Accepted Manuscript



Epigenome-wide Association Study Identifies Methylation Sites Associated With Liver Enzymes and Hepatic Steatosis

Jana Nano, Mohsen Ghanbari, Wenshi Wang, Paul S. de Vries, Klodian Dhana, Taulant Muka, André G. Uitterlinden, Joyce B.J. van Meurs, Albert Hofman, Oscar H. Franco, Qiuwei Pan, Sarwa Darwish Murad, Abbas Dehghan

PII: S0016-5085(17)35737-2
DOI: 10.1053/j.gastro.2017.06.003
Reference: YGAST 61238

To appear in: *Gastroenterology*
Accepted Date: 8 June 2017

Please cite this article as: Nano J, Ghanbari M, Wang W, de Vries PS, Dhana K, Muka T, Uitterlinden AG, van Meurs JBJ, Hofman A, BIOS consortium, Franco OH, Pan Q, Darwish Murad S, Dehghan A, Epigenome-wide Association Study Identifies Methylation Sites Associated With Liver Enzymes and Hepatic Steatosis, *Gastroenterology* (2017), doi: 10.1053/j.gastro.2017.06.003.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Epigenome-wide Association Study Identifies Methylation Sites Associated With Liver Enzymes and Hepatic Steatosis

Short Title: Epigenome wide association study on liver enzymes

Jana Nano[1], Mohsen Ghanbari[1,2], Wenshi Wang[3], Paul S. de Vries[4], Klodian Dhana[5], Taulant Muka[1], André G. Uitterlinden[1,6], Joyce B.J. van Meurs[6], Albert Hofman[1,7], BIOS consortium, Oscar H. Franco[1], Qiuwei Pan[3], Sarwa Darwish Murad[3*], Abbas Dehghan[1]*

[1] Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands

[2] Department of Genetics, School of medicine, Mashhad University of Medical Sciences, Mashhad, Iran

[3] Department of Gastroenterology and Hepatology, Erasmus University Medical Center, Rotterdam, the Netherlands

[4] Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA

[5] Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

[6] Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands

[7] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA.

* These authors contributed equally in this work.

Corresponding author: Abbas Dehghan, Department of Epidemiology, Erasmus University Medical Center, Dr. Molewaterplein 50, Office NA2906, PO Box 2040, 3000 CA Rotterdam, the Netherlands. Tel: 010-7043997. Email: a.dehghan@erasmusmc.nl

Mohsen Ghanbari, Department of Epidemiology, Erasmus University Medical Center, Dr. Molewaterplein 50, Office NA2905, PO Box 2040, 3000 CA Rotterdam, the Netherlands. Email: m.ghanbari@erasmusmc.nl

Conflict of Interest The authors declare no conflict of interest.

Authors Contributions

JN, MG, PSV gave substantial contributions to conception and design or analysis and interpretation of data. MG, WW and QP designed and performed the functional experiments. JN drafted the article and was revised critically for important intellectual content by MG, WW, PSV, KDh, TM, AGU, JBM, SDM, AH, OF, QP, and AD. Final approval of the version to be published was given from all co-authors.

Acknowledgment

ACCEPTED MANUSCRIPT

The authors are grateful to the Rotterdam Study participants, the staff involved with the Rotterdam Study and the participating general practitioners and pharmacists. The generation and management of the Illumina 450K methylation array data (EWAS data) for the Rotterdam Study was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Netherlands. The EWAS data was funded by the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, and by the Netherlands Organization for Scientific Research (NWO; project number 184021007) and made available as a Rainbow Project (RP3; BIOS) of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL). We thank Mr. Michael Verbiest, Ms. Mila Jhamai, Ms. Sarah Higgins, Mr. Marijn Verkerk, and Lisette Stolk PhD for their help in creating the methylation database. We thank Pascal Arp, Mila Jhamai, MarijnVerkerk, Lizbeth Herrera and Marjolein Peters for their help in creating the GWAS database. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Abstract

Background & aims: Epigenetic mechanisms might be involved in the regulation of liver enzyme level. We aimed to identify CpG sites at which DNA methylation levels are associated with blood levels of liver enzymes and hepatic steatosis.

