Table 3. Associations of two SNPs (rs2395309, rs9277535) with HBV infection and clearance in Han Chinese populations.

	South of china		North of china	
	Control group	**Case group**	**Control group**	**Case group**
HLA-DPB1 (rs2395309)- dominant model (AA+AGvsGG)				
AA/AG/GG	57/234/288†	112/709/1367‡	52/193/138†	63/249/302‡
P value OR (95%CI)	Reference	3.36×10^{-8} 0.57 (0.47,0.70)	Reference	1.23×10^{-4} 0.50 (0.35,0.71)
AA/AG/GG	112/709/1367‡	35/235/257‖	63/249/302‡	56/130/121‖
P value OR (95%CI)	Reference	9.63×10^{-7} 1.31 (1.17,1.45)	Reference	0.021 1.20 (1.03,1.40)
HLA-DPA1 (rs9277535)- dominant model (AA+AGvsGG)				
AA/AG/GG	80/277/216†	177/830/1195‡	97/203/80†	118/287/206‡
P value OR (95%CI)	Reference	2.68×10^{-10} 0.52 (0.43,0.64)	Reference	1.74×10^{-5} 0.50 (0.36,0.68)
AA/AG/GG	177/830/1195‡	67/251/208‖	118/287/206‡	67/165/75‖
P value OR (95%CI)	Reference	1.67×10^{-7} 1.33 (1.20,1.49)	Reference	8.37×10^{-3} 1.26 (1.06,1.49

†Healthy control group.
‖HBV clearance group.
‡HBV infection groups, including Asymptomatic HBV carriers, Chronic active hepatitis B group, HBV-related liver cirrhosis group, HBV-related heptocellular carcinoma group.
The *P* values, odds ratios (OR), and 95% confidence intervals (CI) were calculated on the basis of the binary logistic regression analysis, adjusted for sex and age.
doi:10.1371/journal.pone.0024221.t003

HBV-related liver cirrhosis group (OR = 1.11; 95% CI : 0.82, 1.52; $P=0.499$ at rs2395309; OR = 1.24; 95% CI : 0.92, 1.67; $P=0.163$ at rs9277535, in southern Chinese population; OR = 0.74; 95% CI : 0.48, 1.16; $P=0.189$ at rs2395309; OR = 1.29; 95% CI : 0.81, 2.06; $P=0.286$ at rs9277535, in northern Chinese population) and HBV-related heptocellular carcinoma group(OR = 0.85; 95% CI : 0.63, 1.16; $P=0.305$ at rs2395309; OR = 0.98; 95% CI : 0.73, 1.31; $P=0.881$ at

Table 4. Results of the association test for two SNPs(rs2395309,rs9277535) haplotypes in Han Chinese populations.

South of china						
Haplotype	**Health(2n = 1106)**	**Clear(2n = 1048)**	**AsC(2n = 1342)**	**CHB(2n = 1486)**	**LC(2n = 754)**	**HCC(2n = 632)**
A-A	247(22.3)	205(17.0)	227(16.8)	238(16.0)	122(16.2)	92(14.5)
A-G	88(8.0)	97(11.8)	62(4.6)	95(6.4)	36(4.8)	23(3.7)
G-A	173(15.6)	177(19.5)	140(10.4)	143(9.6)	95(12.6)	78(12.4)
G-G	598(54.1)	569(51.7)	913(68.1)	1010(68.0)	501(66.4)	439(69.4)
P value‖	Reference		1.47×10^{-6}	6.47×10^{-8}	2.53×10^{-5}	6.07×10^{-7}
OR (95%CI)			0.60 (0.49,0.74)	0.57 (0.47,0.70)	0.59 (0.46,0.76)	0.51 (0.39,0.66)
P value‡‖		Reference	7.35×10^{-4}	8.92×10^{-5}	2.45×10^{-3}	1.06×10^{-4}
OR (95%CI)			1.45 (1.17,1.80)	1.53 (1.24,1.89)	1.48 (1.15,1.91)	1.72 (1.31,2.27)
North of china						
Haplotype	**Health(2n = 734)**	**Clear(2n = 608)**	**AsC (2n = 422)**	**CHB (2n = 378)**	**LC(2n = 300)**	**HCC(2n = 100)**
A-A	226(30.8)	200(31.5)	118(28.0)	103(27.3)	71(23.6)	21(20.5)
A-G	55(7.5)	41(8.1)	27(6.4)	18(4.8)	16(5.4)	4(3.5)
G-A	157(21.4)	98(17.5)	66(15.6)	60(15.9)	62(20.7)	15(15.4)
G-G	296(40.3)	269(42.9)	211(50.0)	197(52.1)	151(50.3)	60(60.5)
P value‖	Reference		0.032	0.012	0.004	0.003
OR (95%CI)			0.73 (0.55,0.97)	0.68 (0.51,0.92)	0.62 (0.44,0.86)	0.46 (0.27,0.78)
P value‡‖		Reference	0.054	0.021	0.007	0.005
OR (95%CI)			1.33 (0.99,1.78)	1.42 (1.05,1.92)	1.58 (1.13,2.21)	2.12 (1.25,3.61)

‖Two SNPs haplotypes G-G, A-A in Health group compared with those in HBV infection groups.
‡‖Two SNPs haplotypes G-G, A-A in HBV infection groups compared with those in Clearance group.
The *P* values, odds ratios (OR), and 95% confidence intervals (CI) were calculated by Pearson Chi-Square test.
Abbreviations: Clear, HBV clearance group; Health, Healthy control group; AsC, Asymptomatic HBV carriers group; CHB, Chronic active hepatitis B group; LC,HBV-related liver cirrhosis group; HCC, HBV-related heptocellular carcinoma group; OR, odds ratio; CI, confidence interval.
doi:10.1371/journal.pone.0024221.t004

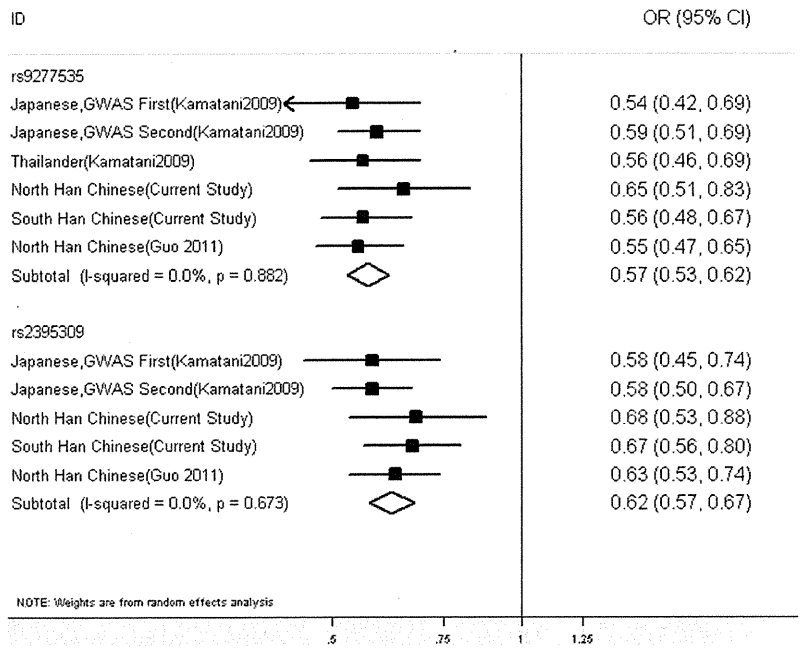

Figure 1. Meta-analasis of the rs9277535 and rs2395309. The meta-analysis combined with the results of previous studies, including more than 2,243 cases and 4,137 controls. Each effect size is shown with its confidence interval. Abbreviations: p, *P* heterogeneity value; OR, odds ratios; 95%CI, 95% confidence interval.
doi:10.1371/journal.pone.0024221.g001

rs9277535, in southern Chinese population; OR = 0.56; 95% CI : 0.28, 1.11; $P=0.097$ at rs2395309; OR = 0.84; 95% CI : 0.42, 1.68; $P=0.615$ at rs9277535, in northern Chinese population), compared with asymptomatic HBV group(Table S5.).

Associations of the *HLA-DP* loci polymorphisms with clinical factors

In order to analyze the associations between two SNPs and clinical factors (HBV-DNA load, ALT and TB), we used the independent-sample Kolmogorov-Smirnov t test in CHB group, LC group and HCC group. Although the GG patients have a higher mean on the HBV-DNA load, no significant difference was found between patients of different genotypes (see Fig. S1). In the analysis of ALT, the associations between two SNPs and the ALT level only be found in HBV-related liver cirrhosis group ($P=0.002$ at rs2395309; $P=0.009$ at rs9277535), rather than in other groups. Meanwhile, for the associations of the TB level, there was no difference between GG patients and AG+AA patients ($P>0.05$ in each group).

