sizes from PLINK.

Because *SLC30A10* Thr95Ile had the strongest effect on ALT and AST of all of our lead variants and 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

Because rare variants are especially prone to errors in array genotyping[36], 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 6**). The only 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 exome data available, 699 (99.6%) were confirmed heterozygous by exome sequencing, while three were called homozygous reference by exome sequencing, suggesting an error either in the array typing or exome sequencing for these three individuals. Overall, these results demonstrate high concordance between array and exome sequencing, implying highly reliable genotyping.

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 activity 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)[7], 5.6% of carriers vs 3.6% of noncarriers had both enzymes elevated at the time of their UK Biobank sample collection, an increased relative risk of 58% (Fisher's $p = 8.1 \times 10^{-4}$) (**Supplementary Table 7**).

Drinking behavior in *SLC30A10* Thr95Ile carriers

The *SLC30A10* Thr95Ile has not been reported as associating with drinking behavior by any of the available studies in the GWAS Catalog. We used the drinking questionnaire taken by UK Biobank participants to assess drinking status at enrollment of *SLC30A10* Thr95Ile carriers (current, former, or never drinkers.) While the rate of current drinkers is higher among carriers vs non-carriers in the entire biobank (93.7% vs 91.7%, Fisher's $p = 0.019$) (**Supplementary Table 7**), this association is highly confounded by genetic ancestry and country of birth

(**Supplementary Table 8**). Limiting to the White British subpopulation and individuals born in England, the rate of current drinking is not detectably different among carriers (94.0% vs 93.4%, Fisher's p = 0.57) while the rate of individuals with elevation of both ALT and AST over the ULN remains significant (5.5% vs. 3.5%, Fisher's $p = 4.6 \times 10^{-3}$) (**Supplementary Table 7**).

Replication of ALT and AST associations

The initial association of rs188273166 with ALT and AST was identified in the White British population. To replicate this association in independent cohorts, we first identified groups besides the White British subpopulation harboring the variant in the UKBB. The only two other populations with a substantial number of *SLC30A10* Thr95Ile carriers were individuals identifying as Other White and as White Irish (**Supplementary Table 8**); we tested for association with ALT and AST in these subpopulations. We then tested the association in two independent cohorts from the DiscovEHR collaboration between the Regeneron Genetics Center and the Geisinger Health System[37]. Meta-analyzing the association results across these four groups (N = 132,992 and N = 131,646, respectively) confirmed the Thr95Ile association with increased ALT and AST ($p = 6.5 \times 10^{-5}$ and $p = 5.4 \times 10^{-6}$, respectively) (**Supplementary Figure 3, Supplementary Table 9**). We also searched repositories of available complete summary statistics for ALT and AST GWAS and found two prior studies that reported associations[21,38]. Although these studies were underpowered to detect significant associations and were not reported in units that allowed their inclusion in our replication analysis, they were consistent with increases in both enzymes (**Supplementary Table 10**).

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 to arrive at a set of lead variants, it was unclear how many independent association signals existed at the *SLC30A10* locus. Revisiting trans-ancestry association results in a window including 1 Mb flanking sequence upstream and downstream of *SLC30A10* revealed 76 variants with genome-wide significant associations with both ALT and AST (**Figure 4**). These 76 variants clustered into three loci: *SLC30A10* (only Thr95Ile, rs188273166); *MTARC1* (mitochondrial amidoxime reducing component 1, lead variant rs2642438 encoding missense Ala165Thr, previously reported to associate with liver disease and liver enzymes[39], and six additional variants together spanning 68 kilobases); and *LYPLAL1-ZC3H11B* (intergenic region between lyophospholipase like 1 and zinc finger CCCH-type containing 11B, with array-genotyped variant rs6541227 and 67 imputed variants spanning 46 kilobases), a locus previously reported to associate with non-alcoholic fatty liver disease (NAFLD)[40].

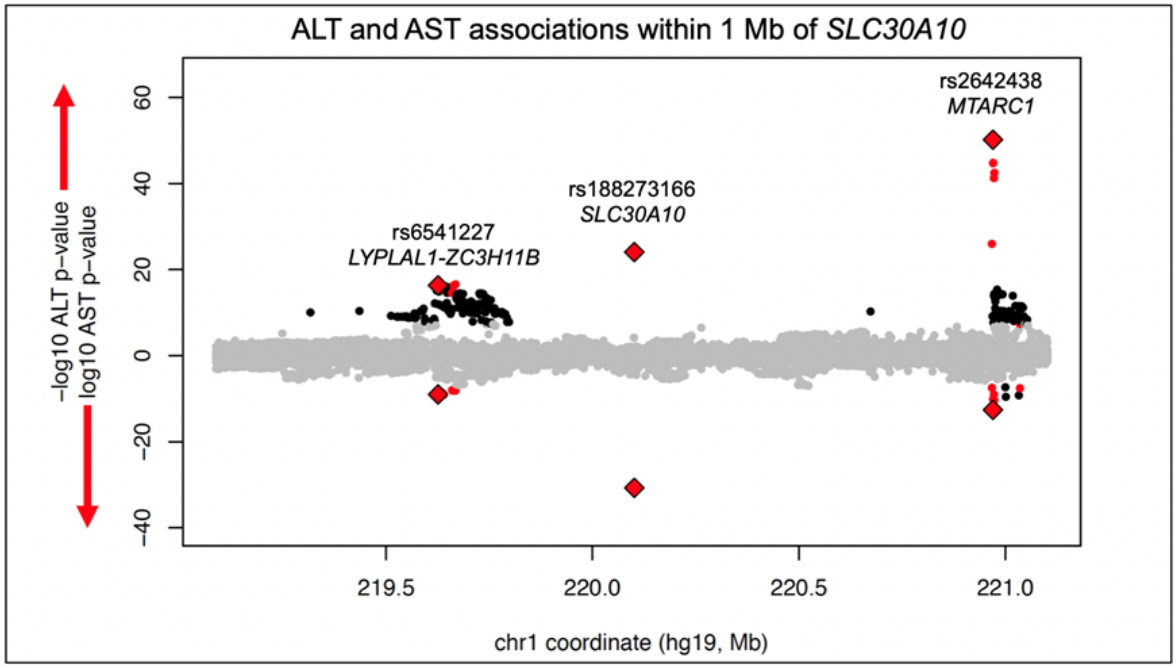


Figure 4: Miami plot of trans-ancestry SAIGE GWAS results within 1 Mb of the gene body of *SLC30A10*. ALT associations are shown in the positive direction and AST associations in the negative direction. Variant associations reaching genome-wide significance for one enzyme are colored black; for both enzymes, colored red; directly-genotyped variants 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 variants 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 11**.) Therefore, the *SLC30A10* Thr195Ile association is statistically independent of the associations at neighboring loci. This statistical independence of the liver enzyme associations does not preclude a long-distance regulatory interaction between the three loci; for example, rs188273166, despite encoding an amino acid

change in *SLC30A10*, could conceivably influence transcription of *MTARC1*, and rs6541227, despite being nearest to *LYPLAL1* and *ZC3H11B*, may influence transcription of *SLC30A10*. However, these three variants are not detected as liver eQTLs for the genes at neighboring loci in published data[41].

Linkage of Thr95Ile to GWAS variants at *SLC30A10*

A GWAS of circulating toxic metals[42] discovered an association between a common intronic variant 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 variant 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 12**); however, genotypes of Thr95Ile in the manganese GWAS or manganese measurements in the UK Biobank would be needed in order to perform conditional analysis or directly measure association of Thr95Ile with serum manganese. We then systematically tested nearby variants 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 13**). In addition to increased blood

Mn[42], these three common alleles have been associated with decreased magnesium/calcium ratio in urine[43], decreased mean corpuscular hemoglobin (MCH)[44-46], increased red blood cell distribution width[44-46], and increased heel bone mineral density (BMD)[46-49]. A recent study, not yet in the GWAS catalog, reported an association between another common intronic variant in *SLC30A10* (rs759359281; MAF in White British, 5.6%) and liver MRI-derived iron-corrected T1 measures (cT1)[50]. However, the reported cT1-increasing 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 4,398 ICD10 diagnosis codes within the White British population (**Supplementary Tables 14 and 15**). We were particularly interested in testing associations with phenotypes related to HMNDYT1, the known syndrome caused by homozygous loss of function of SLC30A10. Besides ALT and AST elevation, rs188273166 was associated with other indicators of hepatobiliary damage such as decreased HDL cholesterol and apolipoprotein A (ApoA)[51], decreased albumin, and increased gamma glutamyltransferase (GGT). Other phenome-wide significant quantitative trait associations were increases in hemoglobin concentration and hematocrit (**Table 1**); increased hematocrit, or polycythemia, is a known symptom of HMNDYT1. Liver iron-corrected T1 by MRI (cT1), although only measured in seven carriers, was above the population median value in all seven **(Supplementary Figure 4)**.

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
Liver MRI: iron corrected T1 (cT1)	3.31E-03	*	1.22	0.48	1.96	2407

Table 1: Association results of *SLC30A10* Thr95Ile with selected quantitative traits. Association tests were performed in the White British population. 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 associations of diagnoses with *SLC30A10* Thr95Ile were C24.0, extrahepatic bile duct carcinoma, and C22.1, intrahepatic bile duct carcinoma. There are eight Thr95Ile carriers of each type of cancer, and six carriers with both types of cancer, for a total of ten carriers (1% of the 1,001 total carriers in the White British population) with bile duct carcinoma. Strikingly, over 5% of individuals with extrahepatic bile duct carcinoma (8 in 148) carry Thr95Ile. (**Table 2, Supplementary Table 15**).

Among hematological manifestations of HMNDYT1, iron deficiency anemia was enriched among carriers (OR = 1.5, 95% CI, 1.1 to 1.9; $p = 4.0 \times 10^{-3}$). Searching for neurological manifestations similar to HMNDYT1, we find no association with Parkinson's disease or dystonia but note that, as with liver diseases, we are powered to exclude only strong effects because of the small case number for these traits (**Supplementary Table 15**).