Methods: We conducted an epigenome-wide association study in whole blood for liver enzymes levels including gamma-glutamyl transferase (GGT), alanine aminotransferase (ALT), and aspartate aminotransferase (AST), among a discovery set of 731 participants of the Rotterdam Study and sought replication in a non-overlapping sample of 719 individuals. Significant DNA methylation changes were further analysed to evaluate their relation with hepatic steatosis. Expression levels of the top identified gene were measured in 9 human liver cell lines and compared with expression profiles of its potential targets associated with lipid traits. The candidate gene was subsequently knocked down in human hepatoma cells using lentiviral vectors expressing small hairpin RNAs.

Results: Eight probes annotated to *SLC7A11, SLC1A5, SLC43A1, PHGDH, PSORS1C1, SREBF1, ANKS3* were associated with GGT and one probe annotated to *SLC7A11* was associated with ALT after Bonferroni correction (1.0×10^{-7}). No probe was identified for AST levels. Four probes for GGT levels including cg06690548 (*SLC7A11),* cg11376147 (*SLC43A1)*, cg22304262 (*SLC1A5)* and cg14476101 (*PHGDH)*, and one for ALT cg06690548 (*SLC7A11)* were replicated. DNA methylation at *SLC7A11* was associated with reduced risk of hepatic steatosis in participants (odds ratio, 0.69; 95% CI= (0.55 - 0.93; *P*-value: 2.7×10^{-3}). In functional experiments, *SLC7A11* was highly expressed in human liver cells; its expression is positively correlated with expression of a panel of lipid-associated genes, indicating a role of *SLC7A11* in lipid metabolism.

Conclusions: Our results provide new insights into epigenetic mechanisms associated with markers of liver function and hepatic steatosis, laying the groundwork for future diagnostic and therapeutic applications.

Key words: DNA methylation; gamma glutamyl transferase; alanine aminotransferase; aspartate aminotransferase.

ACCEPTED MANUSCRIPT

Introduction

High concentrations of liver enzymes including gamma-glutamyl transferase (GGT), alanine aminotransferase, and aspartate aminotransferase (ALT, AST) are used as markers of liver injury (1). There is strong evidence of prospective associations between liver enzymes and hepatic steatosis (2), cancer (3), type 2 diabetes (4), cardiovascular disease (5) and all-cause mortality (6). Moreover, abnormal liver function is a common reason for terminating new clinical therapeutic agents, representing a major challenge for the global pharmaceutical industry (7).

Liver enzymes can be influenced by drug therapy or lifestyle factors such as diet, physical activity, alcohol consumption and smoking (8). Several studies have also revealed a genetic impact on liver enzyme levels (9). Genome-wide association studies identified a total of 42 genetic loci associated with liver enzymes, explaining up to 2.3% of the variation (10, 11). The role of epigenetic determinants are increasingly being recognized as an important link between environmental exposures, genetic determinants and disease risk (12). DNA methylation, the major type of genetic mark, refers to the addition of a methyl group to cytosine at Cytosine-Guanine dinucleotides (CpG) that further influences the function of DNA: activating or repressing expression levels of genes (13). Unlike genetic variation, DNA methylation is dynamically remodeled over time and can be affected by environment; methylation could therefore, influence liver enzymes, and vice versa. So far, no previous study has been investigating differential methylation in relation to liver function.

Hence, we performed the first epigenome-wide association studies (EWAS) of DNA methylation in whole blood for liver enzymes GGT, ALT, and AST in 731 participants of the Rotterdam Study, a population-based cohort study. Further, we investigated whether the identified methylation probes are associated with expression levels of nearby genes and whether genetic variants are determining methylation. Subsequently, the relationship between observed DNA methylation changes and hepatic steatosis was explored. Finally, we integrated these results and performed experimental studies to provide evidence for the function of the identified gene in relation to hepatic steatosis.

Methods

ACCEPTED MANUSCRIPT

Study population

This study was performed among participants of the prospective population-based Rotterdam Study. In 1989, all 10,275 residents aged 55 years or older in Ommoord, a suburb of Rotterdam, were invited to participate in the study. In 2000, the Rotterdam Study was extended by including 3,011 participants that moved to Ommoord or people who turned 55 (RS-II). The third cohort was formed in 2006 and included 3,932 participants 45 years and older (RS-III). Participants have been re-examined every 3-4 years and have been followed up for a variety of diseases. The Rotterdam Study has been approved by the medical ethics committee according to the Population Screening Act: Rotterdam Study, executed by the Ministry of Health, Welfare and Sports of the Netherlands. All participants in the present analysis provided written informed consent to participate and to obtain information from their treating physicians. A more detailed description of the Rotterdam Study can be found elsewhere (14).

