nature genetics

Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma

Concentrations of liver enzymes in plasma are widely used as indicators of liver disease. We carried out a genome-wide association study in 61,089 individuals, identifying 42 loci associated with concentrations of liver enzymes in plasma, of which 32 are new associations ($P = 10^{-8}$ to $P = 10^{-190}$). We used functional genomic approaches including metabonomic profiling and gene expression analyses to identify probable candidate genes at these regions. We identified 69 candidate genes, including genes involved in biliary transport (*ATP8B1* and *ABCB11*), glucose, carbohydrate and lipid metabolism (*FADS1*, *FADS2*, *GCKR*, *JMJD1C*, *HNF1A*, *MLXIPL*, *PNPLA3*, *PPP1R3B*, *SLC2A2* and *TRIB1*), glycoprotein biosynthesis and cell surface glycobiology (*ABO*, *ASGR1*, *FUT2*, *GPLD1* and *ST3GAL4*), inflammation and immunity (*CD276*, *CDH6*, *GCKR*, *HNF1A*, *HPR*, *ITGA1*, *RORA* and *STAT4*) and glutathione metabolism (*GSTT1*, *GSTT2* and *GGT*), as well as several genes of uncertain or unknown function (including *ABHD12*, *EFHD1*, *EFNA1*, *EPHA2*, *MICAL3* and *ZNF827*). Our results provide new insight into genetic mechanisms and pathways influencing markers of liver function.

High concentrations of liver enzymes in plasma are observed in liver injury caused by multiple insults including alcohol misuse, viral and other infections, metabolic disorders, obesity, autoimmune disease and drug toxicity[1]. High liver enzyme concentrations are associated with increased risk of cirrhosis[2], hepatocellular carcinoma[3], type 2 diabetes[4] and cardiovascular disease[5]. Abnormal liver function is a common reason for terminating new clinical therapeutic agents, representing a major challenge for the global pharmaceutical industry[6]. Liver enzyme concentrations in plasma are highly heritable[7], suggesting an important role for genetic factors.

We carried out a genome-wide association study (GWAS) in 61,089 research participants to identify genetic loci influencing liver function measured by concentrations of alanine transaminase (ALT), alkaline phosphatase (ALP) and γ-glutamyl transferase (GGT) in blood. ALT is mainly a marker of hepatocellular damage[1], and may also be high in obesity and fatty liver disease[8]. ALP is a marker of biliary obstruction, and is also released from bone, intestine, leucocytes and other cells[1]. GGT is sensitive to most kinds of liver insult, particularly alcohol[1]. Our study design is summarized in **Figure 1**. Characteristics of participants, genotyping arrays and quality control measures are summarized in **Supplementary Tables 1–4**. Genome-wide significance was inferred at $P < 1 \times 10^{-8}$, allowing a Bonferroni correction for ~10^6 independent SNPs tested[9], and for three separate liver markers; the latter is a conservative adjustment given the correlations between concentrations of the three liver markers (r = 0.19–0.64) and their association test results (r = 0.02–0.19; **Supplementary Table 5**).

We found 1,304 SNPs associated with one or more liver markers at $P < 1 \times 10^{-7}$ across 42 genetic loci (**Table 1** and **Fig. 2**). At 35 of these loci, one or more SNPs reached genome-wide significance ($P < 1 \times 10^{-8}$; **Supplementary Table 6**); at the other seven genetic loci, the top-ranking SNP reached genome-wide significance after further testing in an additional sample of 12,139 research participants (**Supplementary Table 7**). Regional plots for each of the genetic loci are shown in **Supplementary Figures 1–3**. Common variants at chromosome 8q24 were associated with both ALP and ALT, and variants at chromosome 19q13 were associated with both ALP and GGT, at $P < 1 \times 10^{-8}$. Sixteen loci associated with one liver marker at $P < 10^{-8}$ showed additional associations with a second

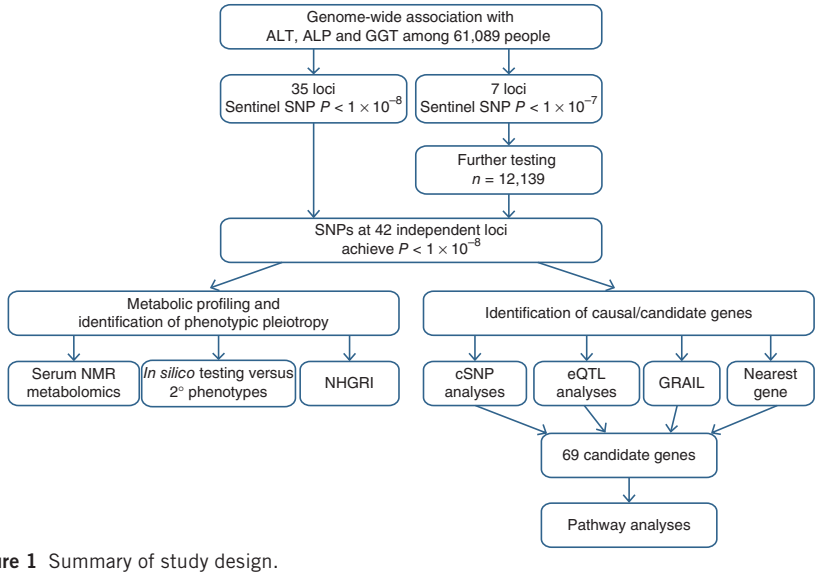

Figure 1 Summary of study design.

A full list of authors and affiliations appears at the end of the paper.

Received 8 March; accepted 12 September; published online 16 October 2011; doi:10.1038/ng.970

Table 1 Genetic loci associated with concentrations of liver enzymes in plasma at $P < 1 \times 10^{-8}$ in the GWAS