	Association with *SLC30A10* Thr95Ile									
Diagnosis	P	sig.	N cases	Prevalence	N cases (carriers)	Prevalence (carriers)	N cases expected (carriers)	OR	95% CI	
C24.0: Malignant neoplasm of extrahepatic bile duct	2.21E-09	***	148	0.04%	8	0.80%	0.4	18.2	7.4	44.7
C22.1: Intrahepatic bile duct carcinoma	6.22E-06	**	424	0.10%	8	0.80%	1.0	6.9	3.1	15.5
K83.0: cholangitis	1.64E-05	*	1297	0.32%	13	1.30%	3.2	3.4	1.8	6.6
K83.1: obstruction of bile duct	1.12E-04	*	1585	0.39%	13	1.30%	3.9	2.7	1.4	5.3
D50.9: Iron deficiency anemia, unspecified	9.86E-04	*	14687	3.60%	55	5.49%	36.0	1.7	1.3	2.3
K81.0: Acute cholecystitis	1.50E-03	*	2270	0.56%	14	1.40%	5.6	2.8	1.6	4.8
D50: Iron deficiency anemia	4.02E-03	*	20073	4.92%	69	6.89%	49.2	1.5	1.1	1.9

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 genome-wide significance $p < 5 \times 10^{-8}$; (**) indicates phenome-wide significance of $p < 1.1 \times 10^{-5}$; (*) indicates nominal significance of $p < 0.05$. Case counts are all limited to the White British subpopulation (N = 408,160), the only subpopulation with a substantial number of carriers (N = 1,001).

The top non-cancer hepatobiliary associations with *SLC30A10* Thr95Ile were with K83.0, cholangitis; K83.1, obstruction of bile duct; and K81.0, acute cholecystitis. Because biliary diseases are risk factors for cholangiocarcinoma and co-occur with them in our data, we tested whether *SLC30A10* Thr95Ile was still associated with these biliary diseases, and the other selected quantitative traits and diagnoses, after removing the 148 individuals with extrahepatic bile duct cancer (**Supplementary Table 16, Supplementary Table 17**). All of the associations remained significant except for intrahepatic bile duct carcinoma.

To test whether the association with extrahepatic bile duct cancer was driven by a nearby association, and to assess other risk variants and the potential for false positives given the extreme case-control imbalance, we performed a GWAS of the phenotype; remarkably, *SLC30A10* Thr95Ile was the strongest association genomewide, with minimal evidence for systematic inflation of p-values (λ_{GC} = 1.05; **Figure 5, Supplementary Figure 5**).

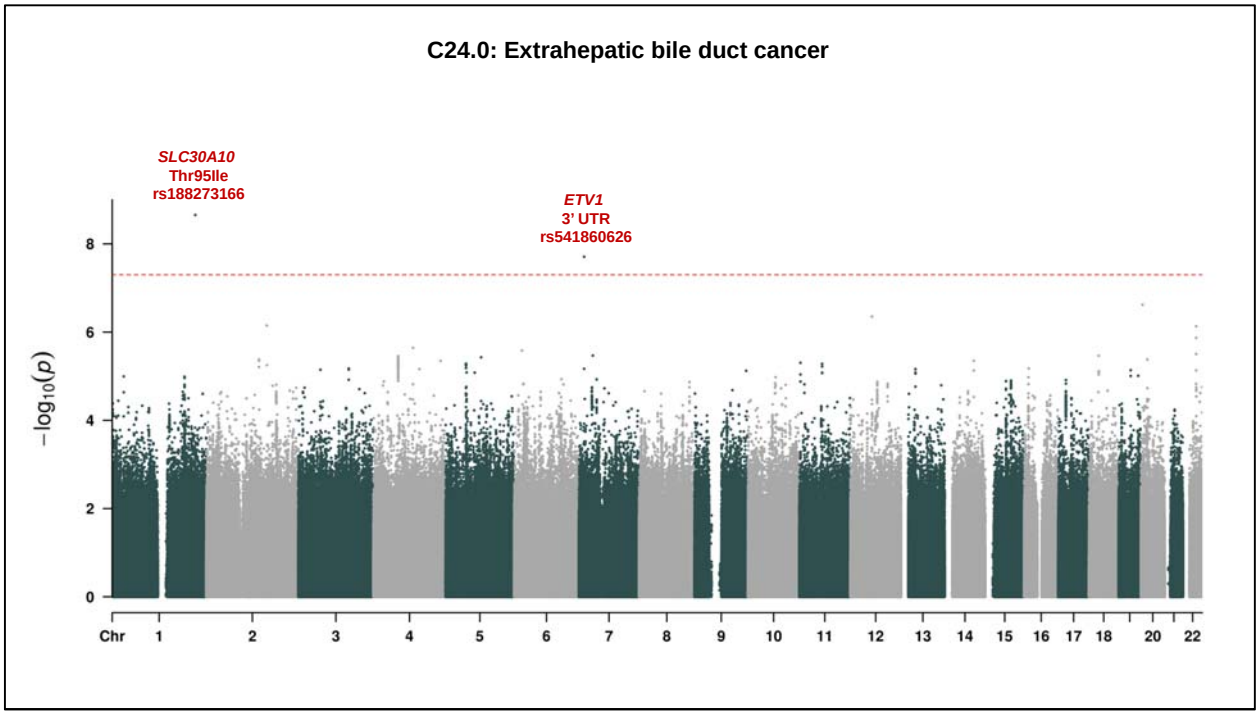


Figure 5. GWAS of extrahepatic bile duct cancer. P values are from SAIGE in White British. Dashed line is genome-wide significance level of $p = 5 \times 10^{-8}$.

Replication of hematocrit association

A key result from the phenome-wide scan that was not related to hepatocellular damage was the association between *SLC30A10* Thr95Ile and increased hematocrit. Polycythemia is a symptom of HMNDYT1 mechanistically related to manganese overload. We meta-analyzed hematocrit

values from the Other White and White Irish populations, the DiscovEHR data, and a non-UKBB population (INTERVAL Study) from a published meta-analysis of hematocrit values[45], and found that the association replicated (N = 179,689, p = 0.013; **Supplementary Table 18**).

SLC30A10 expression in liver cell subtypes

Across organs, SLC30A10 is transcribed at the highest level in liver according to data from the GTEx Project[52]. The association of Thr95Ile with bile duct cancer led us to query expression of SLC30A10 in specific cell types within the liver using data from three single-cell RNA sequencing studies of liver[53-55]. These data show very low expression of SLC30A10 message in individual cells, but all studies detect expression in both hepatocytes and cholangiocytes (**Supplementary Table 19, Supplementary Figure 7**). Immunohistochemistry has established that SLC30A10 protein is present in hepatocytes and bile duct epithelial cells and localizes to the cholangiocyte plasma membrane, facing the lumen of the bile duct[32].

Bioinformatic characterization of *SLC30A10* Thr95Ile

To understand potential functional mechanisms of the Thr95Ile variant, we examined bioinformatic annotations of SLC30A10 Thr95Ile from a variety of databases. 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 (**Supplementary Figure 8**). Several *in silico* algorithms predict that Thr95Ile is a damaging mutation. The CADD (Combined Annotation Dependent Depletion) algorithm, which combines a broad range of functional annotations, gives the variant a score of 23.9, placing it in the top 1% of deleteriousness scores for genome wide potential variants. The algorithm SIFT, which uses sequence homology and physical properties

of amino acids, predicts Thr95Ile as deleterious. The algorithm PolyPhen-2 gives Thr95Ile a HumDiv score of 0.996 (probably damaging), based on patterns of sequence divergence from close mammalian homologs, and a HumVar score of 0.900 (possibly damaging), based on similarity to known Mendelian mutations. 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.

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 (rs281860285) 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 variant proteins are mislocalized in the endoplasmic reticulum (ER)[56]. 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**).

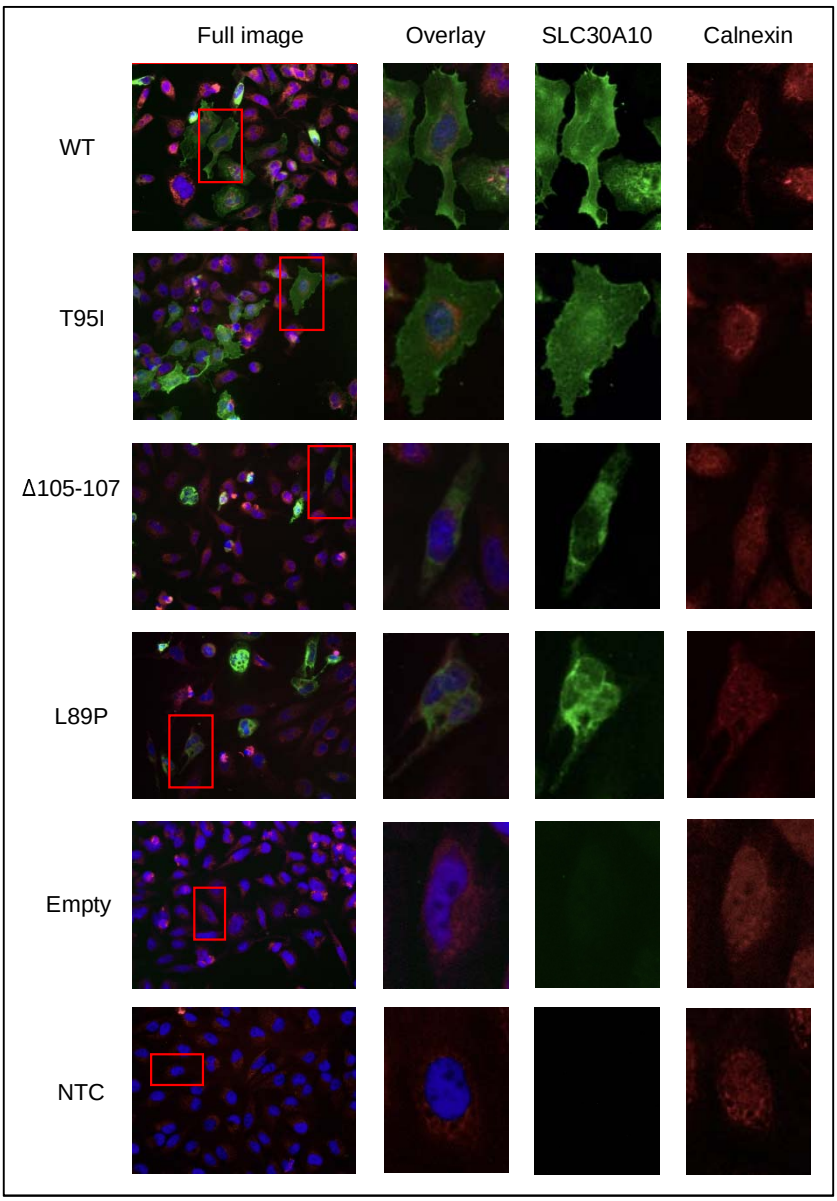


Figure 6: Immunofluorescence imaging of SLC30A10 protein constructs expressed in cultured HeLa cells. WT = wild type; T95I = Thr95Ile; Δ105-107 and L89P, HMNDYT1-causing variants reported previously [56]; empty = transfected with empty vector; NTC = non transfected control. Calnexin staining (in red) indicates the endoplasmic reticulum (ER).