Discovery and replication panel

For the discovery panel, we used data from RSIII-1 compromising 3,932 participants examined between February 2006 and December 2008. EWAS measurements were performed on a random subset of 731 subjects. We sought replication of the identified CpG sites in a set of 719 participants from the third visit of RS-II (468 individuals) and the second visit of RS-III (251 individuals), after excluding 48 participants with missing values on any of the liver enzymes. The individuals in the replication study were not included in the discovery study.

Liver enzyme measurements and hepatic steatosis

Blood samples were collected by venipuncture, and immediately frozen (−20ºC). Serum GGT, ALT and AST levels, were determined within two weeks using a Merck Diagnostica kit on an Elan Autoanalyzer (Merck). Non-fasting samples were considered acceptable as fasting status does not greatly affect serum liver enzyme levels (15). Abdominal ultrasonography was performed by a certified and experienced technician (Pavel Taimr) on Hitachi HI VISION 900. Images were stored digitally and re-evaluated by a single hepatologist with more than ten years of experience in ultrasonography. The diagnosis of steatosis

was determined by the ultrasound technician according to the protocol by Hamaguchi et al (16). Study covariates information is found in the supplementary material.

DNA methylation data

DNA was extracted from whole peripheral blood (stored in EDTA tubes) by standardized salting-out methods. Genome-wide DNA methylation levels were measured using the Illumina Human Methylation 450K array (17). In short, samples (500ng of DNA per sample) were first bisulfite-treated using the Zymo EZ-96 DNA-methylation kit (Zymo Research, Irvine, CA, USA). Next, samples were hybridized to the arrays according to the manufacturers' protocol. The methylation proportion of a CpG site was reported as a beta-value ranging between 0 (no methylation) and 1 (full methylation). The data preprocessing was additionally performed in both datasets using an R programming pipeline based on the pipeline developed by Tost & Toulemat (18), which includes additional parameters and options to preprocess and normalize methylation data directly from idat files. 11,648 probes at X and Y chromosomes were excluded to avoid gender bias. These filtering criteria left us with 731 samples and 463,456 probes in the first dataset, and 767 samples and 419937 probes in the second dataset. The raw beta values were then background-corrected and normalized using the DASEN option of the WateRmelon R-package (19).

Liver cell lines

Human hepatoma cell lines including Hep3B, Huh6, Huh7, PLC/PRF/5, SNU182, SNU398, and SNU449 were cultured in Dulbecco's modified Eagle's medium (Invitrogen-Gibco, Breda, the Netherlands) complemented with 10% (v/v) fetal calf serum (Hyclone, Lonan, UT), 100 IU/ml penicillin, 100 µg/ml streptomycin, and 2 mM L-glutamine (Invitrogen-Gibco). The hepatoblastoma cell line HepG2 was cultured on fibronectin/collagen/albumin-coated plates (AthenaES) in Williams E medium (Invitrogen-Gibco, Breda, the Netherlands) complemented with 10% (v/v) fetal calf serum, 100 IU/ml penicillin, 100 µg/ml streptomycin, and 2 mM L-glutamine. The human liver progenitor cell line HepaRG was cultured in William's E medium supplemented with 10% (v/v) fetal calf serum, 100 IU/ml penicillin, 100 µg/ml streptomycin, 5 µg/ml insulin (Sigma-Aldrich, St. Louis, MO), and 50 µM hydrocortisone hemisuccinate (Sigma-Aldrich, St. Louis, MO). Identity of all cell lines was confirmed by STR genotyping.

Quantitative real-time polymerase chain reaction (qRT-PCR)

RNA was isolated with a Machery-NucleoSpin RNA II kit (Bioke, Leiden, The Netherlands) and quantified using a Nanodrop ND-1000 (Wilmington, DE, USA). cDNA was synthesized from total RNA using a cDNA Synthesis Kit (TAKARA BIO INC). The cDNA of all target genes was amplified for 50 cycles and quantified with a SYBRGreen-based real-time PCR (Applied Biosystems) according to the manufacturer's instructions. GAPDH was considered as a reference gene to normalize gene expression. Relative gene expression was normalized to GAPDH using the formula $2^{-\Delta\Delta CT}$ ($\Delta\Delta CT = \Delta CTsample - \Delta CTcontrol$). All the primer sequences are included in **Supplementary Table 1**.