Region	Sentinel SNP	Position	Alleles (R/E)	EAF	Effect (%, 95% confidence interval)	*P*	Genes of interest
ALT							
4q22	rs6834314	88,432,832	G/A	0.75	2.6 (1.9–3.4)	3.1×10^{-9}	*HSD17B13*[ne], *MAPK10*[e]
8q24	rs2954021	126,551,259	G/A	0.50	1.6 (0.6–2.6)	5.3×10^{-9}	*TRIB1*[n]
10q24[a]	rs10883437	101,785,351	A/T	0.64	2.3 (1.4–3.1)	4.0×10^{-9}	*CPN1*[n]
22q13[a]	rs738409	42,656,060	C/G	0.23	6.0 (5.0–7.0)	1.2×10^{-45}	*PNPLA3*[nc], *SAMM50*[c]
ALP							
1p36.12[a]	rs1976403	21,639,040	A/C	0.40	3.6 (3.0–4.2)	1.8×10^{-50}	*ALPL*[o], *NBPF3*[nce]
2q24	rs16856332	169,548,820	G/T	0.96	3.9 (1.2–6.7)	1.6×10^{-9}	*ABCB11*[ng]
6p22[a]	rs1883415	24,599,454	A/C	0.33	3.1 (2.5–3.7)	5.6×10^{-26}	*ALDH5A1*[e], *GPLD1*[nc]
8p23	rs6984305	9,215,678	T/A	0.11	2.7 (1.1–4.4)	2.1×10^{-10}	*PPP1R3B*[ne]
8q24	rs2954021	126,551,259	G/A	0.50	1.4 (0.5–2.3)	2.3×10^{-13}	*TRIB1*[n]
9q21	rs10819937	103,263,054	G/C	0.17	2.5 (1.4–3.6)	1.0×10^{-9}	*ALDOB*[o], *C9orf125*[n]
9q34[a]	rs579459	135,143,989	C/T	0.80	8.8 (7.4–10.2)	2.6×10^{-123}	*ABO*[n]
10q21[a]	rs7923609	64,803,828	A/G	0.50	2.2 (1.7–2.7)	5.9×10^{-23}	*JMJD1C*[nce], *NRBF2*[e]
11q12	rs174601	61,379,716	C/T	0.35	1.7 (0.8–2.6)	2.6×10^{-9}	*C11orf10*[e], *FADS1*[e], *FADS2*[ne]
11q.24	rs2236653	125,788,995	C/T	0.42	1.5 (0.6–2.5)	1.8×10^{-9}	*ST3GAL4*[n]
16q22	rs7186908	70,777,874	G/C	0.24	2.0 (1.1–2.9)	4.8×10^{-9}	*HPR*[e], *PMFBP1*[n]
17p13	rs314253	7,032,374	T/C	0.33	2.1 (1.5–2.8)	8.4×10^{-12}	*ASGR1*[o], *DLG4*[n]
19q13[a]	rs281377	53,898,415	C/T	0.43	1.8 (0.8–2.8)	1.1×10^{-15}	*FUT2*[nc]
20p11	rs7267979	25,246,087	A/G	0.57	1.5 (0.9–2.0)	7.4×10^{-10}	*ABHD12*[ne],*GINS1*[ce], *PYGB*[o]
GGT							
1p36.13	rs1497406	16,377,907	A/G	0.56	3.8 (2.7–4.8)	2.8×10^{-19}	*RSG1*[e], *EPHA2*[ne]
1p22	rs12145922	88,918,822	C/A	0.61	2.8 (2.2–3.4)	3.8×10^{-11}	*CCBL2*[e], *PKN2*[n]
1p13	rs1335645	111,485,799	G/A	0.88	4.3 (3.5–5.2)	7.3×10^{-9}	*CEPT1*[ne], *DENND2D*[e]
1q21	rs10908458	153,393,572	C/T	0.58	3.7 (3.1–4.2)	1.7×10^{-15}	*DPM3*[n], *EFNA1*[ce], *PKLR*[o]
2p23	rs1260326	27,584,444	C/T	0.38	3.2 (2.4–4.0)	3.9×10^{-13}	*C2orf16*[e], *GCKR*[nc]
2q12	rs13030978	191,825,483	C/T	0.32	3.7 (2.8–4.6)	1.1×10^{-11}	*MYO1B*[ne], *STAT4*[e]
2q37	rs2140773	233,221,419	C/A	0.61	2.9 (2.3–3.5)	1.1×10^{-9}	*EFHD1*[ne], *LOC100129166*[c]
3q26	rs10513686	172,208,236	G/A	0.14	4.9 (4.0–5.7)	6.1×10^{-11}	*SLC2A2*[nc]
4q31	rs4547811	147,014,071	T/C	0.18	6.4 (5.0–7.9)	2.5×10^{-27}	*ZNF827*[n]
5p15	rs6888304	31,056,278	G/A	0.74	2.7 (2.0–3.5)	1.2×10^{-9}	*CDH6*[n]
5q11	rs4074793	52,228,882	A/G	0.07	5.5 (3.3–7.7)	3.4×10^{-10}	*ITGA1*[n]
6p12	rs9296736	54,032,656	C/T	0.31	3.0 (2.1–4.0)	2.6×10^{-9}	*MLIP*[ne]
7q11	rs17145750	72,664,314	T/C	0.86	4.5 (2.9–6.3)	2.9×10^{-9}	*MLXIPL*[nce]
10q23	rs754466	79,350,440	A/T	0.24	3.5 (2.2–4.8)	6.4×10^{-10}	*DLG5*[n]
12q24[a]	rs7310409	119,909,244	A/G	0.59	6.8 (5.7–7.8)	7.0×10^{-45}	*HNF1A*[nc], *C12orf27*[e]
14q32	rs944002	102,642,568	A/G	0.21	6.3 (4.9–7.7)	5.8×10^{-29}	*C14orf73*[nc]
15q21	rs339969	58,670,573	C/A	0.62	4.5 (3.9–5.1)	6.6×10^{-20}	*RORA*[n]
15q23	rs8038465	71,765,390	C/T	0.39	2.4 (1.8–3.0)	1.4×10^{-9}	*CD276*[ne]
16q23	rs4581712	79,055,102	C/A	0.27	3.2 (2.5–3.9)	3.1×10^{-9}	*DYNLRB2*[n]
17q24	rs9913711	67,609,756	G/C	0.65	2.4 (1.8–3.0)	1.3×10^{-9}	*FLJ37644*[e], *SOX9*[n]
18q21.31	rs12968116	53,473,500	T/C	0.87	4.8 (2.8–6.7)	8.9×10^{-10}	*ATP8B1*[ncg]
18q21.32	rs4503880	54,235,034	C/T	0.21	3.6 (2.5–4.7)	3.0×10^{-12}	*NEDD4L*[n]
19q13[a]	rs516246	53,897,984	C/T	0.47	2.3 (1.8–2.9)	7.6×10^{-10}	*FUT2*[nc]
22q11.21	rs1076540	16,819,958	T/C	0.78	4.8 (3.5–6.1)	9.6×10^{-17}	*MICAL3*[ne]
22q11.23	rs2739330	22,625,286	C/T	0.42	3.7 (2.7–4.6)	1.7×10^{-9}	*DDT*[e], *DDTL*[e], *GSTT1*[e], *GSTT2B*[n], *MIF*[e]
22q11.23[a]	rs2073398	23,329,104	C/G	0.34	12.3 (10.9–13.7)	1.1×10^{-109}	*GGT1*[ne], *GGTLC2*[e]

Alleles are given as the reference (R) allele/effect (E). EAF, effect allele frequency; effect is change in concentration of liver enzyme in plasma per copy of effect allele.
[a]Previously reported associations. Annotation for genes of interest: [n]nearest; [e]expression QTL; [c]coding SNP; [g]GRAIL; [o]known biology.

marker at $P < 6 \times 10^{-4}$ (corresponding to $P < 0.05$ after Bonferroni correction for testing 42 loci against two alternate liver markers; **Supplementary Fig. 4** and **Supplementary Table 8**). The loci previously reported to be associated with liver markers in GWASs were replicated in the current study, except for variants at the *ALDH2* locus reported in Japanese populations, which have low allele frequency in European populations[10,11].

We used coding variation, expression quantitative trait loci (eQTL) and GRAIL analyses to identify possible candidate genes at the 42 loci associated with liver enzymes (**Table 1** and **Supplementary Table 9**). There are 19 nonsynonymous SNPs (nsSNPs) that are in linkage disequilibrium (LD) with one or more of the sentinel SNPs at $r^2 \geq 0.5$ in the HapMap phase II CEU data set[12] (see URLs), representing a ~3.5-fold enrichment compared with the number expected under the null hypothesis ($P = 0.004$). We considered the gene containing the nsSNP to be a strong candidate when (i) the nsSNP and the sentinel SNPs were in LD ($r^2 > 0.5$) and (ii) there was no evidence for heterogeneity of effect on phenotype. The genes with coding variants identified as

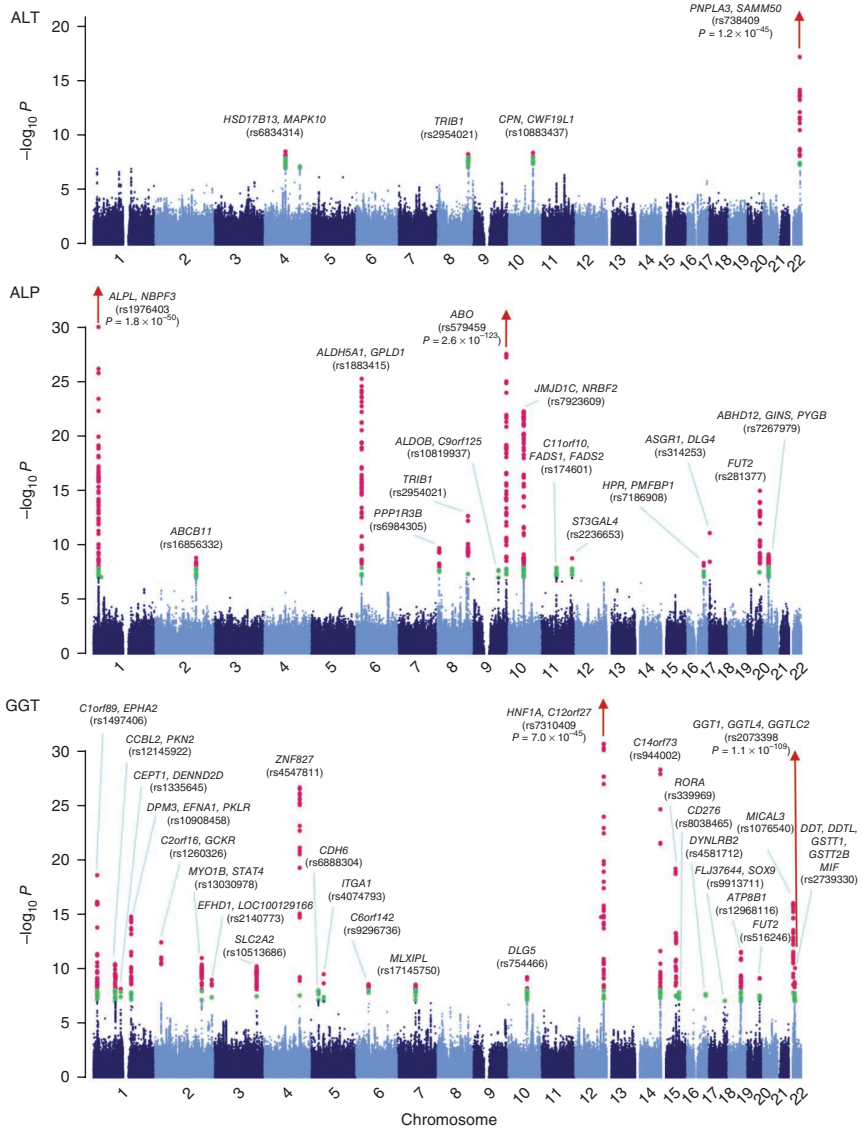

Figure 2 Manhattan plots of association of SNPs with ALT, ALP and GGT in the GWAS. SNPs reaching genome-wide significance ($P < 1 \times 10^{-8}$) are red; SNPs with $P > 1 \times 10^{-8}$ and $P < 1 \times 10^{-7}$ are green.

candidates for mediating the observed associations with liver markers (**Supplementary Table 10**) encode proteins involved in biliary transport (ATP8B1)[13], cell surface glycobiology, endoplasmic trafficking and susceptibility to gastrointestinal infection (FUT2 and GPLD1)[14,15], carbohydrate and lipid metabolism, including susceptibility to type 2 diabetes (GCKR, HNF1A and SLC2A2)[16–18] and inflammation as measured by circulating concentrations of C-reactive protein (CRP) (GCKR and HNF1A)[19]. Mutations in *ATP8B1* are responsible for progressive familial intrahepatic cholestasis and are associated with high GGT concentrations[20]; the coding variant identified is predicted to be nonconservative (**Supplementary Fig. 5**). At chromosome 14q32, rs944002 is in LD ($r^2 = 0.86$) with two nsSNPs in *C14orf73*, a gene strongly expressed in liver. C14orf73 has strong sequence homology with SEC6, a protein that interacts with the actin cytoskeleton and vesicle transport machinery[21]. Of the two nsSNPs reported in *C14orf73*, p.Arg77Trp is predicted to be a nonconservative change from a polar basic residue to a nonpolar hydrophobic residue (**Supplementary Fig. 5**).