Discussion

Expanded genetic landscape of risk for hepatocellular damage

Our trans-ancestry 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 serum activities of both enzymes from 10 (currently in the GWAS Catalog) to 100. Two loci had been previously reported in majority-European ancestry GWAS of ALT and AST as associating with both enzymes: *PNPLA3*[14,15,18-20,57-59] and *HSD17B13*[14,18,26]; we detect ALT and AST signals at both of these loci. Broadening beyond majority-European ancestry GWAS, an additional eight loci have been previously identified: *PANX1*[18], *ALDH2*[17,19], *CYP2A6*[18,57], *ABCB11*[18], *ZNF827*[18], *EFHD1*[18], *AGER-NOTCH4*[18,60], and *AKNA*[18]; we replicate four of these in our trans-ancestry 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 GWAS of both ALT and AST, such as *SERPINA1* (associated with alpha-1 antitrypsin deficiency[35]), *HFE* (homeostatic iron regulator, associated with hemochromatosis[61]), and *TM6SF2* (transmembrane 6 superfamily member 2, associated with NAFLD[62-64]). The *MTARC1* lead variant was discovered in a GWAS of cirrhosis, and then found to associate with lower ALT and AST[39]. Others are known to associate with risk of gallstones (*ABCG8*, *ANPEP*, and *HNF1B*)[65,66] or increased GGT (*EPHA2*, *CDH6*, *DLG5*, *CD276*, *DYNLRB2*, and *NEDD4L*)[14,18]. 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[67,68], and several loci associated with expression of genes in muscle or kidney but not liver (*SHMT1*, *BRD3*, *DLG5*, *EYA1*, *IFT80*, *IL32*, *EIF2AK4*, and *SLC2A4*). 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.

The significant sex heterogeneity we observe at the ALT- and AST- associated loci *HKDC1*, *SERPINA1*, and *FRK* warrants further investigation and is consistent with a prior study that found significant genotype-by-sex interactions in the genetic architecture of circulating liver enzymes[6]. *HKDC1* has been associated with glucose metabolism and notably, this effect is specific to pregnancy[69,70].

SLC30A10 had not previously been identified in prior GWAS of circulating liver enzymes. Because *SLC30A10* Th95Ile is so rare, it is not surprising that these scans were underpowered to detect its large effect, due either to insufficient study size, lack of inclusion on the genotyping arrays used, or lack of power to impute its genotype. For example, in the UK Household Longitudinal Study (N = 5,458 and N = 5,321, respectively) effects were reported that were consistent with a strong effect size but were not statistically significant (**Supplementary Table 10**).

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[71].

The Thr95Ile variant occurs in the third of six transmembrane domains of the SLC30A10 protein[72], the same domain affected by a previously reported loss-of-function variant causing HMNDY1 (hypermanganesemia with dystonia 1), Leu89Pro (rs281860284)[56] (**Figure 6**). In vitro, Leu89Pro abolishes trafficking of SLC30A10 to the membrane[56], and another study pointed to a functional role of polar or charged residues in the transmembrane domains of SLC30A10 for manganese transport function[73]. Bioinformatic analysis suggests 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, or otherwise affects SLC30A10 stability, translation, or transcription.

Comparison of *SLC30A10* Thr95Ile phenotypes to HMNDYT1 phenotypes

SLC30A10 (also known as *ZNT10*, and initially identified through sequence homology to zinc transporters[27]) encodes a cation diffusion facilitator expressed in hepatocytes, the bile duct epithelium, enterocytes, and neurons[32] that is essential for excretion of manganese from the liver into the bile and intestine[28,32]. 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[29-34,56]. 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)[74] and *SLC39A14* (solute carrier family 39 member 14, implicated in hypermanganesemia with dystonia 2)[75], along with experiments in transgenic mice [76,77], have confirmed the critical role of each of these genes in maintaining whole-body manganese homeostasis[78]. 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[78].

We detect two key aspects of HMNDYT1 – increased circulating liver enzymes and increased hematocrit – exceeding phenome-wide significance in heterozygous carriers of *SLC30A10* Thre95Ile. Among other hepatic phenotypes that have been reported in HMNDYT1 cases, we also detect an association with 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

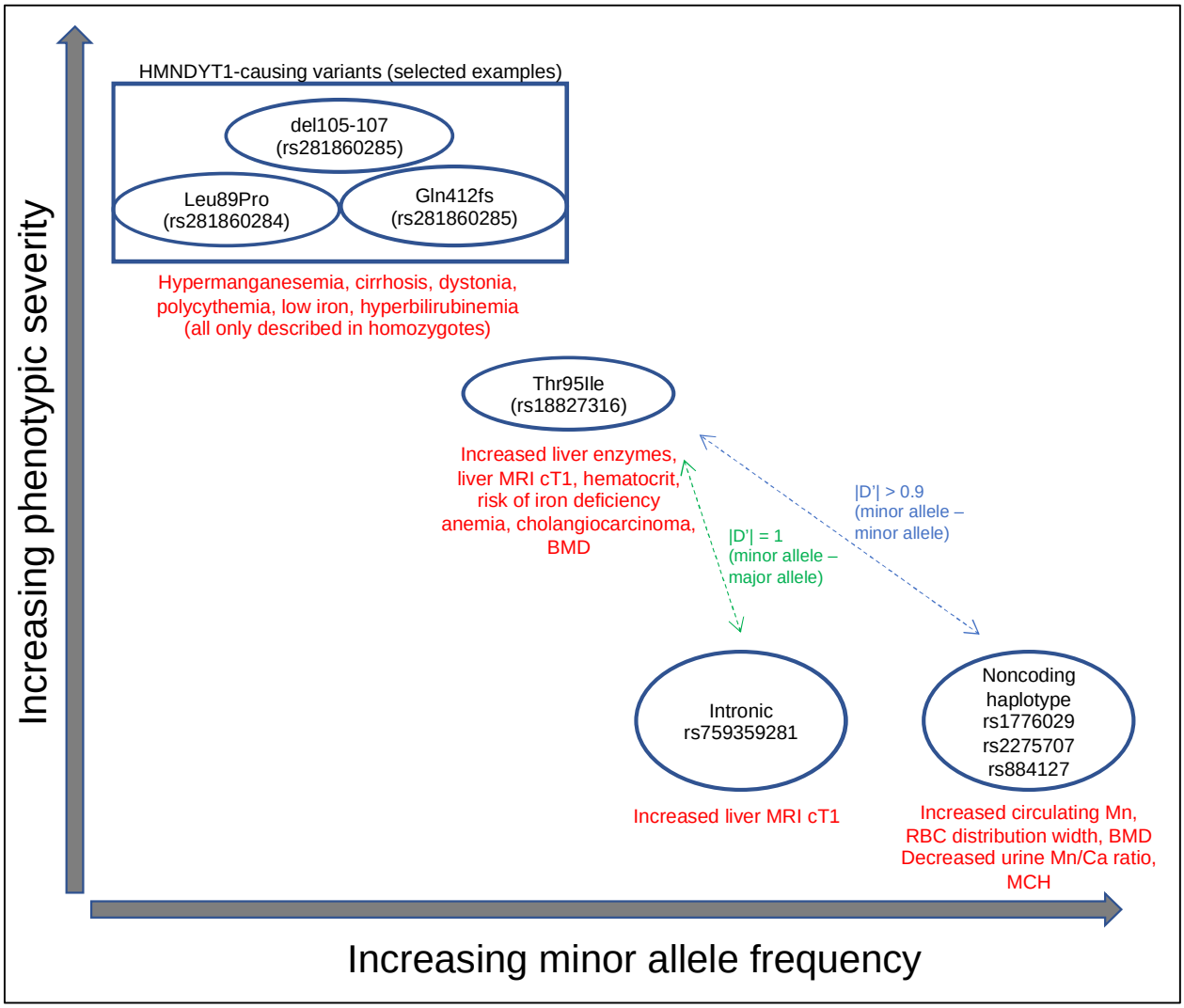
intriguing to consider that carrier status of Thr95Ile may represent a very mild manifestation of HMNDYT1.

The quantitative trait with the largest effect associated with *SLC30A10* Thre95Ile is liver MRI cT1 (+1.2 SD; 95% CI, +0.5 to +2.0; p = 0.0032). Liver MRI cT1 has been recently explored as a non-invasive diagnostic of steatohepatitis and fibrosis[79,80]. However, MRI T1 signal has also been used to detect manganese deposition in the brain, and it is unclear the extent to which hepatic manganese overload could confound the association of liver cT1 with liver damage[81].