Results of the Haplotype analysis and Meta-analysis

To further understand the contributions of these loci to HBV susceptibility, two-locus haplotypes were constructed for two SNPs rs2395309 and rs9277535 (Table 4.). Pairwise linkage disequilibrium (LD) analyses performed using all individuals from the health group showed that rs2395309 and rs9277535 SNPs were in LD with each other ($D' = 0.57$, $r^2 = 0.23$ in southern Chinese population; $D' = 0.58$, $r^2 = 0.20$ in northern Chinese population). In trying to derive HBV infection-specific haplotypes, the haplotype frequencies of two SNPs (rs2395309 and rs9277535) were evaluated in both Chinese populations. Four haplotypes were observed, and among them three haplotypes had frequencies more than 5% (Table 4.). Compared with protective A-A haplotype homozygotes, only G-G haplotype homozygotes had a significant increased risk for HBV infection (*P* value and odds ratios were shown in Table 4). Then, we summarized a meta-analysis combined with the results of related studies [12,17], including more than 2,243 cases and 4,137 controls. As shown in Figure 1 and Table S6, these odds ratios were quite similar among the three ethnic groups (Japanese, Thais and Chinese) and no heterogeneity was observed (P het = 0.673 at rs2395309; P het = 0.882 at rs 9577535).

Discussion

In this analysis, we confirmed that two SNPs sites (rs2395309 and rs9277535) in the *HLA-DPA1* and *HLA-DPB1* genes were significantly associated with HBV infection in southern and northern Han Chinese populations. Again, our haplotype analysis showed the frequency of G-G haplotype had a significant increase in the HBV infected populations, as compared with the healthy control group or HBV clearance group. As a result, we inferred that these persons with G-G haplotype have a higher risk of HBV infection than those persons with A-A haplotype. Meanwhile, the A-A haplotype could be strongly predictive for HBV clearance in HBV infection populations. Although our manuscript suggested that the genotype distributions of both sites (rs2395309 and rs9277535) were different between southern and northern Chinese population, the frequencies of two protective alleles A in Chinese populations were also similar to those in Asian populations, compared with European and Central American populations (data from public databases, HapMap). The results of the genetic association in our study were consistent with the previous study [12]. Hence, we could confirm that the polymorphisms of *HLA-DPA1* and *HLA-DPB1* gene play a very important role in chronic hepatitis B virus infection in southern and northern Han Chinese populations.

It has been well documented that men are more likely than women to be infected with HBV and develop liver cirrhosis and hepatocellular carcinoma [18,19]. The reasons for the gender distinction between HBV populations and health populations are complex, including occupation, alcohol drinking, tobacco smoking, family history of HBV infection and so on. Some previous reports suggested that sex hormones might interact with HBV in the infection process and lead to a dominant sex disparity in HBV populations. Naugler et al. [20] found that estrogen-mediated inhibition of interleukin-6 production by Kupffer cells reduced the risk of liver cancer in females. Wang et al. [21] study demonstrated that the androgen pathway could increase the transcription of HBV through direct binding to the androgen-responsive element sites in viral enhancer. Consequently, to decrease the bias of sex in population sampling, we further conducted the stratified analysis for sex. Although we found that male and female northern Chinese showed a different susceptibility to HBV infection, it only had 25% and 21% statistical power to detect these ORs of 0.73 and 0.74, which may lead to the false-negative results of rs2395309 and rs9277535 in northern female Chinese. The small sample for female HBV patients in this study might be the major reason for the non-significant associations in female Chinese. Hence, we only concluded that the genetic variants of *HLA-DPA1* and *HLA-DPB1* loci differ slightly between male and female Chinese, and the reasons why there is different between male and female for HBV infection need to be further studied.

And indeed, by consulting previous studies [22,23], we found that there are different distributions in some *HLA* alleles among Han Chinese populations. For instance, *HLA-DRB1**0301 [8], a risk-allele with respect to chronic HBV infection in Han Chinese, markedly has higher frequency in southern Han Chinese population than those in northern Han Chinese population. Since the frequency distribution of *HLA-DP* alleles were barely reported in China, it could be inferred only indirectly that there were also different distributions at *HLA-DP* alleles between two Han Chinese populations. And, it was the difference that led to the distinct distributions of both SNPs (rs9277535 and rs2395309) between southern and northern Han Chinese population. Nevertheless, this explanations why the distributions of the *HLA* alleles (or SNPs) differed between Han Chinese populations were complicated, such as evolution and migration history of the Chinese population [24,25,26], MHC-based mate choice [27], pathogen-driven selection at *HLA* alleles [28,29] and so on. Taking into account the different distributions of HBV genotypes [30] and HBV carrier rate [31] in China, as well as recent studies [12,17] and our results, we deduced that the mechanism of pathogen-driven selection (HBV and/or other pathogens) might be the leading cause of the different distributions at *HLA-DP* alleles between two Han Chinese populations.

Moreover, after infection with hepatitis B virus (HBV), the host's inflammatory immune response induces hepatocellular damage and is followed by the pathogenesis of liver cirrhosis and cancer [32]. Liver cancer arises most frequently in the setting of chronic liver inflammation [33]. Considering the function of *HLA-DP* molecules, HBV antigen presentation on *HLA-DP* molecules may be critical for virus elimination and has an important role in the progression of hepatitis B [34]. Therefore, we further analysed the possible association between the polymorphisms in *HLA-DP* gene and the disease progression of chronic hepatitis B. Unfortunately, compared with asymptomatic HBV carrier, there were no associations in chronic active hepatitis B group, HBV-related liver cirrhosis group and HBV-related hepatocellular carcinoma group. Although chronic HBV infection is the most important cause of HCC worldwide and contributes to at least 70% of cases of HCC in Asian-Africa [35], only a tiny fraction of chronic HBV carriers develop HCC in their lifetime [36]. It is suggested that the risk of HCC is caused by a complex interplay between multiple genetic and environmental factors. Recently, Zhang et al. have conducted the first liver GWAS for HCC in Chinese ancestry and identified a single susceptibility locus in the UBE4B-KIF1B-PGD region on 1p36.22 [37]. Since the region involve in these aspects of vesicles transport, cell apoptosis, DNA repair, and other intracellular pathways, it seems likely that different genes play disparate roles in HBV infection and HBV progression. For example, immune pathway (*HLA-DP* or other genes) is the primary cause of HBV infection, but intracellular pathway (Ubiquitin or other pathways) is the major reason of HBV progression. Thus, by combining our results with the aforementioned discussion, we inferred that the polymorphisms in *HLA-DPA1* and *HLA-DPB1* gene influence the infection of HBV in Chinese populations, rather than the progression of HBV disease.

Since the early 1970s [38], classical human leukocyte antigen loci have stood out as the leading candidates for infectious disease susceptibility. The classical *HLA* loci are the class I (*HLA-A*, *-B*, *-C*, *-E*, *-F*, and *-G*) and class II (*HLA-DR*, *-DQ*, *-DM*, and *-DP*) molecules. *HLA* class II molecules are the central part in the immune system by presenting peptides to the antigen receptor of CD4+ T cells [39]. Antigen presentation is not only crucial for the regulation of protective immune responses against invading pathogens, but also necessary for the maintenance of self-tolerance. It is therefore perhaps not surprising to find that the human MHC class II gene region holds the largest number, and some of the longest recognised, associations with a autoimmune, inflammatory and infectious diseases [40,41]. Although *HLA-DPs* have a structure similar to other classical *HLA* class II molecules, *HLA-DP* molecule roles in the immune response have not been well characterized until now. In a previous study, Hirayama et al. [42] indicated that the *HLA* class II genes for the *HLA-DR-DQ* alleles were associated with protection against early changes in liver fibrosis, whereas *HLA-DP* alleles were associated with protection from the late phase of schistosomal hepatic fibrosis. Owing to lack of replication of the previously report, more studies are essential to provide conclusive genetic and functional evidence to support a role for *HLA-DP* in HBV disease susceptibility.

In summary, in this multicenter case-control study, we have confirmed that the G alleles of two SNPs sties in the *HLA-DPA1* and *HLA-DPB1* were significantly associated with hepatitis B virus (HBV) infection in Han Chinese populations, and both A alleles (rs2395309 and rs9277535) also showed a strong protective effect for HBV clearance. Furthermore, we found that the genotype distributions of both sites (rs2395309 and rs9277535) were clearly different between southern and northern Han Chinese population. By using asymptomatic HBV carrier as control group, our study showed that there were no associations of *HLA-DP* variants (rs2395309 and rs9277535) with HBV progression. Although HBV disease is not determined solely by genetic factors, the experimental results offer the foundation for further study of genetic variations in the *HLA-DPA1* and *HLA-DPB1* for the prevention and therapy of chronic HBV infection.