We repeated the search for coding variants using available results from the 1000 Genomes Project[22] (see URLs) and identified coding variants in two additional genes, *NBPF3* (chromosome 1p36.12) and *MLXIPL* (chromosome 7q11). Both genes are separately implicated as candidates for genes mediating the associations of sentinel SNPs with liver markers through eQTL analyses.

We examined the association of the sentinel SNPs with eQTL data from liver, fat and peripheral blood leucocytes[23–25] (**Supplementary Tables 11–14**). We tested SNPs for association with expression of nearby (within 1 Mb) genes (at $P < 0.05$ after Bonferroni correction for number of SNP expression associations tested). When we identified probable eQTLs, we tested whether the sentinel SNP and the SNP most closely associated with the eQTL were coincident ($r^2 > 0.5$ and absence of heterogeneity at the phenotype or eQTL). This strategy identified eQTLs at 23 of the 42 loci, representing genes implicated in glutathione metabolism and drug detoxification (*GSTT1* and *GGT1*), carbohydrate and lipid metabolism (*MLXIPL*, *PPP1R3B*, *FADS1* and *FADS2*), cell signaling (*ABHD12* and *EPHA2*) and inflammation and immunity (*STAT4*, *MAPK10*, *CD276* and *HPR*). The functions of the other candidate genes identified by eQTLs (including *EFHD1*, *MICAL3*, *DENND2D*, *CEPT1*, *MLIP* (also known as *C6orf142*) and *RSG1* (also known as *C1orf89*)) are poorly understood.

We also carried out a literature analysis using the GRAIL algorithm[26] (see URLs), initially using the 2006 data set to avoid studies of the GWAS era. At chromosome 2q24, GRAIL identified *ABCB11* as the most plausible candidate (**Supplementary Table 15**). ABCB11 activity is a major determinant of bile formation and bile flow[27]; mutations in *ABCB11* cause progressive familial intrahepatic cholestasis type 2 and are associated with increased risk of hepatocellular carcinoma[28,29]. We repeated the GRAIL analysis using the 2010 PubMed data set. This also identified *ABCB11* as the plausible candidate at chromosome 2q24 but additionally identified *ABO*, *GCKR*, *MLXIPL* and *PNPLA3* as probable candidates at other loci (**Supplementary Table 15**), replicating our findings from coding variant and eQTL analyses.

Through our coding variant, expression and GRAIL analyses, we identified 44 genes as strong candidates at the 42 loci associated with concentrations of liver enzymes in plasma. We also considered the gene nearest to the sentinel SNP at each locus to be a potential candidate. Together these approaches identified 69 candidate genes. Pathway analyses showed subnetworks of closely interconnected genes (**Supplementary Fig. 6**) from core metabolic pathways and processes including carbohydrate metabolism, insulin signaling and diabetes (*GCKR*, *SLC2A2*, *PPP1R3B*, *FUT2*, *ALDOB*, *HNF1A* and *MLXIPL*), lipid metabolism (*CEPT1*, *FADS1*, *FADS2*, *HNF1A*, *PNPLA3* and *ALDH5A1*), glycosphingolipid biosynthesis

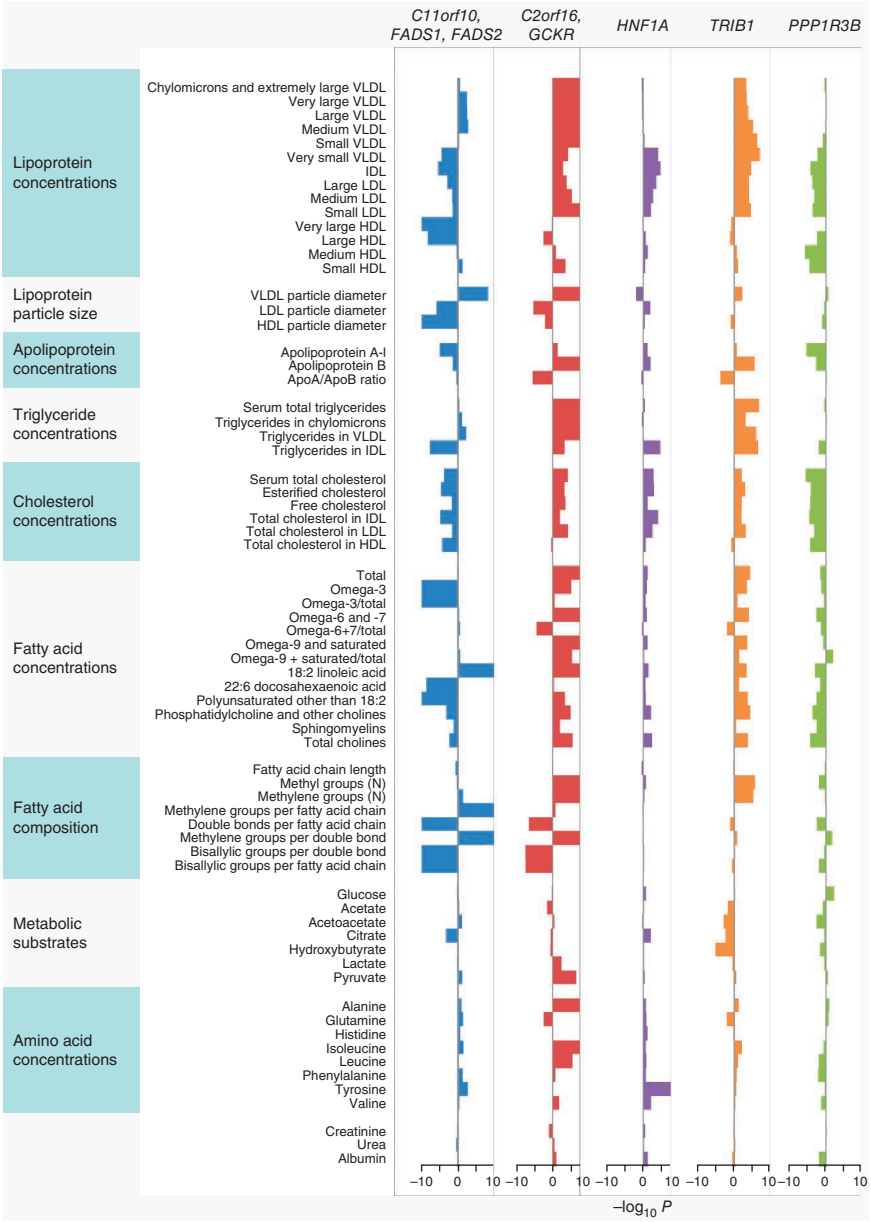

Figure 3 Association of *FADS1*, *FADS2*, *GCKR*, *HNF1A*, *TRIB1* and *PPP1R3B* loci with NMR metabonome. Bars are for $-\log_{10} P$ value, signed for direction of effect.

and glycosylation (*ST3GAL4*, *FUT2* and *ABO*) and glutathione metabolism (*ALDHA5*, *GGT1* and *GSTT1*).

Of the 42 liver marker loci, 24 have been reported to be associated with other phenotypes in genome-wide studies (**Supplementary Table 16**). At 12 of the loci, the lead SNP for the liver marker and the phenotype are the same or in LD at $r^2 \geq 0.5$, suggesting shared biological pathways. The phenotypes include Crohn's disease, pancreatic carcinoma, type 2 diabetes, waist circumference and concentrations of glucose, insulin, total, high-density lipoprotein (HDL) and low-density lipoprotein (LDL) cholesterol, triglycerides, fatty acids, uric acid and C-reactive protein. At other loci, the sentinel SNP from the liver marker GWAS and the lead SNP in the US National Human Genome Research Institute (NHGRI) catalog[30] (see URLs) are in low LD, suggesting that these likely represent different underlying mechanisms. We also ascertained the relationships of the 42 loci with quantitative anthropometric and metabolic traits in published genome-wide meta-analyses (**Supplementary Table 17**). We found that the loci associated with liver enzymes are enriched in SNPs associated with lipid concentrations, fasting glucose and inflammation as measured by CRP.