Comparison of Thr95Ile phenotypes to *SLC30A10* common variant phenotypes

Apart from rare variants in *SLC30A10* causing HMNDYT1, Thr95Ile can also be compared to common variants in *SLC30A10* that have been associated with phenotypes by GWAS (**Figure 7**). We find that the minor allele of Thr95Ile is in almost complete linkage with a common intronic variant associated with increased blood manganese. Other GWAS variants 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[28]. Increased BMD may reflect the protective role of manganese in bone maintenance[82,83]. Looking for the subset of these phenotypes available in our scan of Thr95Ile, we do find a nominally significant increase in BMD but no detectable increase in MCH or erythrocyte distribution width. By contrast, we find that a common intronic variant in *SLC30A10* recently reported to associate with liver MRI cT1[50] is in complete linkage with the major allele of Thr95Ile, suggesting an independent genetic mechanism but also

480 providing independent evidence of the role of *SLC30A10* variants in liver health and/or hepatic
481 manganese content.
482



483

484 **Figure 7: Relationship of Mendelian and common-variant GWAS phenotypes at the *SLC30A10* locus.**
485 Phenotypes are summarized for HMNDYT1-causing variants, *SLC30A10* Thr95Ile, and GWAS variants.
486

487 The linked GWAS variants may be interpreted through two mechanistic hypotheses: first, the
488 associations may all be causally driven by Thr95Ile carriers in the studies, which the GWAS
489 variants tag; alternatively, the associations may be driven by effects of the common variants

490 themselves, which are noncoding but may influence *SLC30A10* (or another gene in cis) by
491 modulating expression or post-transcriptional regulation; or some combination of both. To
492 distinguish between these, measurements of Mn would need to be available to perform
493 conditional analyses. If the GWAS variants have an effect independent of Thr95Ile, *SLC30A10*
494 still seems likely (although not certain) to be the causal gene at the locus, due to the similarity in
495 phenotypes to HMNDYT1 and Thr95Ile. A putative regulatory mechanism could be through
496 transcriptional or post-transcriptional regulatory elements, as the haplotype includes a variant
497 (rs2275707) overlapping both the 3'-UTR of *SLC30A10* and H3K4me1 histone modifications
498 characteristic of enhancers active only in brain and liver[84].
499

500 ## Clinical relevance: manganese homeostasis in health and disease

501 Manganese (Mn) is a trace element required in the diet for normal development and function,
502 serving as a cofactor and regulator for many enzymes. However, it can be toxic in high doses;
503 because Mn(II) and Mn(III) can mimic other cations such as iron, it can interfere with systemic
504 iron homeostasis and disrupt in other biochemical processes[85,86]; at the cellular level, it is
505 cytotoxic and poisons the mitochondria by interfering with the electron transport chain enzymes
506 that rely on Fe-S clusters[87]. The hallmark of occupational exposure through inhalation is
507 neurotoxicity manifesting as parkinsonism and dystonia (manganism, or Mn intoxication)[85,86].
508 Neurotoxicity is an aspect of the Mendelian syndromes caused by loss of function of all three of
509 the hepatic manganese transporters; interestingly, GWAS has also identified a common missense
510 variant in *SLC39A8* as a risk factor for schizophrenia and many other diseases[88,89]; altered
511 function of glycosyltransferases due to manganese overload in the neurons is a proposed
512 mechanism for neurological manifestations of this variant[90]. Because manganese is excreted

513 through the liver into the bile, increased circulating manganese secondary to liver damage may
514 be a contributing factor to the neurological manifestations of chronic acquired hepatocerebral
515 degeneration (CAHD)[91-93]. However, liver toxicity is not a hallmark of environmental or
516 occupational exposure. Importantly, of the Mendelian syndromes of genes encoding manganese
517 transporters, only *SLC30A10* (causing HMNDYT1) involves hepatic symptoms[78,94].
518 Hepatotoxicity in HMNDYT1 is thought to be due to cytotoxic manganese overload within
519 hepatocytes; polycythemia is thought to be caused by upregulation of erythropoietin by
520 manganese; and iron anemia through systemic dysregulation of iron homeostasis by excess
521 manganese[94,95]. Our results suggest that polymorphism in *SLC30A10* is a risk factor for
522 manganese-induced hepatocellular damage, polycythemia, and iron anemia in a much broader
523 population beyond the rare recessive syndrome HMNDYT1.
524
525 The association of *SLC30A10* Thr95Ile with extrahepatic bile duct cancer was unexpected, as
526 this disease has not been described in conjunction with HMNDYT1. Bile duct cancer
527 (cholangiocarcinoma) is a rare disease (age-adjusted incidence of 1 – 3 per 100,000 per year);
528 cirrhosis, viral hepatitis, primary sclerosing cholangitis, and parasitic fluke infection have been
529 identified as risk factors[96,97]. It is unclear whether low levels of manganese in the bile, or high
530 levels of manganese in the hepatocytes and bile duct epithelial cholangiocytes, could be directly
531 carcinogenic; manganese-dependent superoxide dismutase (MnSOD, or SOD2) is a tumor
532 suppressor[98]. A simpler possibility is that cytotoxic manganese overload in hepatocytes and
533 cholangiocytes causes localized inflammation that predisposes to cancer through similar
534 mechanisms as other hepatobiliary risk factors. We do detect an association with cholangitis, but
535 the effect of this association is weaker than the association with cholangiocarcinoma. To our

knowledge, *SLC30A10* Thr95Ile would be the strongest genetic cholangiocarcinoma risk factor identified to date, being carried by 5% of the extrahepatic bile duct cancer cases in the White British subset of the biobank. 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[99]. While the appropriate clinical management of carriers of *SLC30A10* Thr95Ile is unclear and would require further studies to determine whether monitoring of hepatobiliary function is warranted, evidence from HMNDYT1 patients has demonstrated that chelation therapy combined with iron supplementation is effective at reversing the symptoms of SLC30A10 insufficiency[100]. 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, 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 experienced homozygous loss-of-function mutation.

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

559 gap between common and rare disease genetics[101,102]. These discoveries are possible only by
560 combining massive, biobank-scale genotype and phenotype datasets such as the UK Biobank.

561 # Methods

562 ## Sub-population definition and PC calculation

563 Sub-populations for analysis were obtained through a combination of self-reported ethnicity and
564 genetic principal components. First, the White British population was defined using the
565 categorization performed previously by the UK Biobank[103] (Field 22006 value "Caucasian");
566 briefly, this analysis selected the individuals who identify as White British (Field 21000),
567 performed a series of subject-level QC steps (to remove subjects with high heterozygosity or
568 missing rate over 5%, removing subjects with genetic and self-reported sex discrepancies and
569 putative sex chromosome aneuploidies, and removing subjects with second or first degree
570 relatives and an excess of third degree relatives), performed Bayesian outlier detection using the
571 R package aberrant [104] to remove ancestry outliers using principal components (PCs) 1+2, 3+4,
572 and 5+6 (calculated from the global UK Biobank PCs stored in Field 22009), selected a subset of
573 variants in preparation for PCA by limiting to directly-genotyped (info = 1), missingness across
574 individuals < 2%, MAF > 1%, regions of known long range LD, and pruning to independent
575 markers with pairwise LD < 0.1. Based on this procedure used by the UK Biobank to define the
576 "White British" subset, we defined three additional populations, using other self-reported
577 ancestry groups as starting points (Field 21000 values "Asian or Asian British", "Black or Black
578 British", and "Chinese"). Principal components were estimated in PLINK using the unrelated
579 subjects in each subgroup. We then projected all subjects onto the PCs. For the majority of
580 downstream analyses (calculation of per-variant allele frequency and missingness thresholds,
581 calculation of LD, and for association analyses performed in PLINK), just the unrelated subset of

582 people in the subpopulation was used, the unrelated sets were used. The exception was
583 association analyses performed in SAIGE[105], a generalized mixed model method that allows
584 inclusion of related individuals; for SAIGE, related individuals were retained in the
585 subpopulations.
586
587 For validation in an independent subpopulation of the UK Biobank, two other self-reported
588 ethnicity groups with a sufficient number of *SLC30A10* Thr95Ile carriers were assembled, who
589 were not included in "White British" (Field 21000 values "White" subgroup "Irish", and "White"
590 subgroup "Any other white background" or no reported subgroup).
591

592 ## Array genotype data for association analysis

593 Data were obtained from the UK Biobank through application 26041. Genotypes were obtained
594 through array typing and imputation as described previously[103]. For genome-wide association
595 analysis, variants were filtered so that imputation quality score (INFO) was greater than 0.8.
596 Genotype missingness, Hardy-Weinberg equilibrium (HWE), and minor allele frequency (MAF)
597 were then each calculated across the unrelated subset of individuals in each of the four sub-
598 populations. For each sub-population a set of variants for GWAS was then defined by filtering
599 missingness across individuals less than 2%, HWE p-value > 10^{-12}, and MAF > 0.1%.
600

601 ## Phenotype data

602 For genome-wide analysis, blood biochemistry values were obtained for ALT (Field 30620) and
603 AST (Field 30650) and $\log_{10}$ transformed, consistent with previous genetic studies[14,106], resulting
604 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 general practitioner (GP) clinical event records (Field 42040). A selection of 135 quantitative traits was obtained from other fields (**Supplementary Table 14**), encompassing anthropomorphic measurements, blood and urine biochemistry, smoking, exercise, sleep behavior, and liver MRI; all were inverse rank normalized using the RNOmni R package[107]. All quantitative traits and cancer registry diagnoses were downloaded from the UK Biobank Data Showcase on March 17, 2020. The GP clinical events, inpatient diagnoses, and death registry were available in more detail or in more recent updates than was available through the Data Showcase and were downloaded as separate tables; data for GP clinical records were downloaded on March 30, 2019, data from the death registry was downloaded on June 12, 2020, and data from hospital diagnoses was downloaded on July 15, 2020.

Genome-wide association studies of ALT and AST

Because of the high level of relatedness in the UK Biobank participants[108], to maximize power by retaining related individuals we used SAIGE software package[105] 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[109]. The association

results for each enzyme were meta-analyzed across the four populations using the METAL software package[110] 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[111], the PLINK software package v1.90[112] 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. All PLINK and SAIGE association tests were performed using the REVEAL/SciDB translational analytics platform from Paradigm4.

Identifying independent, linked association signals between the two GWAS

Meta-analysis results for each enzyme were LD clumped using the PLINK software package, v1.90[112] 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 variants (the variants 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 variants 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 variant between the two GWAS were treated as potentially colocalized association signals between the two GWAS. In these cases, a representative index variant was chosen to represent the results of both GWAS by choosing the index variant of the GWAS with the more significant p-value. Next, these putative colocalized

association signals were then distance pruned by iteratively removing neighboring index variants within 500 kilobases of each index variant 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[113] .