Gene knockdown by lentiviral vectors

Lentiviral pLKO knockdown vectors (Sigma–Aldrich) targeting *SLC7A11* or control were obtained from the Erasmus Biomics Center and produced in HEK293T cells. After a pilot study, the shRNA vectors exerting optimal gene knockdown were selected (shRNA sequence: CCGGCCCTGGAGTTATGCAGCTAATCT CGAGATTAGCTGCATAACTCCAGGGTTTTTG; target sequence: CCCTGGAGTTATGCAGCTAAT). Stable gene knockdown cells were generated after lentiviral vector transduction and puromycin (2.5 µg/ml; Sigma) selection.

Statistical analyses

GGT, ALT and AST levels were log-transformed using a natural-log to obtain normal distribution. The characteristics of the discovery and replication population were compared using linear regression models adjusted for age, sex and cohort (when appropriate) for continuous variables and logistic regression models for categorical variables. In the discovery stage, we modeled associations between Dasen normalized beta-values of the CpG sites as the dependent variable and log-transformed measure of GGT, ALT, and AST as the independent variable. We used linear mixed effect models adjusting for age, sex, white blood cell proportions and technical covariates (array number and position on array). Technical covariates were modeled as random effects. Estimated leukocyte proportions (B-cells, CD4+ T-cells, CD8+ T-cells, granulocytes, monocytes and NK-cells) were calculated as described by Houseman and implemented in the minfi package in R (20, 21). We corrected for multiple testing using a robust Bonferroni corrected *P*-value of 1.0×10^{-7} as the threshold for significance (0.05 / 463,456 probes). The

probes identified in the discovery analysis were tested for replication in the independent samples from the Rotterdam study. We used identical models with the addition of cohort (RS-II or RS-III) as a variable in the model to adjust for a potential cohort effect. A Bonferroni corrected *P*-value of 0.05 divided by the number of significant findings in the discovery study was used as a threshold of significant replication. In addition, we performed a fixed effects meta-analysis using the inverse-variance weighted method implemented in METAL combining the discovery and replication samples (22). The first meta analyzed model was adjusted for age, sex, technical covariates and cell counts, whereas the second model was further adjusted for BMI, smoking history and alcohol consumption.

Integration of EWAS with existing genetic variation and expression data

Since DNA methylation may have an effect on gene expression, we tested the association between DNA methylation and mRNA expression levels of genes from the replicated CpG sites. We examined the association of significant CpG sites using data from five Dutch biobanks in 3,841 whole blood samples (http://www.genenetwork.nl/biosqtlbrowser/), integrating information on genetic variants influencing methylation levels of close or far-away genes (cis- and trans- methylation quantitative loci (meQTL)). Additionally, we explored if DNA methylation levels were affecting expression of near-by genes (cis-expression quantitative methylation (eQTM)). We subsequently tested the identified genetic variants for further associations with corresponding liver enzyme levels in the Rotterdam Study. Investigation of genetic confounding was carried out to identify whether the observed associations between liver enzymes and methylation levels were due to genetic variants being associated with both liver enzyme levels and DNA methylation. Enrichment analysis was performed to test whether the findings were over-represented for liver enzymes-related genes (**Supplementary material**). Furthermore, to investigate whether any of the identified methylation changes are a cause or a consequence of liver enzymes, we performed bi-directional Mendelian Randomization (MR) approach using the Rotterdam Study data as well as a recently published epigenome-wide association study. In Mendelian Randomization, causality is inferred from associations between genetic variants that mimic the influence of a modifiable exposure and affect the outcome of interest. Similar to a clinical trial, two randomly assigned groups with different levels of exposure (carriers of genetic variants) are compared to each other. Because gene variants do not change

over time and are assigned randomly during gamete formation, they are not prone to reverse causation and are free from confounding. More details about the MR analysis can be found in supplementary material.