Supporting Information

Figure S1 Associations of these two SNPs (rs2395309, rs9277535) genotypes with HBV DNA levels. *P* values of independent-sample Kolmogorov-Smirnov t test for dominant model (AA+AG vs GG). Abbreviations:SNPs, single nucleotide polymorphisms.
(TIF)

Table S1 Diagnosis criteria for Healthy control group (Health), HBV clearance group (Clear), Asymptomatic chronic HBV carriers group (AsC), Chronic active hepatitis B group (CHB), HBV-related liver cirrhosis group (LC) and HBV-related heptocellular carcinoma group (HCC).
(DOC)

Table S2 TaqMan probes and Primers for two SNPs (rs2395309 and rs9277535).
(DOC)

Table S3 The stratified analysis of gender between two SNPs (rs2395309, rs9277535) genotypes and different populations. Male and female patients showed different genotype distributions in these two SNPs (rs2395309 and rs9277535), specially in the northern Chinese population. The *P* values, odds ratios (OR), and 95% confidence intervals (CI) were calculated on the basis of the binary logistic regression analysis, adjusted for age.
(DOC)

Table S4 The stratified analysis of age between two SNPs (rs2395309, rs9277535) genotypes in south Chinese population and north Chinese population. Most cases were no significant difference in genotype distributions of two SNPs sites between patients with age≤45 years and patients with age>45 years. The *P* values, odds ratios (OR), and 95% confidence intervals (CI) were calculated on the basis of the binary logistic regression analysis, adjusted for sex.
(DOC)

Table S5 Associations of two SNPs (rs2395309, rs9277535) with HBV progression in Han Chinese populations. Compared with asymptomatic HBV group, those two sites (rs2395309 and rs9277535) in HLA-DPA1 or HLA-DPB1 gene had no associations with the chronic active hepatitis B, the HBV-related liver cirrhosis, and the HBV-related heptocellular carcinoma in southern and northern Chinese population.
(DOC)

Table S6 A Meta-analysis for previous study and current study (more than 2,243 cases and 4,137 controls). Genotype distributions of rs9277535 and rs2395309 in three ethnic groups (Japanese, Thais, Chinese) between healthy control group and chronic active hepatitis B group. *P* values of Pearson's x^2 test for allele model. Odds ratios (OR) and 95% confidence intervals (CI) of minor allele from two-by-two allele frequency table.
(DOC)

Acknowledgments

We thank all subjects for their ongoing participation in this study. We also thank Dr. Xinguo Peng (The Affiliated Hospital of Binzhou Medical College), Xinjuan Kong (Qingdao Infectious Disease Hospital), Xiaoliang Li, Bin Wen, Zhiyu Li (The Third People's Hospital of Shenzhen), Lu Yuan(The Fourth People's Hospital of Shenzhen) and Yan Gao (The Sixth People's Hospital of Shenzhen) for the help in recruitment of patients; Mr. Yuncheng Zhang, Zhixian Liu and Mrs. Hong Yuan, Ying Zhang, Lizhen Yang, Juzhen Zhong (The Third People's Hospital of Shenzhen) for the laboratory assistance.

Author Contributions

Conceived and designed the experiments: J.Lin Y.Chen J.Li. Performed the experiments: J.Li MW LL J.Yao QS MC HY. Analyzed the data: J.Li Y.Chang X.He. Contributed reagents/materials/analysis tools: J.Yu J.Li Xiaorui Jiang SS QL Xiang Jiang. Wrote the paper: J.Li. Provided the unpublished data: KM YN. Subject recruitment, biological sample collection and medical records in Guangdong province: DY J.Li J.Yao QS. Subject recruitment, biological sample collection and medical records in Shandong province: ZW LL MC HY. Subject recruitment, biological sample collection and medical records in Hubei province: YH FT X.Hu JW J.Yu QS MC HY.

References

1. World Health Organization. Hepatitis B. World Health Organization Fact Sheet No. 204. Available: http://who.int/mediacentre/factsheets/fs204/en. (Revised August 2008).
2. Ganem D, Prince A (2004) Hepatitis B virus infection–natural history and clinical consequences. New England Journal of Medicine 350: 1118.
3. Lu FM, Zhuang H (2009) Management of hepatitis B in China. Chin Med J (Engl) 122: 3–4.
4. Wright TL (2006) Introduction to Chronic Hepatitis B Infection. The American Journal of Gastroenterology 101: S1–S6.
5. Blackwell JM, Jamieson SE, Burgner D (2009) HLA and Infectious Diseases. Clinical Microbiology Reviews 22: 370–385.
6. Godkin A, Davenport M, Hill AVS (2005) Molecular analysis of HLA class II associations with hepatitis B virus clearance and vaccine nonresponsiveness. Hepatology 41: 1383–1390.
7. Thursz MR, Kwiatkowski D, Allsopp CE, Greenwood BM, Thomas HC, et al. (1995) Association between an MHC class II allele and clearance of hepatitis B virus in the Gambia. N Engl J Med 332: 1065–1069.
8. Singh R, Kaul R, Kaul A, Khan K (2007) A comparative review of HLA associations with hepatitis B and C viral infections across global populations. WORLD JOURNAL OF GASTROENTEROLOGY 13: 1770.
9. Ahn SH, Han KH, Park JY, Lee CK, Kang SW, et al. (2000) Association between hepatitis B virus infection and HLA-DR type in Korea. Hepatology 31: 1371–1373.
10. Consortium TWTCC (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661–678.
11. Ober C, Thompson E (2005) Rethinking genetic models of asthma: the role of environmental modifiers. Current Opinion in Immunology 17: 670–678.
12. Kamatani Y, Wattanapokayakit S, Ochi H, Kawaguchi T, Takahashi A, et al. (2009) A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. Nature Genetics 41: 591–595.
13. Zeng Z, Guan L, An P, Sun S, O'Brien SJ, et al. (2008) A population-based study to investigate host genetic factors associated with hepatitis B infection and pathogenesis in the Chinese population. BMC Infectious Diseases 8: 1.
14. He X, Chang Y, Jiang H, Tang F, Meng F, et al. (2010) Persistent Effect of IFNAR-1 Genetic Polymorphism on the Long-Term Pathogenesis of Chronic HBV Infection. Viral Immunology 23: 251–257.
15. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10: 564–567.
16. Xu S, Yin X, Li S, Jin W, Lou H, et al. (2009) Genomic Dissection of Population Substructure of Han Chinese and Its Implication in Association Studies. The American Journal of Human Genetics 85: 762–774.
17. Guo X, Zhang Y, Li J, Ma J, Wei Z, et al. (2011) Strong influence of human leukocyte antigen (HLA)-DP gene variants on development of persistent chronic hepatitis B virus carriers in the Han Chinese population. Hepatology 53: 422–428.
18. Prieto J (2008) Inflammation, HCC and sex: IL-6 in the centre of the triangle. Journal of Hepatology 48: 380–381.
19. Chen S, Yeh J, Chang M, Liao Y, Hsiao L, et al. (2010) Gender Difference of Alanine Aminotransferase Elevation May Be Associated with Higher Hemoglobin Levels among Male Adolescents. PLoS ONE 5: e13269.
20. Naugler W, Sakurai T, Kim S, Maeda S, Kim K, et al. (2007) Gender disparity in liver cancer due to sex differences in MyD88-dependent IL-6 production. Science 317: 121.
21. Wang S-H, Yeh S-H, Lin W-H, Wang H-Y, Chen D-S, et al. (2009) Identification of androgen response elements in the enhancer I of hepatitis B virus: A mechanism for sex disparity in chronic hepatitis B. Hepatology 50: 1392–1402.
22. Trachtenberg E, Vinson M, Hayes E, Hsu YM, Houtchens K, et al. (2007) HLA class I (A, B, C) and class II (DRB1, DQA1, DQB1, DPB1) alleles and haplotypes in the Han from southern China. Tissue antigens 70: 455–463.
23. Yang G, Deng YJ, Hu SN, Wu DY, Li SB, et al. (2006) HLA-A, -B, and -DRB1 polymorphism defined by sequence-based typing of the Han population in Northern China. Tissue antigens 67: 146–152.
24. Wen B, Li H, Lu D, Song X, Zhang F, et al. (2004) Genetic evidence supports demic diffusion of Han culture. Nature 431: 302–305.
25. Zhang F, Su B, Zhang YP, Jin L (2007) Genetic studies of human diversity in East Asia. Philos Trans R Soc Lond B Biol Sci 362: 987–995.
26. Zhao TM, Lee TD (1989) Gm and Km allotypes in 74 Chinese populations: a hypothesis of the origin of the Chinese nation. Hum Genet 83: 101–110.
27. Chaix R, Cao C, Donnelly P (2008) Is mate choice in humans MHC-dependent? PLoS Genet 4: e1000184.

28. Barreiro LB, Quintana-Murci L (2010) From evolutionary genetics to human immunology: how selection shapes host defence genes. Nat Rev Genet 11: 17–30.
29. Prugnolle F, Manica A, Charpentier M, Guegan JF, Guernier V, et al. (2005) Pathogen-driven selection and worldwide HLA class I diversity. Curr Biol 15: 1022–1027.
30. Zeng G, Wang Z, Wen S, Jiang J, Wang L, et al. (2005) Geographic distribution, virologic and clinical characteristics of hepatitis B virus genotypes in China. J Viral Hepat 12: 609–617.
31. Huang ZJ, Zhou MG, Wang LJ (2007) Study on the geographic distribution of liver cancer mortality and HBsAg carrier rate in China. Disease Surveillance (Chinese) 22: 242–245.
32. Giovanna F, Bortolotti F, Francesco D (2008) Natural history of chronic hepatitis B: Special emphasis on disease progression and prognostic factors. Journal of Hepatology 48: 335–352.
33. Yang HI, Lu SN, Liaw YF, You SL, Sun CA, et al. (2002) Hepatitis B e antigen and the risk of hepatocellular carcinoma. N Engl J Med 347: 168–174.
34. Handunnetthi L, Ramagopalan SV, Ebers GC, Knight JC (2009) Regulation of major histocompatibility complex class II gene expression, genetic variation and disease. Genes and Immunity 11: 99–112.
35. Llovet JM, Burroughs A, Bruix J (2003) Hepatocellular carcinoma. Lancet 362: 1907–1917.
36. Chen JD, Yang HI, Iloeje UH, You SL, Lu SN, et al. (2010) Carriers of inactive hepatitis B virus are still at risk for hepatocellular carcinoma and liver-related death. Gastroenterology 138: 1747–1754. e1741.
37. Zhang H, Zhai Y, Hu Z, Wu C, Qian J, et al. (2010) Genome-wide association study identifies 1p36.22 as a new susceptibility locus for hepatocellular carcinoma in chronic hepatitis B virus carriers. Nat Genet 42: 755–758.
38. Thorsby E (1974) The human major histocompatibility system. Transplant Rev 18: 51–129.
39. Pieters J (2000) MHC class II-restricted antigen processing and presentation. Adv Immunol 75: 159–208.
40. Jones EY, Fugger L, Strominger JL, Siebold C (2006) MHC class II proteins and disease: a structural perspective. Nat Rev Immunol 6: 271–282.
41. Grivennikov SI, Greten FR, Karin M (2010) Immunity, Inflammation, and Cancer. Cell 140: 883–899.
42. Hirayama K, Chen H, Kikuchi M, Yin T, Gu X, et al. (1999) HLA-DR-DQ alleles and HLA-DP alleles are independently associated with susceptibility to different stages of post-schistosomal hepatic fibrosis in the Chinese population. Tissue antigens 53: 269–274.