Annotation of associated loci and variants

Index variants 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[114], impact on protein-coding genes using the ENSEMBL Variant Effect Predictor (VEP) software package[115] with the LOFTEE plugin to additionally predict protein-truncating variants[116]; 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[117]; a published meta-analysis of four liver eQTL studies[41]; the eQTLGen meta-analysis of blood eQTL studies[118]; and GWAS results from the NHGRI-EBI GWAS Catalog (r2020-01-27)[119], filtered to associations with $p < 5 \times 10^{-8}$.

Association of ALT- and AST-associated loci with liver disease

Index variants were tested for association with any liver disease using ICD10 codes K70-K77 in inpatient hospital diagnoses, causes of death, and GP clinical event records, using SAIGE, with the same covariates used for the liver enzymes (age, sex, and genetic PCs 1-12) plus a covariate for each of the following: whether the subject was recruited in Scotland, whether the subject was recruited in Wales, and whether the patient had GP clinical event records available. Association results were meta-analyzed across the four sub-populations using METAL using the default

674 method (combining p-values) to obtain the final p-value. To obtain effect sizes and standard
675 errors, the same procedure was performed but using PLINK (on the unrelated subset of each
676 population) and using the classical method in METAL (combining effects and standard errors.)
677
678 ## Sequencing-based validation of rs188273166 array genotyping
679 Whole exome sequencing was available for 301,473 of the 487,327 array-genotyped samples.
680 DNA was extracted from whole blood and was prepared and sequenced by the Regeneron
681 Genetics Center (RGC). A complete protocol has been described elsewhere[120]. Briefly, the xGen
682 exome capture was used and reads were sequenced using the Illumina NovaSeq 6000 platform.
683 Reads were aligned to the GRCh38 reference genome using BWA-mem[121]. Duplicate reads were
684 identified and excluded using the Picard MarkDuplicates tool (Broad Institute)[122]. Variant calling
685 of SNVs and indels was done using the WeCall variant caller (Genomics Plc.)[123] to produce a
686 GVCF file for each subject (GVCF files are files in the VCF Variant Call Format that are
687 indexed for fast processing). GVCF files were combined to using the GLnexus joint calling
688 tool[124]. Post-variant calling filtering was applied as described previously[120].
689
690 ## Replication of *SLC30A10* Thr95Ile associations
691 ALT and AST association tests were repeated as described for the genome-wide scans, using
692 SAIGE and PLINK, in the “Other White” and “White Irish” populations, for the *SLC30A10*
693 Thr95Ile (rs188273166) variant. In the two DiscovEHR Geisinger Health Service (GHS) cohorts,
694 association tests were performed using BOLT[125] with covariates for age, age squared, age x sex,
695 sex, and the first ten principal components of genetic ancestry; ALT, AST, and hematocrit values

were taken from the median of lab values available. Results were meta-analyzed across the four populations. A forest plot was created using the forestplot package in R[126].

Testing linkage of *SLC30A10* Thr95Ile to common GWAS variants

To test linkage of *SLC30A10* Thr95Ile (rs188273166) to common GWAS variants, the GWAS Catalog was searched for all results where "Mapped Gene" was assigned to *SLC30A10*; because of the very relevant phenotype, blood Mn-associated variant rs1776029, an association that is not in the GWAS Catalog, was also included in the analysis, as well at cT1-associated variant rs759359281. LD calculations were performed in PLINK, using the White British unrelated subpopulation, between rs188273166 and the GWAS variants 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[46] 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 variant from GWAS Atlas were used to determine the direction of effect and cited accordingly (**Supplementary Table 13**). Reference epigenome information for the GWAS variants was obtained by searching for rs1776029 in HaploReg v4.1[127].

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 using the covariates 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, GP availability, or country of recruitment covariates: at least 100 diagnoses overall, and at least one diagnosis in each of the following subgroups, to avoid collinearity with covariates while running SAIGE: the with- and without-GP data subgroups, men, women, and each of the three recruitment countries. This resulted in 4,397 ICD10 codes to test, serving as the multiple hypothesis burden.

Bioinformatic analysis of *SLC30A10* Thr95Ile

To visualize Thr95Ile on the protein sequence of SLC30A10, UNIPROT entry Q6XR72 (ZN10_HUMAN) was accessed[72]. In UNIPROT, natural variants causing HMNDYT1[32,34,56] and mutagenesis results[56,73,128] were collated from the literature and highlighted. CADD score v1.5[129] was downloaded from the authors' website. SIFT score was obtained from the authors' website using the "dbSNP rsIDs (SIFT4G predictions)" tool[130]. PolyPhen score and multiple species alignment was obtained from the authors' website using the PolyPhen-2 tool[131].

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 dPBS (Gibco, Grand Island, NY), fixed with 4% paraformaldehyde (in water) (Electron Microscopy Sciences, Hatfield, PA) for 10 min, rinsed in 4°C PBS (Invitrogen, Vilnius, LT), 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 dPBS, 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.

Author Contributions

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

Competing Interests

A.D., A.F.C., S.T., L.W., M.P., P.N., C.Q., H.C.T., G.H., and P.H. are employees of Alnylam Pharmaceuticals, Inc. L.L., N.V., M.F., and A.B. are employees of Regeneron Pharmaceuticals, Inc.

Ethical Compliance

Ethics oversight for the UK Biobank is provided by an Ethics and Governance Council which obtained informed consent from all participants for health-related research. All research described was performed within the framework of Application 26041.

Data Availability

Complete summary statistics from the genome-wide association studies of ALT, AST, and extrahepatic bile duct cancer have been submitted to the NHGRI-EBI GWAS catalog (https://www.ebi.ac.uk/gwas/).

References

1 Asrani, S. K., Devarbhavi, H., Eaton, J. & Kamath, P. S. Burden of liver diseases in the world. *J Hepatol* **70**, 151-171, doi:10.1016/j.jhep.2018.09.014 (2019).

2 Younossi, Z. M. *et al.* Changes in the prevalence of the most common causes of chronic liver diseases in the United States from 1988 to 2008. *Clin Gastroenterol Hepatol* **9**, 524-530 e521; quiz e560, doi:10.1016/j.cgh.2011.03.020 (2011).

3 Plenge, R. M., Scolnick, E. M. & Altshuler, D. Validating therapeutic targets through human genetics. *Nat Rev Drug Discov* **12**, 581-594, doi:10.1038/nrd4051 (2013).

4 Stevens, J. L. & Baker, T. K. The future of drug safety testing: expanding the view and narrowing the focus. *Drug Discov Today* **14**, 162-167, doi:10.1016/j.drudis.2008.11.009 (2009).

5 Deaton, A. M. *et al.* Rationalizing Secondary Pharmacology Screening Using Human Genetic and Pharmacological Evidence. *Toxicol Sci* **167**, 593-603, doi:10.1093/toxsci/kfy265 (2019).

6 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).

7 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).

8 Rahmioglu, N. *et al.* Epidemiology and genetic epidemiology of the liver function test proteins. *PloS one* **4**, e4435, doi:10.1371/journal.pone.0004435 (2009).

9 Pilia, G. *et al.* Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS genetics* **2**, e132, doi:10.1371/journal.pgen.0020132 (2006).

10 Makkonen, J., Pietilainen, K. H., Rissanen, A., Kaprio, J. & Yki-Jarvinen, H. Genetic factors contribute to variation in serum alanine aminotransferase activity independent of obesity and alcohol: a study in monozygotic and dizygotic twins. *J Hepatol* **50**, 1035-1042, doi:10.1016/j.jhep.2008.12.025 (2009).

11 Nilsson, S. E., Read, S., Berg, S. & Johansson, B. Heritabilities for fifteen routine biochemical values: findings in 215 Swedish twin pairs 82 years of age or older. *Scand J Clin Lab Invest* **69**, 562-569, doi:10.1080/00365510902814646 (2009).

12 Bathum, L. *et al.* Evidence for a substantial genetic influence on biochemical liver function tests: results from a population-based Danish twin study. *Clin Chem* **47**, 81-87 (2001).

13 Targher, G. Elevated serum gamma-glutamyltransferase activity is associated with increased risk of mortality, incident type 2 diabetes, cardiovascular events, chronic kidney disease and cancer - a narrative review. *Clin Chem Lab Med* **48**, 147-157, doi:10.1515/CCLM.2010.031 (2010).

14 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).

15 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).

16 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).

17 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).

18 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).

19 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).

20 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).

21 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).

22 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).

23 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).

24 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).

25 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).

26 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).

27 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).

28 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).

29 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).

30 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).

31 Lechpammer, M. *et al.* Pathology of inherited manganese transporter deficiency. *Ann Neurol* **75**, 608-612, doi:10.1002/ana.24131 (2014).

32 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).

33 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).

34 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).

35 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).

36 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).

37 Dewey, F. E. *et al.* Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science* **354**, doi:10.1126/science.aaf6814 (2016).

38 Partnership, A. M. *Common Metabolic Diseases Knowledge Portal*, <http://hugeamp.org/> (

39 Emdin, C. A. *et al.* A missense variant in Mitochondrial Amidoxime Reducing Component 1 gene and protection against liver disease. *PLoS genetics* **16**, e1008629, doi:10.1371/journal.pgen.1008629 (2020).

40 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).

41 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).

42 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).

43 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).

44 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).

45 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).

46 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).

47 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).

48 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).

49 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).

50 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).

51 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).

52 Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *bioRxiv*, 787903, doi:10.1101/787903 (2019).

53 Ramachandran, P. *et al.* Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature* **575**, 512-518, doi:10.1038/s41586-019-1631-3 (2019).

54 MacParland, S. A. *et al.* Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat Commun* **9**, 4383, doi:10.1038/s41467-018-06318-7 (2018).

55 Aizarani, N. *et al.* A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature* **572**, 199-204, doi:10.1038/s41586-019-1373-2 (2019).

56 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).

57 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).

58 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).

59 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).

60 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).

61 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).

62 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).

63 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).

64 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).

65 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).

66 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).

67 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).

68 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).