Association of DNA methylation with prevalent hepatic steatosis

To assess the association of the observed liver enzyme- related CpGs with the presence of hepatic steatosis in the replication dataset, generalized linear mixed effects models were fitted using the R package lme4. Three models were analyzed. The first model was adjusted for age, sex, smoking history and whole blood cells proportions. In the second model, we additionally included body mass index, alcohol consumption, coffee intake, hypertension medication, lipid levels (HDL and triglycerides), glucose levels, and in the third model liver enzymes were further included. We used Bonferroni corrected p-value of 0.05 divided by the number of significant CpGs as a threshold for significance for the clinical outcome association. The same was done for fatty liver index in the discovery set.
All analysis was performed using the statistical package R, version 3.0.2. **Figure 1** depicts an overall overview of the study flow.

Results

Characteristics of all 1,450 participants under study are summarized in **Table 1.** Discovery and replication data sets did differ significantly with regards to mean age of participants. Although AST and ALT mean levels were different between the two samples, the cohort effect was not significant in a regression model with liver enzymes as outcome and age, sex, cohort as covariates. In the discovery sample, the mean age of participants was 59.6 years; 54% were female. No differences were observed for BMI, smoking history, alcohol consumption or other cardiometabolic risk factors. All participants were from European ancestry.

Discovery

Eight CpG sites in 7 unique loci were associated with GGT levels (*P*-values varying from 1.1×10^{-8} to 7.6×10^{-14}). **Table 2** shows the DNA methylation probes significantly associated with liver enzymes. The most significant CpG site in *SLC7A11* with GGT levels, cg06690548 (*P*-value = 1.5×10^{-14}), was also the

only epigenome-wide significant hit for ALT (*P*-value = 1.8×10^{-8}). None of the CpG sites investigated was associated with AST levels after correction for multiple testing. Manhattan plots for the association of GGT and ALT with markers of methylation are shown in **Figure 2 A** and **Figure 2 B**.

Replication

We replicated findings from our discovery set in a replication data set of 719 independent participants of the second and third cohort of the Rotterdam Study using a Bonferroni threshold of $< 6.2 \times 10^{-3}$. We significantly replicated four CpG signals (cg06690548, cg11376147, cg22304262, cg14476101) for GGT and one CpG signal (cg06690548) for ALT (**Table 2**).

A meta-analysis combining the two data sets resulted in six new CpG sites significantly associated with GGT, and one for AST (cg23734418) (**Supplementary Table 2)**. We repeated the analyses in a second model further adjusted for BMI, smoking history and alcohol consumption to test for potential confounding. The effect estimates did not change substantially between the two models.

Association with hepatic steatosis

The CpGs associated with liver enzymes were tested for an association with hepatic steatosis as diagnosed by ultrasound measurements in the replication dataset at Bonferroni corrected threshold *P*-value < 0.012. Three models were used and cg06690548, located in *SLC7A11* gene, showed an association with hepatic steatosis in all three models (model 3: OR (95% CI) = 0.69 (0.55 - 0.87); *P*- value $< 2.2 \times 10^{-3}$) (**Table 3).**

Integration of EWAS, genetic variation, and expression data

The top four CpG sites were further explored in association with expression levels (eQTM) or genetic variants (meQTL) in near-by or far-away genes (**Supplementary Table 3**). Except for *SLC7A11,* all other CpGs showed to be associated with genetic variants of near-by genes. Two SNPs located on chromosome 2 revealed genetic effects with methylation levels at *SLC1A5*. Using data from the Rotterdam study, we found a suggestive association between rs41276626 in *PHGDH* and GGT levels [rs41276626: beta (se): 0.03 (0.01); *P*-value= 0.008)] (**Supplementary Table 4).** Adjusting the model for

this SNP did not affect the association between GGT levels and DNA methylation at *PHGDH*. Cg14476101 (*PHGDH*) was the only CpG associated with expression levels of nearby genes (*P*-value= 2.05×10^{-55}) **(Supplementary Figure 1)**.

Using information provided by the Illumina array manufacturer, we functionally annotated the top hits and found that four of them (cg06690548, cg11376147, cg14476101, cg22304262) are located in the respective gene bodies. One of them (cg11376147) represents an enhancer element (bioinformatically determined) and another (cg14476101) is located within a reprogramming-specific differentially methylated region.