We used metabonomic profiling, the systematic characterization of a metabolite panel, to better understand the relationships of the 42 liver enzyme loci with intermediary and lipoprotein metabolism. We carried out quantitative nuclear magnetic resonance (NMR) spectroscopy on serum samples from 6,516 participants from the London Life Sciences Population[31] (LOLIPOP) and Northern Finland Birth Cohort 1966 (ref. 32; NFBC1966) studies. Significance was inferred at $P < 1 \times 10^{-5}$, corresponding to $P < 0.05$ after Bonferroni correction for the 42 independent SNPs tested, and for the 69 primary NMR measures. At chromosomes 2p23 (*C2orf16* and *GCKR*) and 8q24 (*TRIB1*), effect alleles of the sentinel SNPs are associated with high very low-density lipoprotein, intermediate-density lipoprotein and LDL concentration and VLDL particle size, high lipoprotein triglyceride and cholesterol concentration, omega-3 and omega-6 fatty acid concentrations, and concentrations of metabolic substrates citrate, pyruvate and branch chain amino acids (**Fig. 3**). At chromosome 12q24 (*HNF1A*), rs7310409 is associated with lipoprotein concentration and composition, and with tyrosine concentrations. At chromosomes 11q12 (*C11orf10*, *FADS1* and *FADS2*) and 8p23 (*PPP1R3B*), the effect alleles are associated with low concentrations of cholesterol and HDL cholesterol and with low concentrations of omega-3 and other unsaturated fatty acids. Our results from the NMR confirm and extend previous studies using mass spectroscopy, which showed strong association of *GCKR* and *FADS1* with absolute and relative abundances of polyunsaturated fatty acids[33,34].

We examined the contribution of the 42 genetic loci to concentrations of liver enzymes in plasma among the 8,112 participants of the LIFELINES population study[35]. SNPs at 41 loci showed consistent direction of effect ($P = 4 \times 10^{-13}$, sign test; **Supplementary Table 18**). Together the SNPs associated with each liver enzyme account for 0.1%, 3.5% and 1.9% of population variation in plasma concentrations of ALT, ALP and GGT, respectively (**Supplementary Table 19**). We then constructed a SNP score as the unweighted sum of the effect allele counts for the SNPs associated with each liver marker. Participants in the top quartile of distribution for SNP score for ALT, ALP or GGT were ~1.4, ~2.4 and ~1.8 times more probable to have greater than the upper limit of normal concentrations of ALT, ALP and GGT, and on average had concentrations of ALT, ALP and GGT that were 7%, 13% or 26% higher, respectively, than participants in the lowest quartile of SNP score (**Supplementary Table 19**).

Finally we tested the relationship of the liver enzyme–associated loci with the presence of structural changes in the liver indicative of hepatic steatosis, as determined by computerized axial tomography (CT) scanning in a population sample of 9,610 participants of the Genetics of Liver Disease (GOLD) study[36]. SNPs at five loci were associated with hepatic steatosis at $P < 0.05$, including *PNPLA3*, *PPP1R3B*, *GCKR*, *TRIB1*, *HNF1A* and *SOX9* loci (**Supplementary Table 20**); of these, *PNPLA3*, *PPP1R3B* and *GCKR* were associated with hepatic steatosis at $P < 0.0012$ (that is, $P < 0.05$ after Bonferroni correction for 42 loci).

We identify 42 independent loci associated with ALP, ALT or GGT and 69 genes as candidates for the associations observed (**Supplementary Table 9**). The candidate genes include *ATP8B1* and *ABCB11*, encoding biliary transporters with a key role in bile formation and flow[20,37], and many genes involved in carbohydrate and lipid metabolism, including *GCKR*, *MLXIPL*, *SLC2A2*, *HNF1A*, *PNPLA3*, *FADS1*, *FADS2* and *PPP1R3B*[17,38,39]. *PNPLA3*, *PPP1R3B* and *GCKR* influence accumulation of hepatic triglycerides[40,41]. We identify *GSTT1*, *GSTT2* and *GGT* as candidates encoding key enzymes in glutathione synthesis and drug metabolism[42,43]; these observations may be relevant to pharmacogenetics and drug development. We also identify a set of genes involved in inflammation and immunity, including *CD276*, *CDH6*, *GCKR*, *HPR*, *ITGA1*, *MAPK10*, *RORA* and *STAT4*. Whether these genes influence hepatic inflammatory responses to accumulation of triglycerides, viral infection or other exogenous challenges remains to be determined. Finally we identify a set of genes involved in glycoprotein biology, including *ABO*, *ASGR1*, *FUT2*, *GPLD1* and *ST3GAL4*. The products of these genes influence synthesis, cell surface binding and turnover of glycoproteins. These pathways are linked to susceptibility to pancreatic[44] and gastric malignancy[45], intestinal and other infections[46] and vitamin B_{12} metabolism[47]. The pleiotropic nature of the genes we identified suggests that their relationships with ALP, ALT or GGT may also be mediated by pathways operating outside of the liver.

In summary, we report a GWAS for concentrations of liver enzymes in plasma, providing new insight into the genetic variation and pathways influencing ALP, ALT and GGT. Our findings provide the basis for further studies investigating the biological mechanisms involved in liver injury.

URLs. 1000 Genomes, http://www.1000genomes.org/ (July 2010 data set); HapMap CEU, http://hapmap.ncbi.nlm.nih.gov/downloads/genotypes/latest_phaseII_ncbi_b36/ (release 07-July-2009); GRAIL, http://www.broadinstitute.org/mpg/grail/grail.php; NHGRI, http://www.genome.gov/gwastudies/ (accessed 2 September 2010).

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the many colleagues who contributed to collection and phenotypic characterization of the clinical samples, as well as genotyping and analysis of data. We also thank the research participants who took part in these studies. Major support for the work came from European Commission (FP5, FP6 and FP7); European Science Foundation; European Science Council; US NIH; US National Institute of Mental Health; US NIDDK; Genetic Association Information Network; US National Institute on Aging; US National Human Genome Research Institute; US NHLBI; UK NIHR; NIHR Comprehensive Biomedical Research Centre Imperial College Healthcare NHS Trust; NIHR Comprehensive Biomedical Research Centre Guy's and St. Thomas' NHS Trust; UK Biotechnology and Biological Sciences Research Council; UK MRC; British Heart Foundation; Wellcome Trust; Swiss National Science Foundation; Academy of Finland; Finnish Cardiovascular Research Foundation; Swedish Research Council; Swedish Heart-Lung Foundation; Helmholtz Zentrum München; German Research Center for Environmental Health; German Federal Ministry of Education and Research; German National Genome Research Network; Netherlands Organization for Scientific Research; Dutch Ministries of Economic Affairs, of Education, Culture and Science, for Health, Welfare and Sports; Netherlands Organization for Health Research and Development; Economic Structure Enhancing Fund of the Dutch government; Dutch Kidney Foundation; Dutch Diabetes Research Foundation; Dutch Brain Foundation; Dutch Research Institute for Diseases in the Elderly; Netherlands Genomics Initiative; Canadian Institutes for Health Research; Ontario Research Fund; The Barts and the London Charity; University Medical Center Groningen; University of Groningen; University of Oulu, Biocenter Oulu; University Hospital Oulu; Biocentrum Helsinki; Erasmus Medical Center and Erasmus University, Rotterdam; Karolinska Institutet; Stockholm County Council; Municipality of Rotterdam; Federal State of Mecklenburg-West Pomerania; AstraZeneca; GlaxoSmithKline; Siemens Healthcare; Novo Nordisk Foundation; Yrjö Jahnsson Foundation; Biomedicum Helsinki Foundation; Gyllenberg Foundation; Knut and Alice Wallenberg Foundation; Torsten and Ragnar Söderberg Foundation; Robert Dawson Evans Endowment, Boston University School of Medicine; Instrumentarium Science Foundation; Jenny and Antti Wihuri Foundation and the Canadian Primary Biliary Cirrhosis Society. A full list of acknowledgments is provided in the **Supplementary Note**.