69 Hayes, M. G. *et al.* Identification of HKDC1 and BACE2 as genes influencing glycemic traits during pregnancy through genome-wide association studies. *Diabetes* **62**, 3282-3291, doi:10.2337/db12-1692 (2013).

70 Guo, C. *et al.* Coordinated regulatory variation associated with gestational hyperglycaemia regulates expression of the novel hexokinase HKDC1. *Nat Commun* **6**, 6069, doi:10.1038/ncomms7069 (2015).

71 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).

72 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).

73 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).

74 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).

75 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).

76 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).

77 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).

78 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).

79 Pavlides, M. *et al.* Multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease severity. *Liver Int* **37**, 1065-1073, doi:10.1111/liv.13284 (2017).

80 Pavlides, M. *et al.* Multiparametric magnetic resonance imaging predicts clinical outcomes in patients with chronic liver disease. *J Hepatol* **64**, 308-315, doi:10.1016/j.jhep.2015.10.009 (2016).

81 Kim, Y. High signal intensities on T1-weighted MRI as a biomarker of exposure to manganese. *Ind Health* **42**, 111-115, doi:10.2486/indhealth.42.111 (2004).

82 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).

83 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).

84 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).

85 Crossgrove, J. & Zheng, W. Manganese toxicity upon overexposure. *NMR Biomed* **17**, 544-553, doi:10.1002/nbm.931 (2004).

86 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).

87 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).

88 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).

89 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).

90 Mealer, R. G. *et al.* A schizophrenia risk locus alters brain metal transport and plasma glycosylation. *bioRxiv*, 757088, doi:10.1101/757088 (2019).

91 Krieger, D. *et al.* Manganese and chronic hepatic encephalopathy. *Lancet* **346**, 270-274, doi:10.1016/s0140-6736(95)92164-8 (1995).

92 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).

93 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).

94 Anagianni, S. & Tuschl, K. Genetic Disorders of Manganese Metabolism. *Curr Neurol Neurosci Rep* **19**, 33, doi:10.1007/s11910-019-0942-y (2019).

95 Ebert, B. L. & Bunn, H. F. Regulation of the erythropoietin gene. *Blood* **94**, 1864-1877 (1999).

96 Tyson, G. L. & El-Serag, H. B. Risk factors for cholangiocarcinoma. *Hepatology* **54**, 173-184, doi:10.1002/hep.24351 (2011).

97 Razumilava, N. & Gores, G. J. Cholangiocarcinoma. *Lancet* **383**, 2168-2179, doi:10.1016/S0140-6736(13)61903-0 (2014).

98 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).

99 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).

100 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).

101 Bastarache, L. *et al.* Phenotype risk scores identify patients with unrecognized Mendelian disease patterns. *Science* **359**, 1233-1239, doi:10.1126/science.aal4043 (2018).

102 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).

103 Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv*, 166298, doi:10.1101/166298 (2017).

104 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).

105 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).

106 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).

107 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).

108 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).

109 Turner, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv*, 005165, doi:10.1101/005165 (2014).

110 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).

111 Zhou, W. (2018).

112 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).

113 Yin, L. *A high-quality drawing tool designed for Manhattan plot of genomic analysis*, <https://github.com/YinLiLin/R-CMplot> (2018).

114 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).

115 McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122, doi:10.1186/s13059-016-0974-4 (2016).

116 Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443, doi:10.1038/s41586-020-2308-7 (2020).

117 Consortium, G. T. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213, doi:10.1038/nature24277 (2017).

118 Võsa, U. *et al.* Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv*, 447367, doi:10.1101/447367 (2018).

119 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).

120 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).

121 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

122 *Picard*, <http://broadinstitute.github.io/picard/> (

123 *WeCall*, <https://github.com/Genomicsplc/wecall> (

124 Lin, M. F. *et al.* GLnexus: joint variant calling for large cohort sequencing. *bioRxiv*, 343970, doi:10.1101/343970 (2018).

125 Loh, P. R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nature genetics* **50**, 906-908, doi:10.1038/s41588-018-0144-6 (2018).

126 Gordon, M. & Lumley, T. forestplot: Advanced Forest Plot Using'grid'Graphics. *R package version* **1** (2015).

127 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).

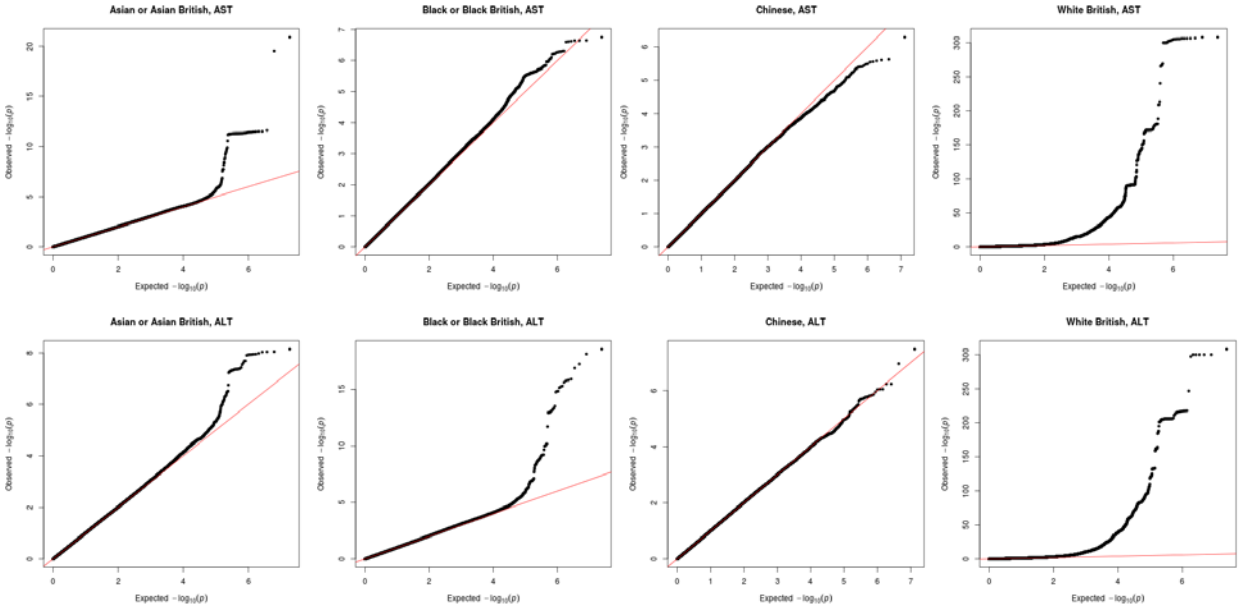
128 Zhao, Y., Feresin, R. G., Falcon-Perez, J. M. & Salazar, G. Differential Targeting of SLC30A10/ZnT10 Heterodimers to Endolysosomal Compartments Modulates EGF-Induced MEK/ERK1/2 Activity. *Traffic* **17**, 267-288, doi:10.1111/tra.12371 (2016).

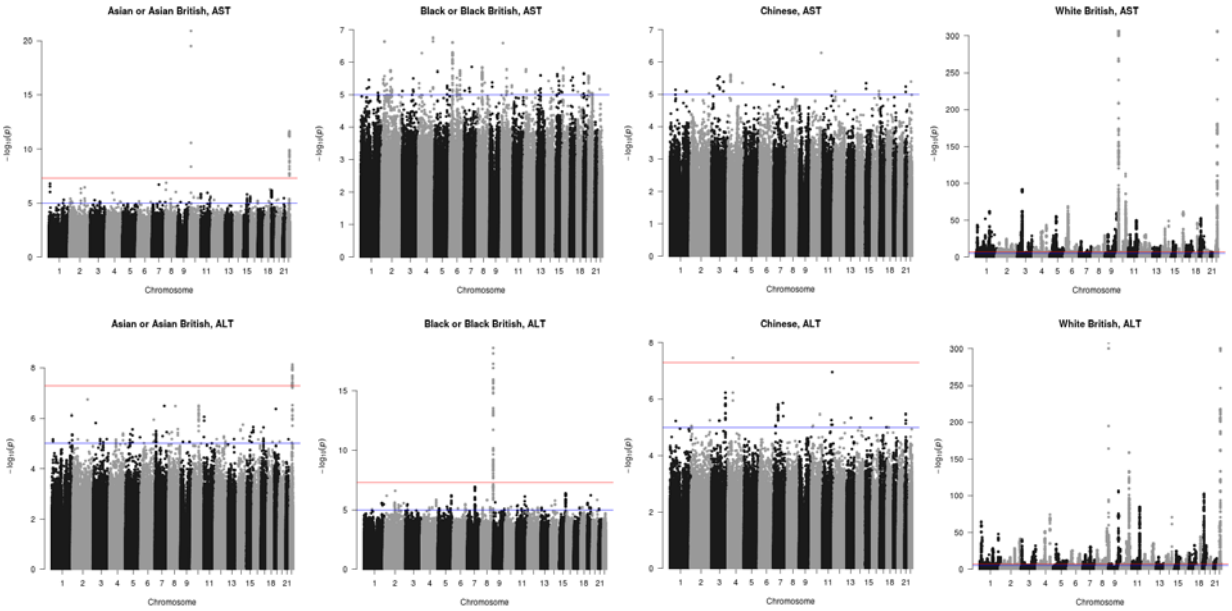
129 Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886-D894, doi:10.1093/nar/gky1016 (2019).

130 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).