nature genetics

Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma

Vinod Kumar[1,2], Naoya Kato[3], Yuji Urabe[1], Atsushi Takahashi[2], Ryosuke Muroyama[3], Naoya Hosono[2], Motoyuki Otsuka[4], Ryosuke Tateishi[4], Masao Omata[4], Hidewaki Nakagawa[2], Kazuhiko Koike[4], Naoyuki Kamatani[2], Michiaki Kubo[2], Yusuke Nakamura[1,2] & Koichi Matsuda[1]

To identify the genetic susceptibility factor(s) for hepatitis C virus–induced hepatocellular carcinoma (HCV-induced HCC), we conducted a genome-wide association study using 432,703 autosomal SNPs in 721 individuals with HCV-induced HCC (cases) and 2,890 HCV-negative controls of Japanese origin. Eight SNPs that showed possible association ($P < 1 \times 10^{-5}$) in the genome-wide association study were further genotyped in 673 cases and 2,596 controls. We found a previously unidentified locus in the 5′ flanking region of *MICA* on 6p21.33 (rs2596542, $P_{combined} = 4.21 \times 10^{-13}$, odds ratio = 1.39) to be strongly associated with HCV-induced HCC. Subsequent analyses using individuals with chronic hepatitis C (CHC) indicated that this SNP is not associated with CHC susceptibility ($P = 0.61$) but is significantly associated with progression from CHC to HCC ($P = 3.13 \times 10^{-8}$). We also found that the risk allele of rs2596542 was associated with lower soluble MICA protein levels in individuals with HCV-induced HCC ($P = 1.38 \times 10^{-13}$).

It is estimated that more than 170 million people are infected with HCV worldwide[1]. Persistent HCV infection causes CHC and, subsequently, fatal liver diseases such as liver cirrhosis and HCC. Therefore, the treatment of HCV carriers is an issue of global importance. HCC is the third most common cause of cancer-related deaths[2], and HCV infection accounts for 30–70% of the individuals with HCC[3,4]. HCV-induced HCC is a multistep and progressive liver disease in which disease progression may be influenced by both environmental and genetic risk factors. The impact of host genetic variation on progression to CHC after HCV exposure is well documented by recent genome-wide association studies (GWAS)[5–7]. However, no comprehensive analyses have been performed to explore the genetic basis of HCV-induced HCC. Therefore, we conducted a GWAS for HCV-induced HCC.

We genotyped the DNA of 721 individuals with HCV-induced HCC and 2,890 HCV-negative controls (**Supplementary Table 1**) from BioBank Japan[8]. After the initial standard SNP quality filters, we obtained genotyping results for 432,703 SNPs for association analysis. Because progression from CHC to liver cancer is strongly affected by age and gender[3], we performed a logistic regression analysis by including age and gender as covariates at all tested loci in our analyses. The genetic inflation factor (λ) was 1.03, indicating that there is no or little population stratification (**Supplementary Fig. 1**). Although no SNPs cleared the GWAS significance threshold ($P < 5 \times 10^{-8}$) at this stage, we identified eight independent loci showing possible association ($P < 1 \times 10^{-5}$; **Supplementary Fig. 2**).

In the replication stage, 673 cases from an independent HCC cohort from the University of Tokyo and 2,596 HCV-negative controls from BioBank Japan were genotyped at these eight SNPs. We observed a significant replication of association at rs2596542 on chromosome 6p21.33 ($P = 8.62 \times 10^{-9}$, odds ratio (OR) = 1.44, 95% confidence interval (CI) 1.27–1.63; **Table 1**), whereas the remaining seven SNPs failed to replicate the association (**Supplementary Table 2**). Furthermore, the combination analysis of the GWAS and replication study data at rs2596542 revealed a highly significant association in which the frequency of the risk allele A is higher in cases ($P = 4.21 \times 10^{-13}$, OR = 1.39; **Fig. 1** and **Table 1**) after the age and gender adjustment, without any heterogeneity ($P = 0.24$) between the two stages. To further investigate the impact of rs2596542 on the complex nature of the HCV-induced HCC phenotype, we genotyped 1,730 individuals with CHC who had not developed liver cirrhosis or HCC during their recruitment. As a result, rs2596542 was found to have no association with chronic hepatitis C susceptibility ($P = 0.61$) but was significantly associated with progression from CHC to HCC ($P = 3.13 \times 10^{-8}$, OR = 1.36; **Table 2**).

Because heavy alcohol consumption (>50 g per day) as well as poor response to interferon (IFN) treatment were shown to be the major risk factors for HCC among individuals with CHC[9], we evaluated the effect of alcohol consumption as a confounding factor and found that rs2596542 remained highly significant even after adjustment for this factor (non-HCV versus HCC, OR = 1.39, $P = 1.22 \times 10^{-11}$; CHC versus HCC, OR = 1.25, $P = 2.31 \times 10^{-4}$; **Supplementary Table 3**). The major genotypes of HCV can be determined by a serotyping

[1]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan. [2]Center for Genomic Medicine, The Institute of Physical and Chemical Research (RIKEN), Kanagawa, Japan. [3]Unit of Disease Control Genome Medicine, The Institute of Medical Science, University of Tokyo, Tokyo, Japan. [4]Department of Gastroenterology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan. Correspondence should be addressed to K.M. (koichima@ims.u-tokyo.ac.jp).

Received 20 July 2010; accepted 23 March 2011; published online 17 April 2011; doi:10.1038/ng.809

Table 1 Association results of rs2596542 in the GWAS, replication stage and combined analysis

SNP	Chr. (locus)	Stage	Case RAF	Control RAF	*P*	OR (95% CI)
rs2596542 (A/G)	6 (*MICA*)	GWAS[a]	0.388	0.331	4.50×10^{-6}	1.34 (1.16–1.53)
		Replication[a]	0.413	0.331	8.62×10^{-9}	1.44 (1.27–1.63)
		Combined[a]	0.400	0.331	4.21×10^{-13}	1.39 (1.27–1.52)
				MH test	7.76×10^{-12}	1.35 (1.24–1.47)

We analyzed 1,394 cases with HCC (721 in the GWAS and 673 in the replication) and 5,486 controls (2,890 in GWAS and 2,596 in replication). Chr., chromosome; RAF, risk allele frequency (allele A); OR, odds ratio for the minor allele calculated by considering the major allele as a reference; MH, Mantel-Haenszel.
[a]*P* values and ORs are adjusted for age and gender by logistic regression analysis under an additive model.

assay that is based on the type-specific antibodies produced by the infected host[10]. A subgroup analysis for HCV serotypes or history of IFN therapy indicated that this variation is associated with HCC susceptibility independently of HCV genotypes or treatment response (**Supplementary Fig. 3**). Consistent with this result, rs1051796, which had $r^2 = 0.7$ and $D' = 0.95$ with rs2596542, was not associated with IFN response ($P = 0.89$) according to previously published data in the Japanese population[11].

rs2596542 is located within the class I major histocompatibility complex (MHC) region. The human MHC region encompasses the complex and extended linkage disequilibrium (LD) structure[12,13]. Several HLA alleles and genes within MHC region have been implicated in HCV infection or clearance or in response to treatment[14–16]. Therefore, we searched the whole 7.5-Mb extended MHC region using GWAS data to test the possibility of other associated loci. We found a moderate association peak at rs9275572 ($P = 4.99 \times 10^{-5}$), which is located between *HLA-DQA* and *HLA-DQB* loci (**Supplementary Fig. 4**). Subsequent replication and combination analyses at rs9275572 indicated a significant association with HCV-induced HCC ($P = 9.38 \times 10^{-9}$, OR = 1.30; **Supplementary Table 4**). The multiple logistic regression analysis to control for alcohol consumption along with age and gender also indicated a significant association at rs9275572 ($P = 3.21 \times 10^{-8}$, OR = 1.29; **Supplementary Table 5**). However, rs2596542 was not in high LD with rs9275572 ($D' = 0.41$, $r^2 = 0.16$), and both SNPs remained associated with HCC even after conditional analysis on each other and had small reductions in their ORs upon conditioned analysis (OR = 1.23, $P = 4.43 \times 10^{-6}$ and OR = 1.17, $P = 0.00059$, respectively; **Supplementary Table 6**). A haplotype analysis between these two markers showed four possible haplotypes, with haplotype AA showing higher risk (with OR = 1.44) compared to the major haplotype GG (**Supplementary Table 7**). However, the OR for the risk haplotype was 1.32 with $P = 2.31 \times 10^{-10}$ after comparing against all observed haplotypes in the population (**Supplementary Table 7**), which is weaker than that of rs2596542 alone (OR = 1.39, $P = 4.21 \times 10^{-13}$). Hence, the impact of rs2596542 is much stronger than the haplotype of two SNPs, suggesting that rs2596542 is a principal genetic factor in this region. We also found that rs9275572 has a moderate association with CHC susceptibility as well as progression from CHC to HCC ($P = 0.03$ and $P = 2.58 \times 10^{-5}$, OR = 1.09 and OR = 1.29, respectively; **Supplementary Table 8**). Because *HLA-DQ* and *HLA-DR* alleles were shown to be associated with viral persistence and early liver disease among Japanese individuals[16], further study will be needed to confirm whether the association at rs9275572 is because of its LD with *HLA-DQ* or *DR* alleles.