AUTHOR CONTRIBUTIONS

Study organization and manuscript preparation was done by J.C.C., W.Z., J. Sehmi, X.L., M.N.W., P.V.d.H., H.H., S.S., M.K., M.A.-K., K.S., P.V., H.V., E.E.S., J. Scott, M.-R.J., P.E. and J.S.K. All authors reviewed and had the opportunity to comment on the manuscript. Data collection and analysis in the participating GWASs were done by G.W.M., J.B.W. and N.G.M. (Australian Twin Cohort); C.W., M.C., M.J.B. and P.B.M. (BRIGHT); D.M.W., G. Waeber, P.M.V., P.V., V.M. and X.Y. (CoLaus); D.F.G., G.I.E., G.T., H.H., I.O., K.S. and U.T. (deCODE); J.L., N.G.F., N.J.W. and R.J.F.L. (Fenland); K.H.P. (Finnish Twin Cohort); C.J.O., C.S.F., J.P.L., L.D.A., N.L.H.-C., R.S.V., T.J.L. and W.G. (Framingham Heart Study); A.D., B.K., C.G., C.M. and H.-E.W. (KORA); B.H.R.W., I.M.L., I.P.K., M.M.V.d.K., P.V.d.H. and R.P.S. (LIFELINES); D.D., G.D., H.C.T., I.P., J.C.C., J. Scott, J. Sehmi, J.S.K., M.I.M., P.E., P.F., S.S.-C., W.Z., X.L. and Y.L. (LOLIPOP); B.P.P., B.W.P., B.Z.A., H.S., J.H.S. and V.L. (NESDA); D.I.B., E.J.C.d.G., G. Willemsen, J.-J.H. (Netherlands Twins Register); A.-L.H., A.P., A.R., E.H., M.-R.J. and P.F.O. (Northern Finland Birth Cohort 1966); H. Watkins, J.F.P., M.F. and U.S. (PROCARDIS); A.G.U., A.H., C.M.v.D., H.L.A.J., J.C.M.W., J.N.L.S. and M.K. (Rotterdam Study 1); D.S., F.C., G.R.A., M.U., S.L. and S.S. (SardiNIA); G.H., H.V., H. Wallaschofski, J.P.K., M.M.L., N.F., R.P. and S.E.B. (SHIP); K.R.A., N.R. and T.D.S. (TwinsUK). Biologic associations of loci and bioinformatics were carried out by G.D., W.T., K. Matsuda, V.K., Y.N. and by G.S., L.J.C., P.C. (AlcGen Consortium), C.X., G.M.H., K.A.S. (Canadian Primary Biliary Cirrhosis Consortium), K. Musunuru, T.M.T. (Global Lipids Consortium), E.K.S., I.B.B., L.M.Y.A., T.B.H. (GOLD consortium) and G.E. and T.J. (ICBP-GWAS). Gene expression analyses were done by E.E.S., A.L.D., H.H., G.T., L.L., M.F.M., M.L., S.H. and W.O.C. Metabonomic analyses were done by A.J.K., M.A.-K., M.J.S., P.S., P.W. and T.T. Structural biology was done by M.J.E.S. and M.N.W.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Pratt, D.S. & Kaplan, M.M. Evaluation of abnormal liver-enzyme results in asymptomatic patients. *N. Engl. J. Med.* **342**, 1266–1271 (2000).
2. Söderberg, C. *et al.* Decreased survival of subjects with elevated liver function tests during a 28-year follow-up. *Hepatology* **51**, 595–602 (2010).
3. Xu, K. *et al.* Diagnostic value of serum γ-glutamyl transferase isoenzyme for hepatocellular carcinoma: a 10-year study. *Am. J. Gastroenterol.* **87**, 991–995 (1992).
4. Sattar, N. *et al.* Elevated alanine aminotransferase predicts new-onset type 2 diabetes independently of classical risk factors, metabolic syndrome, and C-reactive protein in the west of Scotland coronary prevention study. *Diabetes* **53**, 2855–2860 (2004).
5. Ioannou, G.N., Weiss, N.S., Boyko, E.J., Mozaffarian, D. & Lee, S.P. Elevated serum alanine aminotransferase activity and calculated risk of coronary heart disease in the United States. *Hepatology* **43**, 1145–1151 (2006).
6. Watkins, P.B. Idiosyncratic liver injury: challenges and approaches. *Toxicol. Pathol.* **33**, 1–5 (2005).

7. Rahmioglu, N. *et al.* Epidemiology and genetic epidemiology of the liver function test proteins. *PLoS ONE* **4**, e4435 (2009).
8. Nugent, C. & Younossi, Z.M. Evaluation and management of obesity-related nonalcoholic fatty liver disease. *Nat. Clin. Pract. Gastroenterol. Hepatol.* **4**, 432–441 (2007).
9. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M.J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* **32**, 381–385 (2008).
10. 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 (2008).
11. Kamatani, Y. *et al.* Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat. Genet.* **42**, 210–215 (2010).
12. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
13. Paulusma, C.C. *et al.* Atp8b1 deficiency in mice reduces resistance of the canalicular membrane to hydrophobic bile salts and impairs bile salt transport. *Hepatology* **44**, 195–204 (2006).
14. Iwamori, M. & Domino, S.E. Tissue-specific loss of fucosylated glycolipids in mice with targeted deletion of alpha(1,2)fucosyltransferase genes. *Biochem. J.* **380**, 75–81 (2004).
15. LeBoeuf, R.C. *et al.* Mouse glycosylphosphatidylinositol-specific phospholipase D (Gpld1) characterization. *Mamm. Genome* **9**, 710–714 (1998).
16. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42**, 105–116 (2010).
17. Voight, B.F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42**, 579–589 (2010).
18. Zeggini, E. *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.* **40**, 638–645 (2008).
19. Elliott, P. *et al.* Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease. *J. Am. Med. Assoc.* **302**, 37–48 (2009).
20. Bull, L.N. *et al.* A gene encoding a P-type ATPase mutated in two forms of hereditary cholestasis. *Nat. Genet.* **18**, 219–224 (1998).
21. Shin, D.M., Zhao, X.S., Zeng, W., Mozhayeva, M. & Muallem, S. The mammalian Sec6/8 complex interacts with Ca(2+) signaling complexes and regulates their activity. *J. Cell Biol.* **150**, 1101–1112 (2000).
22. The 1000 Genomes Projects Consortium. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
23. Dixon, A.L. *et al.* A genome-wide association study of global gene expression. *Nat. Genet.* **39**, 1202–1207 (2007).
24. Schadt, E.E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107 (2008).
25. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
26. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5**, e1000534 (2009).
27. Noé, J., Stieger, B. & Meier, P.J. Functional expression of the canalicular bile salt export pump of human liver. *Gastroenterology* **123**, 1659–1666 (2002).
28. van Mil, S.W. *et al.* Benign recurrent intrahepatic cholestasis type 2 is caused by mutations in ABCB11. *Gastroenterology* **127**, 379–384 (2004).
29. Knisely, A.S. *et al.* Hepatocellular carcinoma in ten children under five years of age with bile salt export pump deficiency. *Hepatology* **44**, 478–486 (2006).
30. Hindorff, L.A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **106**, 9362–9367 (2009).
31. Chambers, J.C. *et al.* Genetic loci influencing kidney function and chronic kidney disease. *Nat. Genet.* **42**, 373–375 (2010).
32. Sabatti, C. *et al.* Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat. Genet.* **41**, 35–46 (2009).
33. Gieger, C. *et al.* Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet.* **4**, e1000282 (2008).
34. Illig, T. *et al.* A genome-wide perspective of genetic variation in human metabolism. *Nat. Genet.* **42**, 137–141 (2010).
35. Stolk, R.P. *et al.* Universal risk factors for multifactorial diseases: LifeLines: a three-generation population-based study. *Eur. J. Epidemiol.* **23**, 67–74 (2008).
36. 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 (2011).
37. Klomp, L.W. *et al.* Characterization of mutations in ATP8B1 associated with hereditary cholestasis. *Hepatology* **40**, 27–38 (2004).
38. Petit, J.M. *et al.* Specifically PNPLA3-mediated accumulation of liver fat in obese patients with type 2 diabetes. *J. Clin. Endocrinol. Metab.* **95**, E430–E436 (2010).
39. Dunn, J.S. *et al.* Examination of PPP1R3B as a candidate gene for the type 2 diabetes and MODY loci on chromosome 8p23. *Ann. Hum. Genet.* **70**, 587–593 (2006).
40. He, S. *et al.* A sequence variation (I148M) in PNPLA3 associated with nonalcoholic fatty liver disease disrupts triglyceride hydrolysis. *J. Biol. Chem.* **285**, 6706–6715 (2010).
41. Saxena, R. *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331–1336 (2007).
42. Bolt, H.M. & Thier, R. Relevance of the deletion polymorphisms of the glutathione S-transferases GSTT1 and GSTM1 in pharmacology and toxicology. *Curr. Drug Metab.* **7**, 613–628 (2006).
43. Zhang, H., Forman, H.J. & Choi, J. γ-glutamyl transpeptidase in glutathione biosynthesis. *Methods Enzymol.* **401**, 468–483 (2005).
44. Wolpin, B.M. *et al.* Variant ABO blood group alleles, secretor status, and risk of pancreatic cancer: results from the pancreatic cancer cohort consortium. *Cancer Epidemiol. Biomarkers Prev.* **19**, 3140–3149 (2010).
45. Edgren, G. *et al.* Risk of gastric cancer and peptic ulcers in relation to ABO blood type: a cohort study. *Am. J. Epidemiol.* **172**, 1280–1285 (2010).
46. Lindesmith, L. *et al.* Human susceptibility and resistance to Norwalk virus infection. *Nat. Med.* **9**, 548–553 (2003).
47. Hazra, A. *et al.* Common variants of FUT2 are associated with plasma vitamin B12 levels. *Nat. Genet.* **40**, 1160–1162 (2008).