131 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

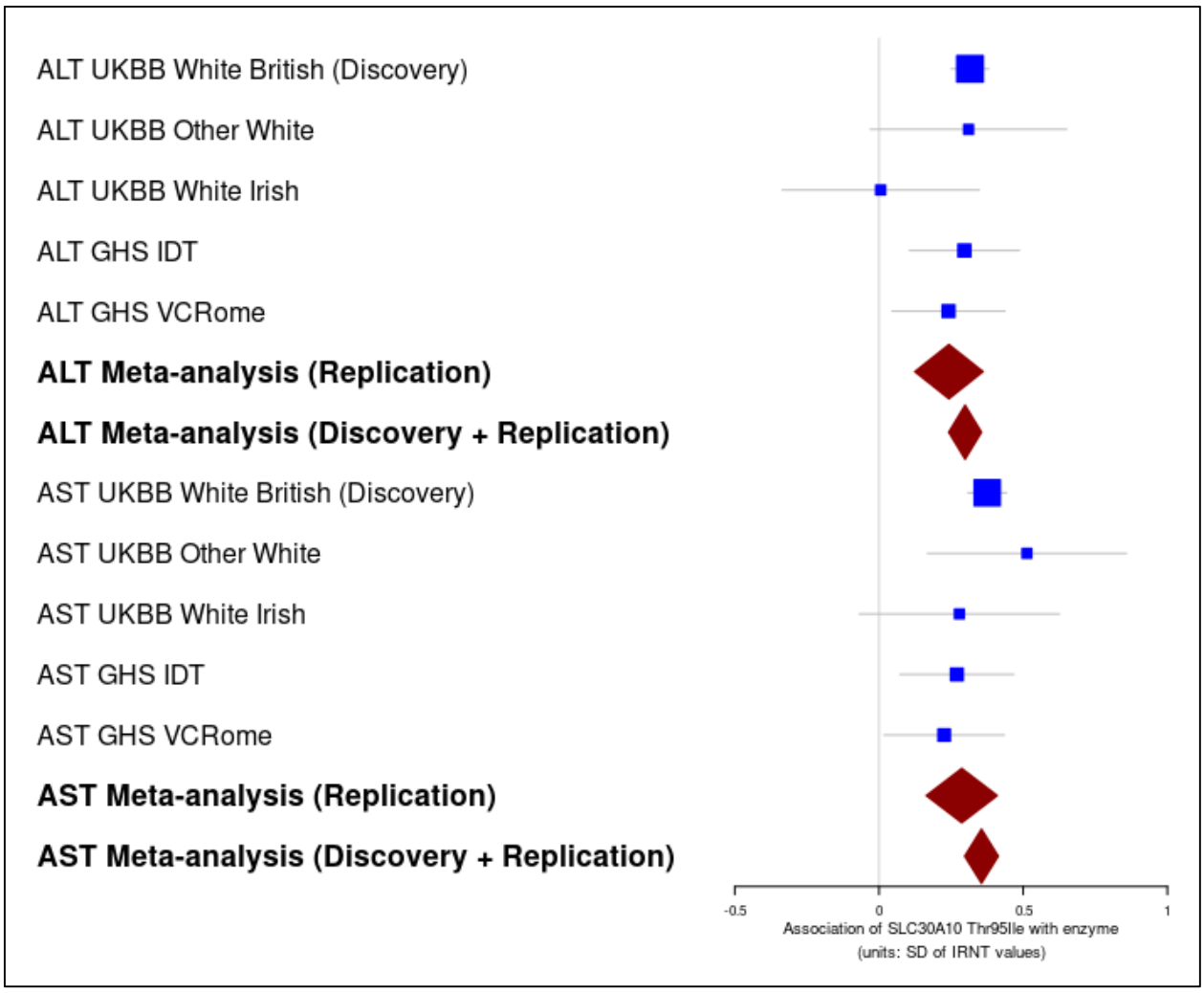




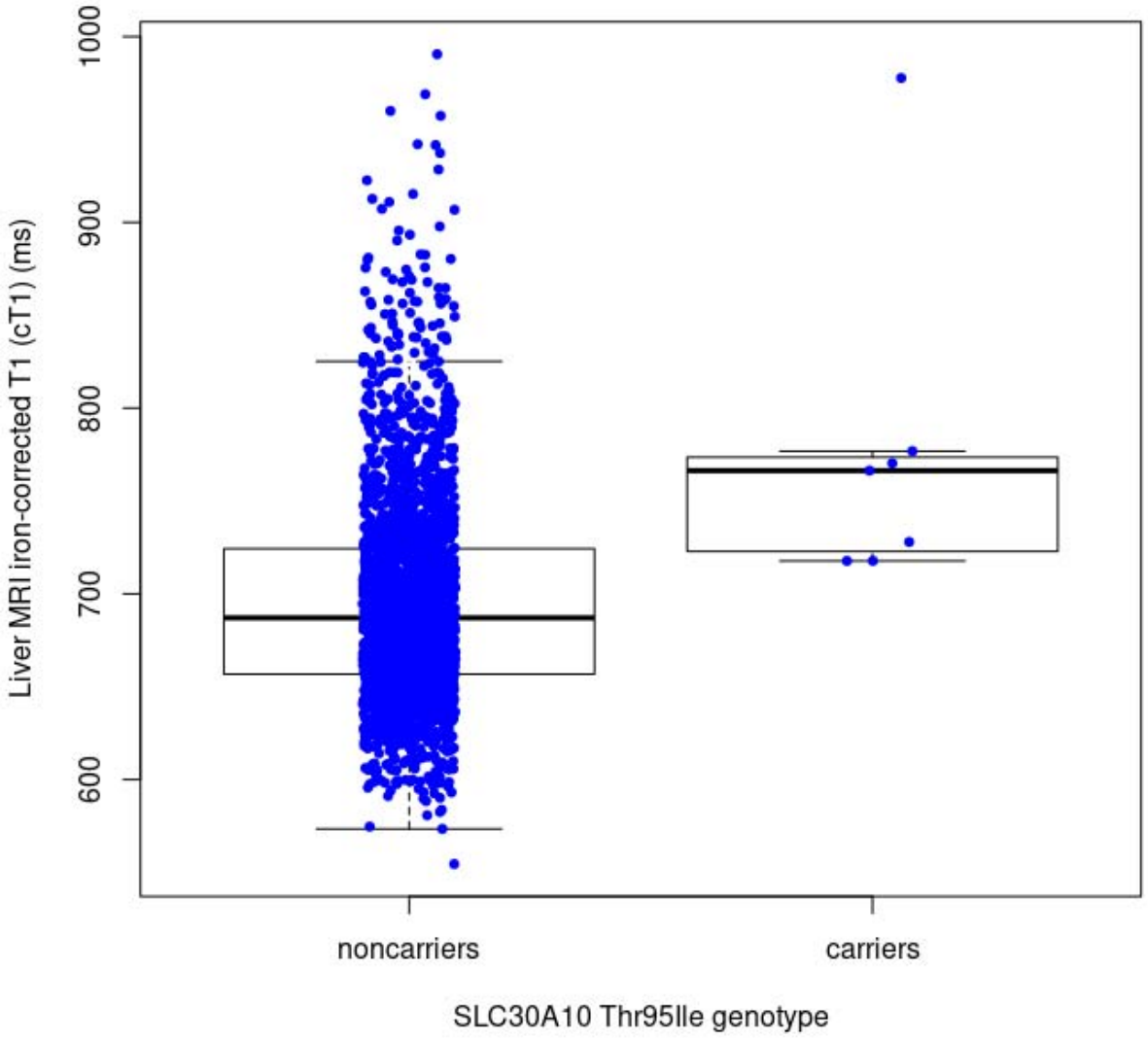
1129

1130

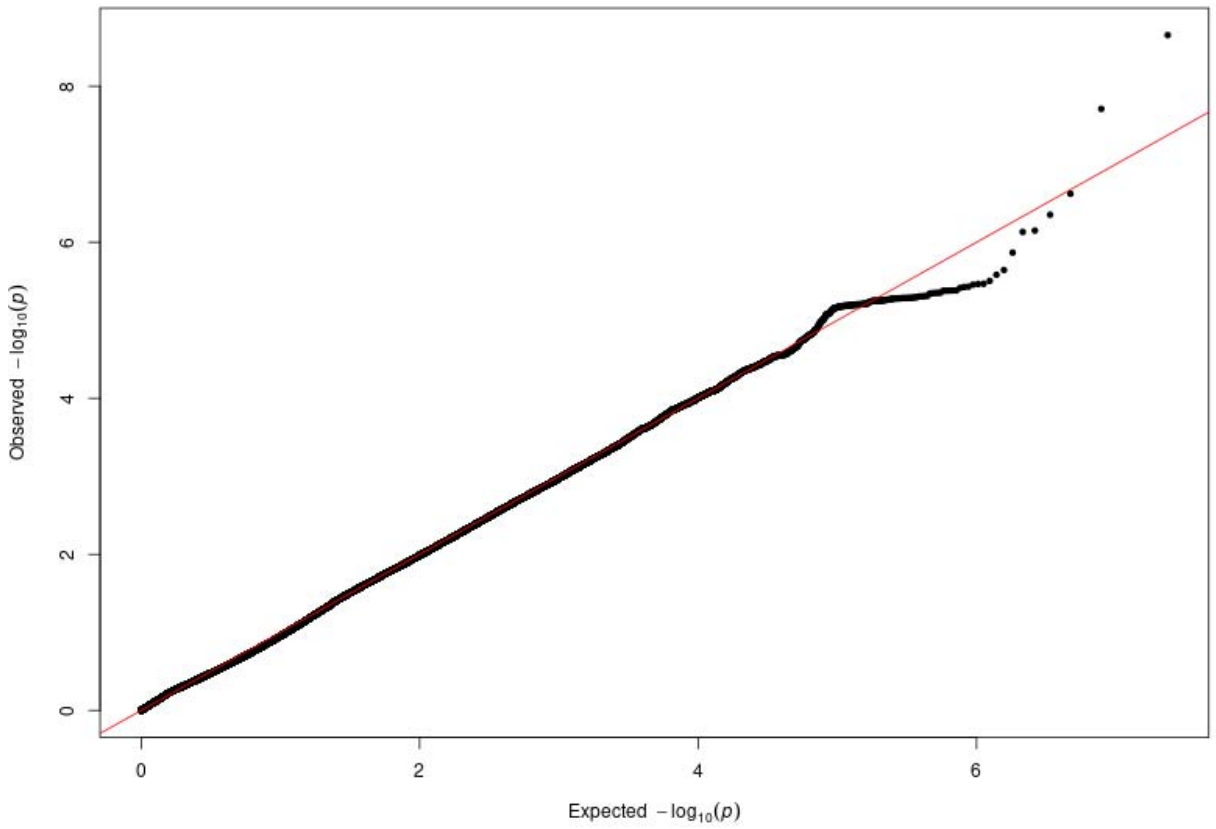
1131



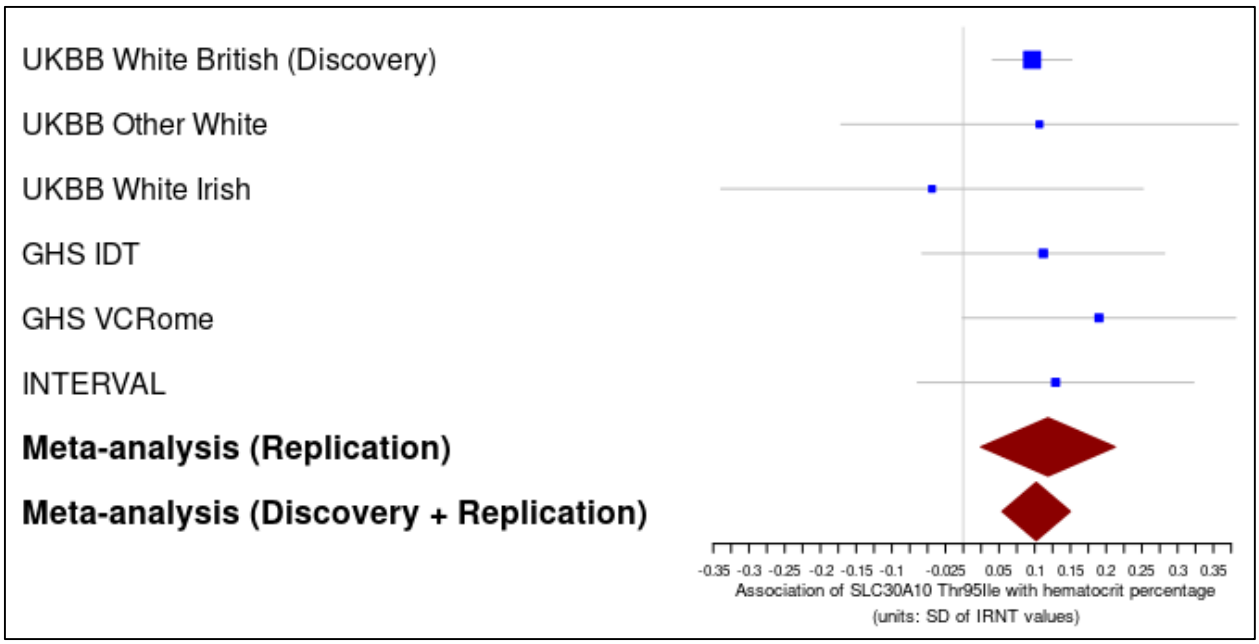
Supplementary Figure 3: Forest plot of *SLC30A10* Thr95Ile (rs188273166) association with ALT and AST in the discovery and replication groups. Boxes show effect size estimates from PLINK, and error bars show 95% confidence intervals.



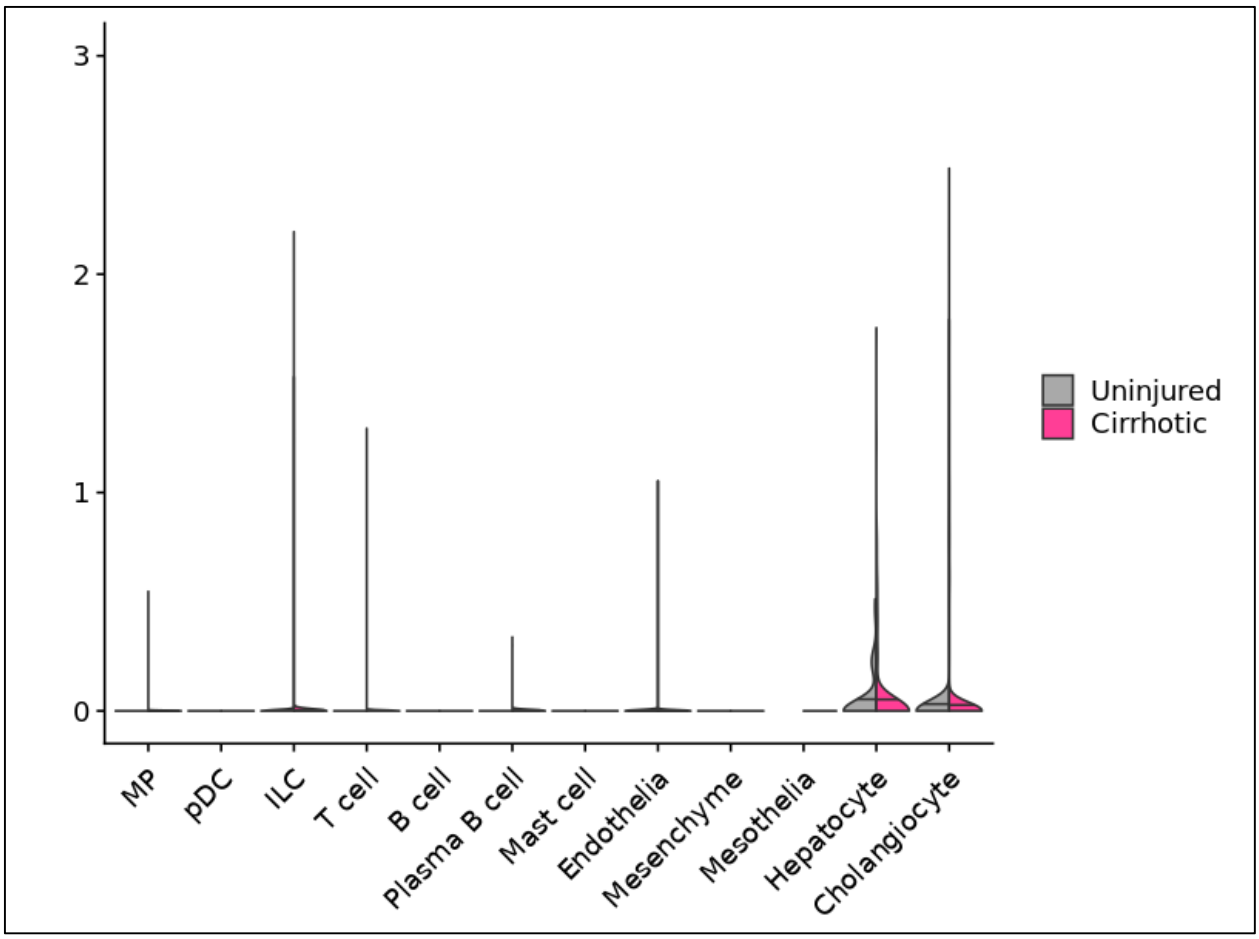
Supplementary Figure 4: iron-corrected T1 (cT1) values from liver MRI in seven SLC30A10 Thr95Ile carriers (limited to the White British population).



Supplementary Figure 5: Q-Q plot of GWAS (p value from SAIGE) for ICD10 diagnosis C24.0, extrahepatic bile duct cancer.