In this regard, it is interesting to note that rs9275572 had a very strong expression quantitative trait locus effect on *HLA-DQB1* ($\log_{10}$ odds (LOD) ≥ 19.48) and *HLA-DRB4* alleles (LOD ≥ 26.88)[17]. Thus, it will be important to test the functional effect of the common haplotype (AA; **Supplementary Table 7**), which tags the risk alleles at these two SNPs.

Two SNPs, rs12979860 and rs8099917, at the *IL28B* locus were reported to be associated with spontaneous clearance of HCV virus[18] and response to pegylated IFN-α and ribavirin therapy[11], respectively. However, we found no association at rs12979860 and rs8099917 in our dataset (**Supplementary Table 9**). Because we used non-HCV control subjects rather than subjects who had cleared HCV infection spontaneously, and because only about 20% of the cases with HCC had been treated with IFN, our study may not be suitable to detect associations at the *IL28B* locus. In addition, the protective C allele at rs12979860 is nearly fixed throughout east Asia, with a frequency of more than 91% in the Japanese population as compared to 67% in European Americans[6], indicating a role for other factors in spontaneous clearance.

The top associated SNP, rs2596542, is located 4.7 kb upstream of *MICA*, the MHC class I polypeptide-related sequence A gene, and 41.7 kb downstream of the *HLA-B* gene (**Supplementary Fig. 5**). The regional association plot at the rs2596542 locus, made using genotype data from the GWAS (**Fig. 1**) and imputation analysis (**Supplementary Fig. 6**), revealed that all of the modestly associated SNPs are tightly

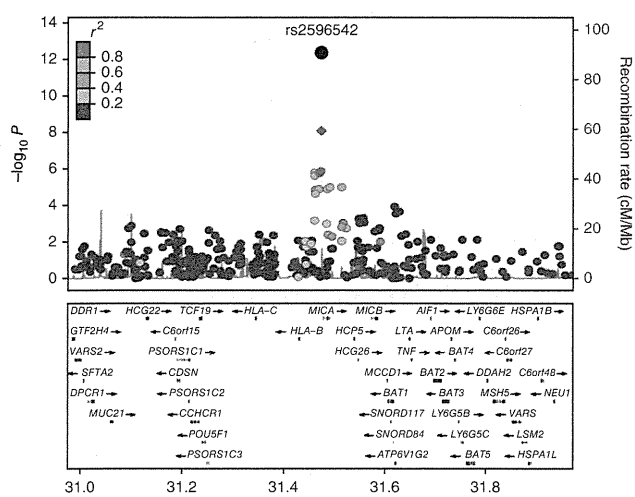

Figure 1 Regional association plot at rs2596542. Above, the *P* values of genotyped SNPs are plotted (as $-\log_{10}$ values) against their physical position on chromosome 6 (NCBI Build 36). The *P* value for rs2596542 at the GWAS stage, replication stage and combination analysis is represented by a purple diamond, circle and diamond, respectively. Estimated recombination rates from the HapMap JPT population show the local LD structure. Inset, the SNP's colors indicate LD with rs2596542 according to a scale from $r^2 = 0$ to $r^2 = 1$ based on pairwise r^2 values from HapMap JPT. Below, gene annotations from the UCSC genome browser.

Table 2 rs2596542 (A/G) is associated with progression from CHC to HCC

Subjects	RAF	(Comparison) *P*[a]	OR[a]	95% CI
Healthy	0.331			
CHC	0.333	(Healthy vs. CHC) 0.61	1.02	0.94–1.10
HCC	0.398	(CHC vs. HCC) 3.13×10^{-8}	1.36	1.22–1.51

We analyzed 5,486 controls, 1,730 CHC cases and 1,394 HCC cases. RAF, risk allele frequency (allele A); OR, odds ratio for the minor allele by considering the major allele as a reference.
[a]Calculated by logistic regression analysis, by PLINK upon age and gender adjustment under additive model.

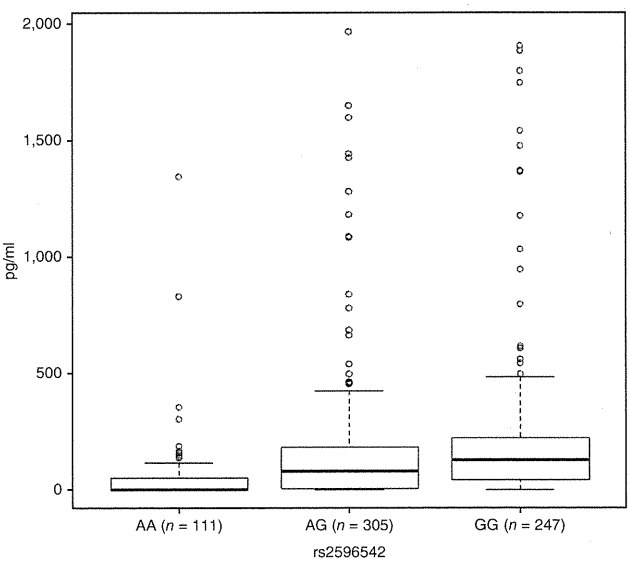

Figure 2 Correlation between soluble MICA levels and rs2596542 genotype. The *x* axis shows the genotypes at rs2596542, and the *y* axis shows the concentration of soluble MICA in pg/ml. The number of independent samples tested in each group is shown in parentheses. Each group is shown as a box plot, and the median values are shown as thick dark horizontal lines (median values of AA = 0, AG = 43.6 and GG = 77.74). The box covers the twenty-fifth to seventy-fifth percentiles, and the whiskers outside the box extend to the highest and lowest value within 1.5 times the interquartile range. Points outside the whiskers are outliers. We tested the difference in the median values among genotypes using the Kruskal-Wallis test ($P = 1.6 \times 10^{-13}$). We plotted the box plots using default settings in R (see URLs).

linked to rs2596542 ($r^2 > 0.4$) and are confined to the *MICA* gene locus. On the other hand, the imputation analysis of *HLA*-tagging SNPs did not show any evidence of linkage with rs2596542 (Online Methods and **Supplementary Table 10**), suggesting that *MICA* is a disease-associated candidate gene at this locus.

MICA is a membrane protein that acts as a ligand for NKG2D to activate anti-tumor effects through natural killer cells and $CD8^+$ T cells[19]. On the other hand, MICA is secreted into the serum by cleavage at the transmembrane domain with matrix metalloproteinases[20,21] and inhibits the anti-tumor effect of natural killer cells and $CD8^+$ T cells by blocking their action[22–24]. Elevated expression of both the membrane-bound and soluble forms of MICA (sMICA) have been reported in several cancers, including HCC[25–27]. Exon 5 of *MICA* encodes the transmembrane domain and contains a variable number of tandem repeats (VNTR) consisting of 4, 5, 6 or 9 repeats of GCT or one additional G nucleotide insertion into the 5-GCT-repeat allele (referred as A4, A5, A6, A9 and A5.1, respectively). The insertion of G (A5.1) causes a premature stop codon and subsequent loss of the transmembrane domain, leading to altered subcellular localization[28]. Therefore, we tested whether rs2596542 is in linkage with functional *MICA* VNTR alleles.

We further genotyped 673 cases with HCV-induced HCC and 890 non-HCV controls for the *MICA* VNTR locus with capillary-based electrophoresis (**Supplementary Fig. 7**). A case-control analysis revealed that the *MICA* VNTR is associated with HCV-induced HCC (global $P = 4.55 \times 10^{-7}$; **Supplementary Table 11**). Particularly, alleles A9 and A6 were associated with conferring a higher risk of HCC (OR = 1.73 and OR = 1.34, respectively), whereas the A5 and A5.1 alleles had a protective effect. Comparison of the genotypes at rs2596542 and the VNTR locus revealed that the A risk allele at rs2596542 is in LD with the A9 and A4 alleles, and the non-risk G allele is in LD with the A5 and A5.1 alleles, whereas we observed no linkage between an A6 allele and rs2596542 (**Supplementary Table 12**). We also genotyped 124 individuals with CHC; however, we observed no significant association between individuals with CHC and controls or individuals with CHC and HCC (**Supplementary Tables 13,14**).