John C Chambers[1–3,113], Weihua Zhang[1,3,113], Joban Sehmi[3,4,113], Xinzhong Li[5,113], Mark N Wass[6,113], Pim Van der Harst[7,113], Hilma Holm[8,113], Serena Sanna[9,113], Maryam Kavousi[10,11,113], Sebastian E Baumeister[12], Lachlan J Coin[1], Guohong Deng[13], Christian Gieger[14], Nancy L Heard-Costa[15], Jouke-Jan Hottenga[16], Brigitte Kühnel[14], Vinod Kumar[17], Vasiliki Lagou[18–20], Liming Liang[21,22], Jian'an Luan[23], Pedro Marques Vidal[24], Irene Mateo Leach[7], Paul F O'Reilly[1], John F Peden[25], Nilufer Rahmioglu[19], Pasi Soininen[26,27], Elizabeth K Speliotes[28,29], Xin Yuan[30], Gudmar Thorleifsson[8], Behrooz Z Alizadeh[18], Larry D Atwood[31], Ingrid B Borecki[32], Morris J Brown[33], Pimphen Charoen[1,34], Francesco Cucca[9], Debashish Das[3], Eco J C de Geus[16,35], Anna L Dixon[36], Angela Döring[37], Georg Ehret[38–40], Gudmundur I Eyjolfsson[41], Martin Farrall[25,42], Nita G Forouhi[23], Nele Friedrich[43], Wolfram Goessling[44–46], Daniel F Gudbjartsson[8], Tamara B Harris[47], Anna-Liisa Hartikainen[48], Simon Heath[49], Gideon M Hirschfield[50–52], Albert Hofman[10,11], Georg Homuth[53], Elina Hyppönen[54], Harry L A Janssen[10,55], Toby Johnson[56], Antti J Kangas[26], Ido P Kema[57], Jens P Kühn[58], Sandra Lai[9], Mark Lathrop[49,59], Markus M Lerch[60], Yun Li[61], T Jake Liang[62], Jing-Ping Lin[63], Ruth J F Loos[23], Nicholas G Martin[64], Miriam F Moffatt[36], Grant W Montgomery[64], Patricia B Munroe[56], Kiran Musunuru[31,65–68], Yusuke Nakamura[17], Christopher J O'Donnell[69], Isleifur Olafsson[70], Brenda W Penninx[71–73], Anneli Pouta[48,74], Bram P Prins[18], Inga Prokopenko[19,20], Ralf Puls[58], Aimo Ruokonen[75], Markku J Savolainen[26,76], David Schlessinger[77], Jeoffrey N L Schouten[55], Udo Seedorf[78], Srijita Sen-Chowdhry[1], Katherine A Siminovitch[50,79–82], Johannes H Smit[71], Timothy D Spector[83], Wenting Tan[13], Tanya M Teslovich[84], Taru Tukiainen[1,26], Andre G Uitterlinden[10,11,85], Melanie M Van der Klauw[86,87], Ramachandran S Vasan[88,89], Chris Wallace[33], Henri Wallaschofski[43], H-Erich Wichmann[37,90,91], Gonneke Willemsen[16,92], Peter Würtz[1,26], Chun Xu[93], Laura M Yerges-Armstrong[94], Alcohol Genome-wide Association (AlcGen) Consortium[95],

Diabetes Genetics Replication and Meta-analyses (DIAGRAM+) Study[95], Genetic Investigation of Anthropometric Traits (GIANT) Consortium[95], Global Lipids Genetics Consortium[95], Genetics of Liver Disease (GOLD) Consortium[95], International Consortium for Blood Pressure (ICBP-GWAS)[95], Meta-analyses of Glucose and Insulin-Related Traits Consortium (MAGIC)[95], Goncalo R Abecasis[84], Kourosh R Ahmadi[83], Dorret I Boomsma[16,92], Mark Caulfield[56], William O Cookson[36], Cornelia M van Duijn[10,11,96], Philippe Froguel[97], Koichi Matsuda[17], Mark I McCarthy[19,20,98], Christa Meisinger[99], Vincent Mooser[30], Kirsi H Pietiläinen[100–102], Gunter Schumann[103], Harold Snieder[18], Michael J E Sternberg[6,87], Ronald P Stolk[104], Howard C Thomas[2,105], Unnur Thorsteinsdottir[8,106], Manuela Uda[9], Gérard Waeber[107], Nicholas J Wareham[23], Dawn M Waterworth[30], Hugh Watkins[25,42], John B Whitfield[64], Jacqueline C M Witteman[10,11], Bruce H R Wolffenbuttel[86,87], Caroline S Fox[69,108], Mika Ala-Korpela[26,27,76,113], Kari Stefansson[8,106,113], Peter Vollenweider[107,113], Henry Völzke[12,113], Eric E Schadt[109,113], James Scott[4,113], Marjo-Riitta Järvelin[1,74,110–113], Paul Elliott[1,112,113] & Jaspal S Kooner[2–4,113]

[1]Epidemiology and Biostatistics, Imperial College London, Norfolk Place, London, UK. [2]Imperial College Healthcare National Health Service (NHS) Trust, London, UK. [3]Ealing Hospital NHS Trust, Middlesex, UK. [4]National Heart and Lung Institute, Imperial College London, Hammersmith Hospital, London, UK. [5]Institute of Clinical Science, Imperial College London, Royal Brompton Hospital, London, UK. [6]Structural Bioinformatics Group, Division of Molecular Biosciences, Imperial College London, South Kensington, London, UK. [7]Department of Cardiology, University Medical Center Groningen, University of Groningen, The Netherlands. [8]deCODE genetics, Reykjavik, Iceland. [9]Istituto di Ricerca Genetica e Biomedica del Consiglio Nazionale delle Ricerche, Monserrato, Cagliari, Italy. [10]Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands. [11]Netherlands Genomics Initiative-Sponsored Netherlands Consortium for Health Aging, Rotterdam, The Netherlands. [12]Institute for Community Medicine, University of Greifswald, Germany. [13]Institute of Infectious Diseases, Southwest Hospital, Third Military Medical University, Chongqing, China. [14]Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [15]Department of Neurology, Boston University School of Medicine, Boston Massachusetts, USA. [16]Department of Biological Psychology, VU University Amsterdam (VUA), Amsterdam, The Netherlands. [17]Laboratory of Molecular Medicine, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [18]Unit of Genetic Epidemiology and Bioinformatics, Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [19]Wellcome Trust Center for Human Genetics, University of Oxford, Oxford, UK. [20]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Oxford, UK. [21]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. [22]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA. [23]Medical Research Council (MRC) Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge UK. [24]Institute of Social and Preventive Medicine (IUMSP), University Hospital and University of Lausanne, Lausanne, Switzerland. [25]Department of Cardiovascular Medicine, The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [26]Computational Medicine Research Group, Institute of Clinical Medicine, University of Oulu and Biocenter Oulu, Oulu, Finland. [27]Nuclear Magnetic Resonance (NMR) Metabonomics Laboratory, Department of Biosciences, University of Eastern Finland, Kuopio, Finland. [28]Department of Internal Medicine, Division of Gastroenterology, University of Michigan, Ann Arbor, Michigan, USA. [29]Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA. [30]Genetics, GlaxoSmithKline, King of Prussia, Pennsylvania, USA. [31]Boston University School of Medicine, Boston, Massachusetts, USA. [32]Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, Saint Louis, Missouri, USA. [33]The Diabetes Inflammation Laboratory, Cambridge Institute of Medical Research, University of Cambridge, Cambridge, UK. [34]Department of Tropical Hygiene, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand. [35]Neuroscience Campus Amsterdam, VUA and VUA Medical Center, Amsterdam, The Netherlands. [36]National Heart and Lung Institute, Imperial College London, London, UK. [37]Institute of Epidemiology I, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [38]Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [39]IUMSP, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland. [40]Cardiology, Department of Medicine, Geneva University Hospital, Geneva, Switzerland. [41]The Laboratory in Mjodd, Reykjavik, Iceland. [42]Department of Cardiovascular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, UK. [43]Institute of Clinical Chemistry and Laboratory Medicine, University of Greifswald, Germany. [44]Genetics and Gastroenterology Divisions, Brigham and Women's Hospital, Gastrointestinal Cancer Center, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [45]Harvard Medical School, Boston, Massachusetts, USA. [46]Harvard Stem Cell Institute, Cambridge, Massachusetts, USA. [47]Laboratory of Epidemiology, Demography, and Biometry, National Institute on Aging, US National Institutes of Health (NIH), Bethesda, Maryland, USA. [48]Institute of Clinical Medicine, University of Oulu, Oulu, Finland. [49]CEA-IG Centre National de Genotypage, Evry Cedex, France. [50]Department of Medicine, University of Toronto, Toronto, Ontario, Canada. [51]Liver Center, Toronto Western Hospital, Toronto, Ontario, Canada. [52]Centre for Liver Research, University of Birmingham, Birmingham, UK. [53]Interfaculty Institute for Genetics and Functional Genomics, University of Greifswald, Greifswald, Germany. [54]Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, London, UK. [55]Department of Gastroenterology and Hepatology, Erasmus Medical Center, Rotterdam, The Netherlands. [56]Clinical Pharmacology and The Genome Center, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [57]Department of Laboratory Medicine, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [58]Institute of Diagnostic Radiology and Neuroradiology, University of Greifswald, Greifswald, Germany. [59]Fondation Jean Dausset Ceph, Paris, France. [60]Department of Medicine A, University Medicine Greifswald, Greifswald, Germany. [61]Department of Genetics, Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina, USA. [62]Liver Diseases Branch, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), NIH, Bethesda, Maryland, USA. [63]Office of Biostatistics Research, Division of Cardiovascular Sciences, National Heart, Lung and Blood Institute (NHLBI), NIH, Bethesda, Maryland, USA. [64]Queensland Institute of Medical Research, Brisbane, Queensland, Australia. [65]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA. [66]Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA. [67]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA. [68]Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [69]NHLBI Framingham Heart Study, Framingham, Massachusetts, USA. [70]Department of Clinical Biochemistry, Landspitali University Hospital, Reykjavik, Iceland. [71]Department of Psychiatry and EMGO Institute for Health and Care Research, VUA Medical Centre, Amsterdam, The Netherlands. [72]Department of Psychiatry, Leiden University Medical Centre, Leiden, The Netherlands. [73]Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [74]Department of Lifecourse and Services, National Institute for Health and Welfare, Oulu, Finland. [75]Institute of Diagnostics, Clinical Chemistry, University of Oulu, Oulu, Finland. [76]Department of Internal Medicine and Biocenter Oulu, Clinical Research Center, University of Oulu, Oulu, Finland. [77]Laboratory of Genetics, National Institute on Aging, Baltimore, Maryland, USA. [78]Gesellschaft für Arteriosklerose forschung, Leibniz-Institut für Arteriosklerose forschung an der Universität Münster, Münster, Germany. [79]Department of Immunology, University of Toronto, Toronto, Ontario, Canada. [80]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. [81]Mount Sinai Hospital Samuel Lunenfeld Research Institute, Toronto, Ontario, Canada. [82]Toronto General Research Institute, Toronto, Ontario, Canada. [83]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [84]Center for Statistical Genetics, Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan, USA. [85]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands. [86]Department of Endocrinology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [87]LifeLines Cohort Study and Biobank, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [88]Section of Preventive Medicine and Epidemiology, Boston University School of Medicine, Boston, Massachusetts, USA. [89]Cardiology, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA. [90]Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, Munich,