Liver enzymes related genes

We analyzed 781 probes representing genes identified by genome-wide association studies associated with GGT and 46 probes representing ALT-related GWAS genes. Collectively, the methylation probes were enriched for association with GGT (Fisher combined probability *P*-value = 0.0024), but not for ALT (*P*-value = 0.3).

Mendelian Randomization approach

In reference to **Supplementary Table 5**, the predicted and observed estimates were in opposing directions for *SLC43A1* and *PHGDH*. For *SLC7A11*, the estimates were consistent but not significantly different (P value =0.07). This analysis suggests supportive evidence of our hypothesis that methylation at *SLC7A11* might be a cause of GGT, but not for the other two genes.

As shown in **Supplementary Table 6**, the predicted and observed beta estimates are opposing for *SLC7A11* suggesting that it's not likely that methylation at the CpG is affected by GGT levels. Notably, for *SLC43A1* and *PHGDH* the expected and observed effect estimates are consistent and not significantly different (P = 0.55 and P = 0.11), supporting the hypothesis that methylation at these CpG sites might be a consequence of GGT levels.

The potential role of SLC7A11 in lipid metabolism

To experimentally show the importance of *SLC7A11* in liver function, we first examined the expression levels of this gene in 9 different human liver cell lines as shown in **Figure 3A**. This experiment showed that *SLC7A11* has relatively high expression levels in all these cell lines with an exceptional abundance in HepaRG cell line (**Figure 3B**). We further selected two of these cell lines, HepaRG and PLC/PRF/5, with opposite profiles of the expression levels of *SLC7A11*, specifically, high and low expression levels, and compared with the relative expression levels of a number of important lipid-associated genes reported in previous studies (23, 24). Compared with PLC/PRF/5 cells, HepaRG cells expressed much higher levels of the lipid-associated genes indicating a positive correlation between expression of *SLC7A11* and the lipid-associated genes (**Figure 3C**). In particular, the expression level of *LPL* was highly correlated with the *SLC7A11* expression, approximately 5000-fold (ΔΔCT = -12.2) higher in HepaRG than PLC/PRF/5. To have more insight into the correlation between *SLC7A11* and the lipid-associated genes, we next knocked down *SLC7A11* in HepaRG cells. After successfully silencing *SLC7A11*, we observed a significant decrease in expression levels of 7 out of 9 lipid-associated genes, specially *ABCA1* and *LPL* (**Figure 3D**). These experiments strongly suggest *SLC7A11* to be involved in lipid metabolism that may take place through inducing the expression of lipid-related genes.

Discussion

In this study, we conducted the first EWAS of DNA methylation in whole blood for liver enzymes levels (GGT, ALT, and AST) in the Rotterdam Study. We found that differential methylation of CpG sites within *SLC7A11, SLC43A1, SLC1A5, PHGDH* were associated with GGT levels. Methylation differences at *SLC7A11* were associated with ALT levels as well. In addition, we demonstrated that cg06690548 (*SLC7A11)* was directly associated with hepatic steatosis. Finally, we showed that *SLC7A11* is abundantly expressed in human liver cells and its expression is positively correlated with lipid-associated genes such as *LPL*. Collectively, our findings provide evidence in support of cross talk between genetic and epigenetic features and liver enzymes at the identified genetic loci.

The most significant replicated CpG is located at *SLC7A11* gene and was found to be associated with GGT, ALT levels and hepatic steatosis. DNA methylation is one of the epigenetic mechanisms that cells use to control gene expression, indicating that the identified CpG may alter *SLC7A11* expression. Our

knockdown experiment showed that reduced expression of *SLC7A11* in liver cells resulted in alterations in lipid metabolism as assessed by a significant decrease of expression levels of lipid-associated genes. These findings suggest the potential role of *SLC7A11* in respect to lipid homeostasis in the liver and consequently, liver related phenotypes such as hepatic steatosis. As an amino acid transporter of cysteine and glutamate, *SLC7A11* might be implicated in hepatic oxidative stress through GGT enzyme. GGT has pro-oxidative properties that may promote lipid peroxidation followed by activation of inflammatory response stellate cells that lead to fibrogenesis in hepatic steatosis (25). Our EWAS results, supported by literature and our experimental findings, suggest that DNA methylation at *SLC7A11* plays a role in lipid metabolism in the general population. We found an inverse strong association between cg06690548 and serum triglycerides levels in the Rotterdam Study (meta-analysis results in 1,485 individuals (Effect=-0.008, *P*-value= 3.9×10^{-8}) (26), and the association was further confirmed by a recent epigenome wide association study of lipid levels in 3,187 individuals (27). Additionally, cg06690548 was implicated in metabolic risk; DNA methylation changes at the site were reported to be inversely associated with diabetes (Effect= 0.16, *P*-value= 0.005) and insulin resistance as measured by HOMA-IR (Effect=-0.16, *P*-value= 8.1×10^{-7})(28). Altogether, we present both molecular and genetic association evidence to support the conclusion that the function of *SLCA11* is important for liver enzymes and hepatic steatosis as a key molecular regulator of lipid-associated genes.