We then tested whether the VNTR alleles, rs2596542 alleles, or VNTR-rs2596542 haplotypes had any association with MICA expression in individuals with HCV-induced HCC. We determined sMICA levels by ELISA using a total of 665 HCC serum samples (**Supplementary Table 15**). Notably, rs2596542 was significantly correlated with sMICA levels, and specifically, the risk genotype AA was associated with low levels of sMICA ($P = 1.38 \times 10^{-13}$; **Fig. 2**), whereas VNTR alleles (**Supplementary Fig. 8**) and VNTR- rs2596542 haplotypes (**Supplementary Table 16**) showed no strong association. The absence of any correlation between *MICA* VNTR alleles and sMICA suggests that sMICA levels are not regulated by post-translational processing or a premature stop codon caused by A5.1 alleles in individuals with HCC. We also examined the sMICA level in different stages of HCV-induced liver disease (in non-HCV subjects and those with CHC and HCV-induced liver cirrhosis) and found that sMICA level was elevated at the early stage of disease and was not correlated with disease progression (**Fig. 3**). Additionally, the risk allele A was also correlated with low sMICA levels in subjects with CHC (**Supplementary Fig. 9**). These findings suggest that MICA expression was induced by factors caused by chronic HCV infection,

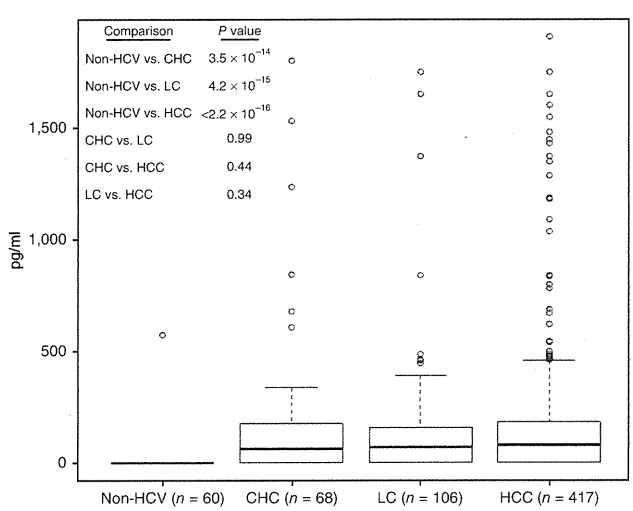

Figure 3 Correlation between soluble MICA and HCV-related diseases. The *x* axis shows the disease stages after HCV infection, and the *y* axis shows the concentration of soluble MICA in pg/ml. The number of independent samples tested in each group is shown in parentheses. Each group is shown as a box plot, and the median values are shown as thick dark horizontal lines (median values of non-HCV = 0, CHC = 64.55, LC = 72.11 and HCC = 77.98). The box covers the twenty-fifth to seventy-fifth percentiles, and the whiskers outside the box extend to the highest and lowest value within 1.5 times the interquartile range. Points outside the whiskers are outliers. We tested the difference in the median values among the disease groups using the Wilcoxon rank test. The box plots were plotted using default settings in R. Non-HCV, individuals not exposed to HCV infection; CHC, individuals with chronic hepatitis C; LC, individuals with liver cirrhosis; HCC, individuals with hepatocellular carcinoma.

similar to various types of stresses such as viral infection, inflammation and heat shock[29,30]. The levels of sMICA were shown to be directly proportional to the level of membrane-bound MICA[25], and membrane bound MICA is essential for activating natural killer cells and $CD8^+$ T cells to eliminate virus-infected cells[19]. Considering the association of the risk allele A with low levels of sMICA, our findings suggest that the individuals who carry the rs2596542 A allele would express low levels of membrane-bound MICA in response to HCV infection, which thus leads to poor or no activation of natural killer cells and $CD8^+$ T cells against virus-infected cells. Eventually, these individuals are likely to progress from CHC to HCC. Notably, several SNPs that are in absolute linkage with rs2596542 are located within the promoter or enhancer region of *MICA* and may alter the binding of stress-inducible transcriptions factors such as heat shock proteins (**Supplementary Table 17**). In this regard, it is important to analyze the factors that regulate *MICA* expression, particularly in the context of CHC. Although, the molecular mechanism whereby *MICA* polymorphisms confer the risk of disease progression should be characterized in the future, our findings reveal a crucial role of genetic variations in the host innate immune system in the development of HCV-induced HCC.

URLs. R, http://cran.r-project.org/; PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/; Primer3 v0.3.0, http://frodo.wi.mit.edu/primer3/; LocusZoom, http://csg.sph.umich.edu/locuszoom/; FastSNP, http://fastsnp.ibms.sinica.edu.tw/pages/input_CandidateGeneSearch.jsp.

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 would like to thank all the subjects and the members of the Rotary Club of Osaka-Midosuji District 2660 Rotary International in Japan who donated their DNA for this work. We also thank the technical staff of the Laboratory for Genotyping Development, Center for Genomic Medicine, RIKEN, and the Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, the University of Tokyo. This work was conducted as a part of the BioBank Japan Project that was supported by the Ministry of Education, Culture, Sports, Science and Technology of the Japanese government.

AUTHOR CONTRIBUTIONS

K.M. and Y.N. conceived of the study; Y.N., V.K., M.K. and K.M. designed the study; V.K., Y.U., R.M. and N.H. performed genotyping; V.K., Y.N. and K.M. wrote the manuscript; A.T. and N. Kamatani performed quality control at the genome-wide phase; Y.N., K.M., H.N. and M.K. managed DNA and serum samples belonging to BioBank Japan; N. Kato, R.T., M. Otsuka, M. Omata and K.K. managed replication DNA and serum samples; V.K. analyzed the data, performed VNTR genotyping, ELISA and summarized the whole results; Y.N. obtained funding for the study.

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://npg.nature.com/reprintsandpermissions/.

1. Global Burden of Hepatitis C Working Group. Global burden of disease (GBD) for hepatitis C. *J. Clin. Pharmacol.* **44**, 20–29 (2004).
2. Parkin, D.M., Bray, F., Ferlay, J. & Pisani, P. Global cancer statistics, 2002. *CA Cancer J. Clin.* **55**, 74–108 (2005).
3. Umemura, T., Ichijo, T., Yoshizawa, K., Tanaka, E. & Kiyosawa, K. Epidemiology of hepatocellular carcinoma in Japan. *J. Gastroenterol.* **44** (Suppl 19), 102–107 (2009).
4. Vong, S. & Bell, B.P. Chronic liver disease mortality in the United States, 1990–1998. *Hepatology* **39**, 476–483 (2004).
5. Ge, D. *et al.* Genetic variation in *IL28B* predicts hepatitis C treatment-induced viral clearance. *Nature* **461**, 399–401 (2009).
6. Thomas, D.L. *et al.* Genetic variation in *IL28B* and spontaneous clearance of hepatitis C virus. *Nature* **461**, 798–801 (2009).
7. Rauch, A. *et al.* Genetic variation in IL28B is associated with chronic hepatitis C and treatment failure: a genome-wide association study. *Gastroenterology* **138**, 1338–1345, 1345.e1–7 (2010).
8. Kamatani, Y. *et al.* A genome-wide association study identifies variants in the *HLA-DP* locus associated with chronic hepatitis B in Asians. *Nat. Genet.* **41**, 591–595 (2009).
9. Schütte, K., Bornschein, J. & Malfertheiner, P. Hepatocellular carcinoma–epidemiological trends and risk factors. *Dig. Dis.* **27**, 80–92 (2009).
10. Tanaka, T. *et al.* Significance of specific antibody assay for genotyping of hepatitis C virus. *Hepatology* **19**, 1347–1353 (1994).
11. Tanaka, Y. *et al.* Genome-wide association of *IL28B* with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. *Nat. Genet.* **41**, 1105–1109 (2009).
12. Anonymous. Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. *Nature* **401**, 921–923 (1999).
13. de Bakker, P.I. *et al.* A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat. Genet.* **38**, 1166–1172 (2006).
14. Kuniholm, M.H. *et al.* Specific human leukocyte antigen class I and II alleles associated with hepatitis C virus viremia. *Hepatology* **51**, 1514–1522 (2010).
15. Wang, J.H. *et al.* Ethnic and geographical differences in HLA associations with the outcome of hepatitis C virus infection. *Virol. J.* **6**, 46 (2009).
16. Singh, R., Kaul, R., Kaul, A. & Khan, K. A comparative review of HLA associations with hepatitis B and C viral infections across global populations. *World J. Gastroenterol.* **13**, 1770–1787 (2007).
17. Dixon, A.L. *et al.* A genome-wide association study of global gene expression. *Nat. Genet.* **39**, 1202–1207 (2007).
18. Ge, D. *et al.* Genetic variation in *IL28B* predicts hepatitis C treatment-induced viral clearance. *Nature* **461**, 399–401 (2009).
19. Bauer, S. *et al.* Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. *Science* **285**, 727–729 (1999).
20. Salih, H.R., Rammensee, H. & Steinle, A. Cutting edge: down-regulation of MICA on human tumors by proteolytic shedding. *J. Immunol.* **169**, 4098–4102 (2002).
21. Waldhauer, I. *et al.* Tumor-associated MICA is shed by ADAM proteases. *Cancer Res.* **68**, 6368–6376 (2008).
22. Jinushi, M. *et al.* Impairment of natural killer cell and dendritic cell functions by the soluble form of MHC class I-related chain A in advanced human hepatocellular carcinomas. *J. Hepatol.* **43**, 1013–1020 (2005).
23. Groh, V., Wu, J., Yee, C. & Spies, T. Tumour-derived soluble MIC ligands impair expression of NKG2D and T-cell activation. *Nature* **419**, 734–738 (2002).
24. Doubrovina, E.S. *et al.* Evasion from NK cell immunity by MHC class I chain-related molecules expressing colon adenocarcinoma. *J. Immunol.* **171**, 6891–6899 (2003).
25. Kohga, K. *et al.* Serum levels of soluble major histocompatibility complex (MHC) class I-related chain A in patients with chronic liver diseases and changes during transcatheter arterial embolization for hepatocellular carcinoma. *Cancer Sci.* **99**, 1643–1649 (2008).
26. Jinushi, M. *et al.* Expression and role of MICA and MICB in human hepatocellular carcinomas and their regulation by retinoic acid. *Int. J. Cancer* **104**, 354–361 (2003).
27. Groh, V. *et al.* Broad tumor-associated expression and recognition by tumor-derived gamma delta T cells of MICA and MICB. *Proc. Natl. Acad. Sci. USA* **96**, 6879–6884 (1999).
28. Ota, M. *et al.* Trinucleotide repeat polymorphism within exon 5 of the MICA gene (MHC class I chain-related gene A): allele frequency data in the nine population groups Japanese, Northern Han, Hui, Uygur, Kazakhstan, Iranian, Saudi Arabian, Greek and Italian. *Tissue Antigens* **49**, 448–454 (1997).
29. Groh, V. *et al.* Costimulation of CD8αβ T cells by NKG2D via engagement by MIC induced on virus-infected cells. *Nat. Immunol.* **2**, 255–260 (2001).
30. Groh, V. *et al.* Cell stress-regulated human major histocompatibility complex class I gene expressed in gastrointestinal epithelium. *Proc. Natl. Acad. Sci. USA* **93**, 12445–12450 (1996).