Germany. [91]Klinikum Grosshadern, Munich, Germany. [92]EMGO+ Institute, VUA Medical Center, Amsterdam, The Netherlands. [93]Samuel Lunenfeld and Toronto General Research Institutes, Toronto, Ontario, Canada. [94]Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland, USA. [95]A full list of members is given in **Supplementary Note**. [96]Center for Medical Systems Biology, Rotterdam, The Netherlands. [97]Genomics of Common Diseases, School of Public Health, Imperial College London, Hammersmith Hospital, London, UK. [98]Oxford National Institute for Health Research Biomedical Research Centre, Churchill Hospital, Oxford, UK. [99]Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [100]Obesity Research Unit, Department of Medicine, Division of Internal Medicine, Helsinki University Hospital, Helsinki, Finland. [101]The Institute for Molecular Medicine FIMM, Helsinki, Finland. [102]Hjelt Institute, Department of Public Health, University of Helsinki, Helsinki, Finland. [103]MRC-Social Genetic Developmental Psychiatry (SGDP) Centre, Institute of Psychiatry, King's College, London, UK. [104]Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [105]Faculty of Medicine, Imperial College London, London, UK. [106]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [107]Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland. [108]Division of Endocrinology, Hypertension, and Metabolism, Brigham and Women's Hospital, Boston, Massachusetts, USA. [109]Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York, USA. [110]Institute of Health Sciences, University of Oulu, Oulu, Finland. [111]Biocenter Oulu, University of Oulu, Oulu, Finland. [112]MRC–Health Protection Agency (HPA) Centre for Environment and Health, Imperial College London, London, UK. [113]These authors contributed equally to this work. Correspondence should be addressed to J.C.C. (john.chambers@ic.ac.uk), P.E. (p.elliott@ic.ac.uk) or J.S.K. (j.kooner@ic.ac.uk).

ONLINE METHODS

Participants. Genome-wide association was done among 61,089 participants from the following published studies: the Australian Twin cohort (n = 425)[48]; the British Genetics of Hypertension study (BRIGHT, n = 1,955)[49]; the Lausanne Cohort (CoLaus, n = 5,636)[50]; deCODE genetics (n = 12,572)[51]; the Fenland study (n = 1,397)[52]; the Finnish Twin cohort study (FinnTwin, n = 32)[53]; the Framingham Heart Study (n = 2,869)[54]; the Monica/KORA Augsburg study (KORA, n = 1,809)[55]; the London Life Sciences Population study (LOLIPOP, n = 10,338)[31]; the Northern Finland Birth Cohort 1966 (NFBC1966, n = 4,562)[32]; the Netherlands Study of Depression and Anxiety (NESDA, n = 1,724)[56]; the Netherlands Twin study (n = 1,721)[57]; the Precocious Coronary Artery Disease study (Procardis, n = 1,239)[58]; the Rotterdam Study 1 (RS1, n = 4,312)[59]; the SardiNIA study (n = 4,302)[60]; the Study of Health in Pomerania (SHIP, n = 4,101)[61] and the TwinsUK study (n = 2,256)[62]. Sample sizes for ALT, ALP and GGT genome-wide analyses were 45,596, 56,415 and 61,089, respectively. Further characteristics of the genome-wide association cohorts are listed in **Supplementary Note** and **Supplementary Tables 1** and **2**. SNPs showing equivocal association with liver markers were further tested among 12,139 participants from the LOLIPOP study, with none included in the genome-wide study (**Supplementary Table 4**).

Genotyping and quality control. Genome-wide association scans were done using Affymetrix, Illumina and Perlegen Sciences arrays (**Supplementary Table 3**). Imputation of missing genotypes was done using phased haplotypes from HapMap build36 and dbSNP build 126. Imputed SNPs with minor allele frequency < 0.01 or low-quality score (r^2 < 0.30 in MACH, or information score <0.3 in IMPUTE) were removed. This generated ~2.6 million directly genotyped or imputed autosomal SNPs. Genotyping for further testing was done by KASPar (K-Biosciences, LTD).

Statistical analysis. Plasma concentrations of ALT, ALP and GGT were $\log_{10}$ transformed to achieve approximate normality. SNPs were tested for association with liver markers by linear regression using an additive genetic model adjusted for age and sex. An additional term was included to indicate case status in case-control studies, and principal component scores (EIGENSTRAT[63]) were used to adjust for substructure in studies of unrelated individuals (**Supplementary Table 3**). Test statistics were corrected for respective genomic control inflation factor (**Supplementary Table 4**) to adjust for residual population structure. Association analyses were carried out separately in each cohort followed by meta-analysis using weighted z scores. Meta-analysis P values were then corrected for the meta-analysis genomic control inflation factors. The GWAS had 80% power to detect SNPs associated with 0.1% of population variation in ALP and 0.06% of population variation in ALT and GGT at $P < 5 \times 10^{-7}$.

In the replication samples, SNP associations were tested by linear regression using an additive genetic model and adjustment for age and sex. Results were combined with findings from the genome-wide association cohorts, using the weighted z scores. Genome-wide significance was inferred at $P < 1 \times 10^{-8}$.

SNP effect sizes were estimated by inverse-variance meta-analysis in the genome-wide association cohorts and available replication cohorts using a fixed effects model.

Coding variant analyses. We identified coding SNPs within 1 Mb and in LD at $r^2 > 0.5$ with the sentinel liver SNPs using HapMap CEU II genotype data (see URLs). We tested for enrichment by permutation testing using 42 randomly selected SNPs from the ~2.6 million genotyped or imputed SNPs studied that had similar minor allele frequency ±0.02), number of nearby genes (±10%) and gene proximity (±20 kb) to the sentinel SNPs. We counted coding SNPs within 1 Mb and in LD at $r^2 > 0.5$ of the random SNPs; this was repeated 1,000 times to generate a distribution for expected, against which we compared the number observed (n = 19, P = 0.004).

We considered a coding SNP to be a strong candidate for the observed association when it was in LD at $r^2 > 0.5$ with the sentinel SNP, with no evidence for heterogeneity of effect on phenotype ($P > 0.05$). Using this approach, we identified 17 coding SNPs in 14 genes as candidates for mediating the observed associations with liver markers (**Supplementary Table 10**). We used PHYRE[64] to model the molecular structure of the protein products and possible pathogenicity of the coding SNPs identified.

Expression analyses. The sentinel SNPs from the liver marker GWAS were tested for association with gene expression in 603 adipose and 745 peripheral blood samples from Icelandic subjects[25], peripheral blood lymphocytes from 206 families of European descent (830 parents and offspring)[23] and 960 human liver samples[24]. Sentinel SNPs were tested for association with transcript levels of genes within 1 Mb; significance was inferred at $P < 0.05$ after Bonferroni correction for number of SNP-transcript combinations tested. We then used the whole-genome genotype data to identify which SNP from the liver locus was most closely associated with the transcript of interest; we defined this as the transcript SNP. We tested whether the sentinel SNP and transcript SNP were coincident, defined as in LD at $r^2 > 0.5$, with no evidence for heterogeneity of effect between the SNPs on transcript expression or liver marker phenotype.