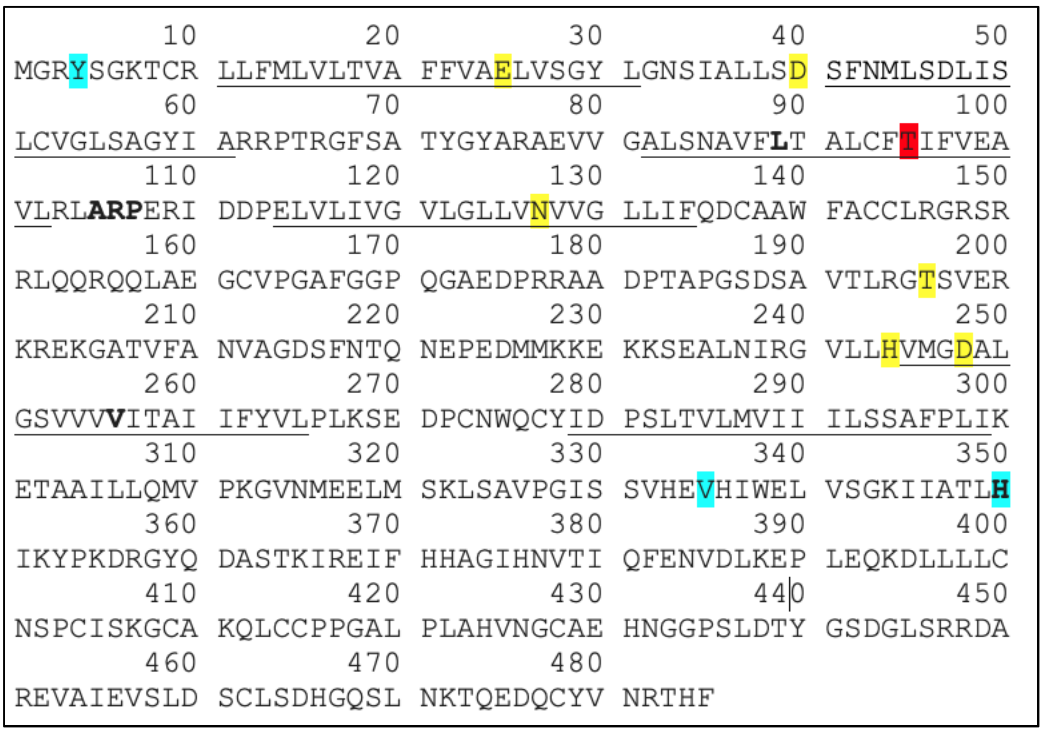


Supplementary Figure 6: Forest plot of *SLC30A10* Thr95Ile (rs188273166) association with ALT and AST in the discovery and replication groups. Boxes show effect size estimates from PLINK, and error bars show 95% confidence intervals.



Supplementary Figure 7: Single-cell expression data, showing distribution of expression values for SLC30A10 from Ramachandran et al.[53] Produced by https://www.livercellatlas.mvm.ed.ac.uk/.

*



Supplementary Figure 8: Protein sequence of SLC30A10 along with experimental evidence cited by UNIPROT[72]. 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[56,73,128]. In bold are variants known to cause SLC30A10 deficiency (HMNDYT1)[32,34,56].