ONLINE METHODS

Sample collections. We obtained DNA from 721 HCV-related HCC cases, 1,730 CHC cases and 5,486 HCV-negative controls from the BioBank Japan project[31]. For replication analysis, DNA from 673 HCV-induced HCC cases was obtained from a prospective HCC study cohort of the University of Tokyo. A diagnosis of CHC, liver cirrhosis or HCC were based on histological, clinical and laboratory findings obtained by trained physicians. Case samples with HBV co-infection were excluded from the analysis. Interferon was administrated to 20.4% of HCC cases and 70.1% of cases were not treated. The remaining 9.5% of the cases lacked information about interferon treatment. The non-HCV controls obtained from BioBank Japan contained case-mixed individuals after excluding all individuals with cancer, chronic hepatitis B, diabetes or tuberculosis. All subjects were of Japanese origin and provided written informed consent. The clinical and demographic details of the samples are summarized in **Supplementary Table 1**. We also obtained serum samples from BioBank Japan and the University of Tokyo (**Supplementary Table 12**). This research project was approved by the ethical committees of the University of Tokyo and RIKEN.

SNP genotyping and quality control. In the GWAS, 721 individuals with HCV-related liver cancer and 2,890 controls were genotyped using Illumina HumanHap610-Quad and Illumina HumanHap550v3 Genotyping BeadChip, respectively. In the replication stage, 673 cases with HCV-related disease, 1,730 cases with CHC and 2,596 controls were genotyped by the multiplex PCR-based Invader assay (Third Wave Technologies) and the Illumina HumanHap610-Quad, respectively. The common SNPs between the Illumina HumanHap550v3 and the Illumina HumanHap610-Quad arrays from all autosomal chromosomes were included for the analysis. We applied standard SNP quality control filters to exclude SNPs with low call rate (<99%), a Hardy-Weinberg equilibrium $P < 1.0 \times 10^{-6}$ for controls and minor allele frequency of <0.01. In the end, we obtained 432,703 SNPs for the analysis. In the replication analysis, the allele discrimination plots were validated by two well-trained researchers (the plots are available on request). We excluded samples with low genotyping rate (<99%) and employed principal component analysis to avoid the population stratification issue, in which individuals belonging only to Hondo cluster were included in the analysis (**Supplementary Fig. 10**)[32].

Statistical analysis. The association of SNPs with the disease phenotype in the GWAS, replication stage and combination analyses was tested using multivariate logistic regression analysis after adjusting for age at recruitment (continuous) and gender by assuming an additive model and using PLINK[33]. In the GWAS, the genetic inflation factor (λ) was derived by applying logistic regressed P values for all the tested SNPs. The quantile-quantile plot was drawn using R. The ORs were calculated by considering the major allele as a reference, unless it was stated otherwise elsewhere. The combined analysis of the GWAS and replication stage was verified by conducting the Mantel-Haenszel method. We considered $P < 5 \times 10^{-8}$ as the genome-wide significance threshold, which is the Bonferroni-corrected threshold for the number of independent SNPs genotyped in HapMap Phase 2 (ref. 34). Heterogeneity across the two stages was examined by using the Breslow-Day test[35].

For multiple logistic regression analysis at rs2596542 using the R program, we considered age at recruitment (≤60 or >60 years)[3], gender (male or female) and alcohol consumption (non-drinkers, ≤ 50 g alcohol per day or >50 g alcohol per day) as covariates from both the GWAS and replication stage cases with HCC and non-HCV controls. Association at the *MICA* VNTR locus was analyzed by Fisher's exact test, and the global P value was calculated using a χ^2 test. Statistical comparisons between genotypes and sMICA levels were performed by Kruskal-Wallis test or Wilcoxon rank test using R. We employed the R package haplo.stats to infer haplotypes and to perform haplotype association analysis. P values for association between sMICA levels and haplotype distribution were obtained by score test under an additive model by using the haplo.score function. ORs and 95% confidence intervals were calculated from the coefficients of the GLM model by considering the major haplotype as a reference. We used the haplo.cc function to calculate these statistical values.

HCV serotype. HCV serotype data was available for 531 cases with HCC from the replication stage. HCV serotype was examined by serotyping assay (SRL Laboratory) according to previously reported methods[36]. According to the Simmonds classification[37], serotype 1 corresponded to disease types 1a and 1b, whereas serotype 2 corresponded to disease types 2a and 2b.

***MICA* VNTR locus genotyping.** We followed the method suggested by Applied Biosystems. Briefly, the 5′ end of the forward primer was labeled with 6-FAM, and the 5′ end of reverse primer was labeled with the GTGTCTT non-random sequence to promote addition of As. The primer sequences were previously reported[28]. The PCR products were mixed with Hi-Di Formamide and GeneScan-600 LIZ size standard and separated using a GeneScan system on a 3730xl DNA analyzer (Applied Biosystems). GeneMapper software (Applied Biosystems) was used to assign the repeat fragment size (**Supplementary Fig. 7**).

Quantification of soluble MICA. sMICA levels were measured by sandwich enzyme-linked immunosorbent assay, as described in the manufacturer's instructions (R&D Systems).

Imputation and association analysis at HLA allele tagging SNPs. We obtained a SNP or a combination of SNPs which can tag HLA alleles in the Japanese population from a previous study[13]. The untyped genotypes of these SNPs were imputed in the GWAS samples by using a hidden Markov model programmed in MACH[38] and haplotype information from HapMap JPT samples. We applied the same SNP quality criteria as in the GWAS for selecting SNPs for the analysis. The association was tested on all SNPs that passed the quality control criteria using logistic regression analysis conditioned on age and gender.

Initially, we obtained the pair-wise LD between HLA alleles tagging SNPs and rs2596542. We performed case-control association analysis in our GWAS dataset. As shown in **Supplementary Table 9**, none of the HLA-tagging SNPs showed evidence of linkage or association except rs2844521, and rs2844521 was in absolute linkage with rs2596542 ($r^2 = 1$, $D' = 1$) and thus showed similar association. We obtained actual genotype data at rs2596501, as this SNP is included on the 550K SNP platform, and inferred the haplotype between rs2844521 and rs2596501. However, the haplotype GT (the G allele of rs2844521 and the T allele of rs2596501), which is reported to tag the *HLA-B*3501* allele ($r^2 = 1$, $D' = 1$), was not associated with HCC in our GWAS dataset ($P = 0.39$). We also performed a conditional logistic regression analysis on rs2596501 (data not shown) and found no effect on the association between rs2596542 and HCV-induced HCC. This data suggested that rs2596542 association is independent of *HLA-B*3501*. Although we observed mild association between other *HLA-B* alleles (*HLA-B*5401*, $P = 0.004$; *HLA-B*6701*, $P = 0.012$) and HCV-induced HCC, the association at rs2596542 alone was the most significant. Taken together, we found no strong evidence for linkage of HLA alleles with rs2596542.

Software. For general statistical analysis, we used R statistical environment version 2.6.1 or plink version 1.06. The Haploview software version 4.2 (ref. 39) was used to calculate LD and to draw Manhattan plots. Primer3 v0.3.0 web tool was used to design primers. We used LocusZoom for plotting regional association plots. We used FastSNP[40] web tool for functional annotation of SNPs (see URLs for all software packages).