GRAIL. We carried out a PubMed literature analysis using GRAIL (see URLs)[65] including all 42 sentinel SNPs simultaneously. We used the 2006 PubMed data set as the primary analysis (**Supplementary Table 15**) but repeated the analysis using the 2010 PubMed data set.

Network analyses. Network analyses were carried out using the Ingenuity Pathway Analysis tool[66]. P values for canonical pathways and functions were calculated from the observed number of candidate genes in the gene set, compared with the number expected under the null hypothesis and corrected (Bonferroni) for the number of pathways tested.

Overlap with other GWAS. We used the NHGRI[30] catalog (see URLs) to identify other phenotypic associations ($P < 5 \times 10^{-8}$) located within 1 Mb of a the SNPs we identified as associated with liver enzymes (**Supplementary Table 16**). Previous studies reporting genetic variants influencing concentrations of liver enzymes in plasma were excluded. Pairwise LD with the sentinel liver marker SNP was determined using HapMap 2 CEU genotype data.

Phenotypic pleiotropy. Relationships of the selected 42 sentinel SNPs with anthropometric and metabolic traits relevant to liver function were tested in the following genome-wide meta-analyses (**Supplementary Table 17**): AlcGen Consortium, alcohol consumption[67]; ICBP-GWAS, systolic and diastolic blood pressure[68]; the Genetics of C-reactive Protein Study (CRP-Gen), C-reactive protein[19]; MAGIC, fasting glucose and related glycemic traits[16]; DIAGRAM+ Study, type 2 diabetes[17]; GIANT Consortium, body mass index[69] and the Global Lipids Genetics Consortium, total cholesterol, LDL cholesterol, HDL cholesterol and triglyceride concentrations[70]. Associations were tested *in silico* using results from the genome-wide association phase and adopting the phenotypic definitions applied in each study. We inferred association of SNP with phenotype at $P < 0.0012$, corresponding to $P < 0.05$ after Bonferroni correction for 42 loci. We tested whether phenotypes were enriched for association with liver marker SNPs using a binomial probability test.

Metabonomic analyses. We carried out quantitative NMR spectroscopy on serum samples from 2,269 LOLIPOP and 4,247 NFBC1966 participants with genome-wide data to investigate the relationships of the identified loci with lipoprotein and intermediary metabolism. NMR assays were carried out using a Bruker AVANCE III spectrometer operating at 500.36 MHz (^{1}H observation frequency; 11.74 T) and equipped with an inverse selective SEI probehead including an automatic tuning and matching unit and a z-axis gradient coil for automated shimming[71,72]. A BTO-2000 thermocouple was used for temperature stabilization of the sample at ~0.01 °C. The high-performance electronics enabled metabolite quantification without per-sample chemical referencing or double-tube systems. The NMR methodology provides information on lipoprotein subclass distribution and lipoprotein particle concentrations, low-molecular-mass metabolites such as amino acids, 3-hydroxybutyrate and creatinine, and detailed molecular information on serum lipids including free and esterified cholesterol, sphingomyelin, saturation, unsaturation, polyunsaturation and omega-3 fatty acids[73]. Associations of SNPs with metabolic measures were tested in each cohort separately using an additive genetic model and were adjusted for age, gender and principal components. Results for

LOLIPOP and NFBC1966 were combined by inverse variance meta-analysis, and significance was inferred at $P < 1 \times 10^{-5}$ (corresponding to $P < 0.05$ after Bonferroni correction for the 42 independent SNPs tested and for 69 primary NMR measures).

Contribution of genetic loci identified to population variation in liver enzymes. This was investigated in the LifeLines Cohort Study[35], a prospective population-based cohort study of 165,000 persons aged 18–90 living in The Netherlands, and independent of the genome-wide association discovery cohorts. Genotyping was carried out in representative samples of 8,112 participants (aged 47.8 ± 11.2, body mass index 26.2 ± 4.3 kg/m^2 (mean ± s.d.), 43% male) using the Illumina CytoSNP12 array, and imputation of missing HapMap2 genotypes was done using Beagle 3.1.0. Liver markers were measured on a Roche/Hitachi Modular System (Roche Diagnostics). Mean ± s.d. concentrations of liver markers were 23.8 ± 16.8, 62.8 ± 18.4 and 26.3 ± 24.5 IU/l for ALT, ALP and GGT, respectively. The contribution of SNPs to population variation in liver markers was examined individually and in aggregate (**Supplementary Tables 18** and **19**). For the latter, SNP scores were calculated for each individual on the basis of the sum of effect (trait-raising) alleles present at each of the genetic loci identified.

Liver imaging for hepatic steatosis. Hepatic steatosis was assessed by CT scanning in 9,610 participants from four population cohorts primarily designed for investigation of cardiovascular disease and its risk factors, (i) AGES-Reykjavik ($n = 4,772$), (ii) the Amish study ($n = 541$), (iii) the Family Heart Study ($n = 886$) and (iv) the Framingham Study ($n = 3,411$)[36]. CT measurements, blind to participant characteristics, were calibrated against phantoms and inverse normally transformed. Genome-wide SNP data were available in each cohort with imputation of missing genotypes. SNP association with hepatic steatosis was tested in each cohort separately by linear regression with age, with age^2 and gender as covariates and taking relatedness into account. Results were combined by fixed-effect inverse-variance meta-analysis (**Supplementary Table 20**).

48. Heath, A.C. *et al.* Genetic and environmental contributions to alcohol dependence risk in a national twin sample: consistency of findings in women and men. *Psychol. Med.* **27**, 1381–1396 (1997).
49. Wallace, C. *et al.* Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. *Am. J. Hum. Genet.* **82**, 139–149 (2008).
50. Firmann, M. *et al.* The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc. Disord.* **8**, 6 (2008).
51. Kong, A. *et al.* Parental origin of sequence variants associated with complex diseases. *Nature* **462**, 868–874 (2009).
52. Watkinson, C., van Sluijs, E.M., Sutton, S., Marteau, T. & Griffin, S.J. Randomised controlled trial of the effects of physical activity feedback on awareness and behaviour in UK adults: the FAB study protocol. *BMC Public Health [ISRCTN92551397]* **10**, 144 (2010).
53. Kaprio, J. Twin studies in Finland 2006. *Twin Res. Hum. Genet.* **9**, 772–777 (2006).
54. Levy, D. *et al.* Genome-wide association study of blood pressure and hypertension. *Nat. Genet.* **41**, 677–687 (2009).
55. Löwel, H. *et al.* The MONICA Augsburg surveys–basis for prospective cohort studies. *Gesundheitswesen* **67** (suppl 1), S13–S18 (2005).
56. Lamers, F *et al.* Comorbidity patterns of anxiety and depressive disorders in a large cohort study: the Netherlands Study of Depression and Anxiety (NESDA). *J. Clin. Psychiatry* **72**, 341–348 (2011).
57. Boomsma, D.I. *et al.* Netherlands Twin Register: from twins to twin families. *Twin Res. Hum. Genet.* **9**, 849–857 (2006).
58. Clarke, R. *et al.* Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N. Engl. J. Med.* **361**, 2518–2528 (2009).
59. Hofman, A. *et al.* The Rotterdam Study: 2010 objectives and design update. *Eur. J. Epidemiol.* **24**, 553–572 (2009).
60. Scuteri, A. *et al.* Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genet.* **3**, e115 (2007).
61. Haring, R. *et al.* Prediction of metabolic syndrome by low serum testosterone levels in men: results from the study of health in Pomerania. *Diabetes* **58**, 2027–2031 (2009).
62. Spector, T.D. & MacGregor, A.J. The St. Thomas' UK Adult Twin Registry. *Twin Res.* **5**, 440–443 (2002).
63. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
64. Bennett-Lovsey, R.M., Herbert, A.D., Sternberg, M.J. & Kelley, L.A. Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins* **70**, 611–625 (2008).
65. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5**, e1000534 (2009).
66. Jimenez-Marin, A., Collado-Romero, M., Ramirez-Boo, M., Arce, C. & Garrido, J.J. Biological pathway analysis by ArrayUnlock and Ingenuity Pathway Analysis. *BMC Proc.* **3** (suppl. 4), S6 (2009).
67. Schumann, G. *et al.* Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. *Proc. Natl. Acad. Sci. USA* **108**, 7119–7124 (2011).
68. Newton-Cheh, C. *et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nat. Genet.* **41**, 666–676 (2009).
69. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
70. Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
71. Inouye, M. *et al.* Metabonomic, transcriptomic, and genomic variation of a population cohort. *Mol. Syst. Biol.* **6**, 441 (2010).
72. Soininen, P. *et al.* High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. *Analyst* **134**, 1781–1785 (2009).
73. Würtz, P. *et al.* Characterization of systemic metabolic phenotypes associated with subclinical atherosclerosis. *Mol. Biosyst.* **7**, 385–393 (2011).