31. Nakamura, Y. The BioBank Japan Project. *Clin. Adv. Hematol. Oncol.* **5**, 696–697 (2007).
32. Yamaguchi-Kabata, Y. *et al.* Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. *Am. J. Hum. Genet.* **83**, 445–456 (2008).
33. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
34. Frazer, K.A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
35. Breslow, N. & Day, N. Statistical methods in cancer research. Volume II–The design and analysis of cohort studies. *IARC Sci. Publ.* 1–406 (1987).
36. Tsukiyama-Kohara, K. *et al.* A second group of hepatitis C viruses. *Virus Genes* **5**, 243–254 (1991).
37. Simmonds, P. *et al.* Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS-5 regions. *J. Gen. Virol.* **75**, 1053–1061 (1994).
38. Scott, L.J. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345 (2007).
39. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
40. Yuan, H.Y. *et al.* FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. *Nucleic Acids Res.* **34**, W635–W641 (2006).

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

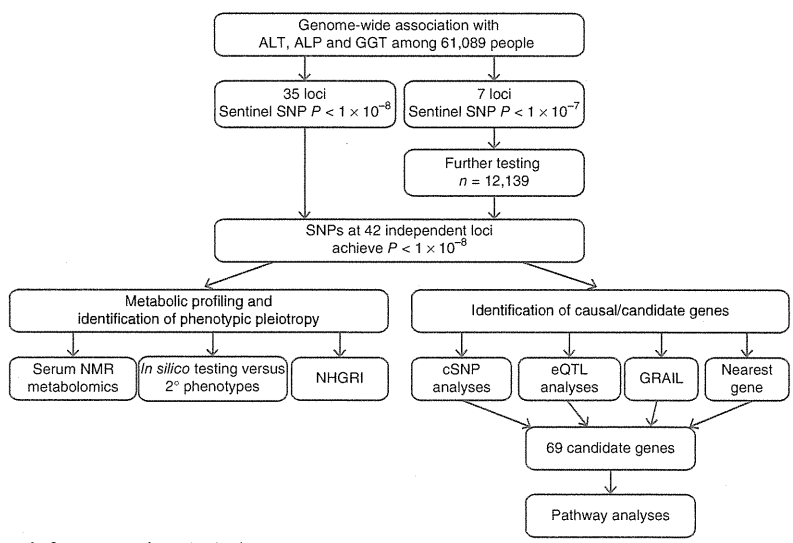

Figure 1 Summary of study design.

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

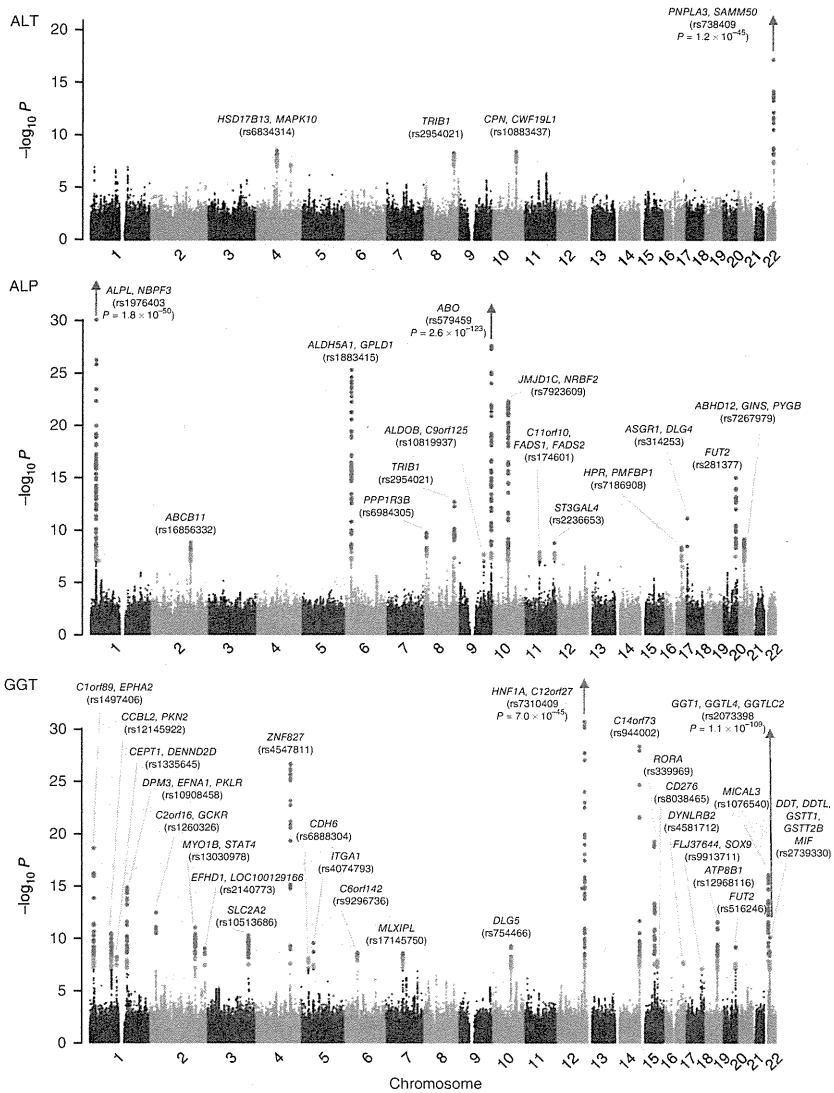

Figure 2 Manhattan plots of association of SNPs with ALT, ALP and GGT in the GWAS. SNPs reaching genome-wide significance ($P < 1 \times 10^{-8}$) are red; SNPs with $P > 1 \times 10^{-8}$ and $P < 1 \times 10^{-7}$ are green.

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

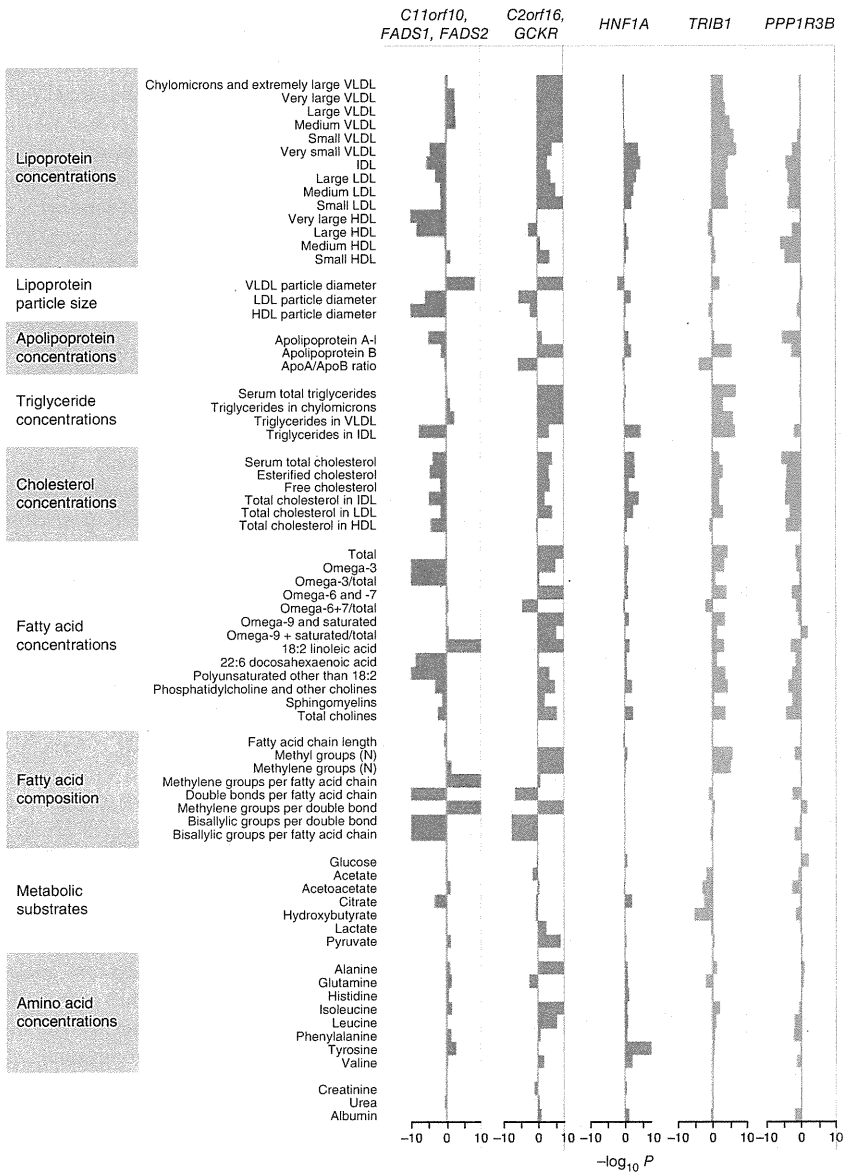

Figure 3 Association of *FADS1*, *FADS2*, *GCKR*, *HNF1A*, *TRIB1* and *PPP1R3B* loci with NMR metabonome. Bars are for $-\log_{10} P$ value, signed for direction of effect.

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], Günter 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).