Our second top hit, *PHGDH*, encodes the phosphoglycerate dehydrogenase enzyme, which catalyzes the first step in the phosphorylation pathway of serine biosynthesis. In fact, serine deficiency has been reported in a reconstructed genome-scale metabolic model of hepatocytes among patients with non-alcoholic fatty liver disease (29). Similarly, serine was demonstrated to ameliorate alcoholic fatty liver by accelerating serine–dependent homocysteine metabolism in mice and rats (30). Methylation of the same locus has been linked to body mass index and waist circumference in recent EWAS studies (31, 32). *PHGDH* represents an interesting gene for further research as methylation and cis gene expression has been reported to be strongly associated not only in blood, as we have shown with our data, but also in liver tissue (32).

Two other replicated loci for GGT, *SLC43A1* and *SLC1A5*, are coding genes for protein transporters including amino acids such as glutamine and asparagine, glucose, bile salts, organic acids and metal ions in intestine, liver and kidney (33, 34). *SLC1A5* is glutamate-transporter reported to facilitate the uptake of glutamine in tumor cells, which in turn, serves multiple metabolic functions within cells depending on the tumor oncogenotype and microenvironment (35). Drugs such as Tamoxifene and Raloxifene targeting *SLC1A5,* are used to suppress the proliferation of estrogen receptor-negative cells through inhibition of glutamine uptake (36). While *SLC1A5* role in oncogenesis is well-characterized, our findings warrant further examination of its role in cardiometabolic risk profile. Previously, *SLC7A11*, *SLC1A5, PHGDH and SLC43A1* have been reported to be associated with adiposity (32). An elegant work by Wahl et al. on epigenome wide association study of adiposity measured by body mass index, reported three of our top hits *SLC7A11, SLC43A1* and *PHGDH* to be significantly associated with BMI levels as well. Methylation at *SLC7A11* showed consistent direction of association between blood, liver and adipose tissue, whereas for *SLC43A1* and *PHGDH,* consistency was shown only between blood and liver tissue suggesting that the relationship between methylation marks are highly likely to be shared across tissues. In another study, *SLC7A11, SLC1A5, PHGDH* were reported to be linked to blood concentrations of metabolites related to steroid hormones that are upregulated in obesity (37, 38). However, the results are suspected to be driven by external common factors such as smoking and BMI seem to be an important mediator. To investigate whether our findings are confounded by such factors, we further adjusted for smoking history, body mass index and alcohol consumption in a second model. Although a decline in beta estimates indicate a potential confounding effect of these factors, the significance of the results imply that other pathways might be linked with these genes and liver enzymes.

Assessing direction of causality remains an important issue to be explored in epigenetics. In an attempt to investigate the nature of the association by employing Mendelian Randomization approach, we found suggestive evidence for methylation at *SLC7A11* to be a cause of altered GGT levels, whereas differential methylation at *SLC43A1* and *PHGDH* might be a consequence of GGT. However, these results should be interpreted with caution and further well-powered studies with adequate sample size should confirm these findings.

The combined analysis of the discovery and replication panel identified further loci including *ANKS3, ABCG1*, and *CPT1A* highlighting a continued benefit of the EWAS approach by using larger sample size studies to infer new biology. While *ANKS3* is a protein coding gene with unknown function, DNA methylation in *ABCG1* has recently been linked to glycemic traits and type 2 diabetes (39, 40). More specifically, *ABCG1* is involved in macrophage cholesterol and phospholipid transport, the dysregulation of which might be involved in the pathogenesis of cardiovascular and malignant diseases (41). *CPT1A* is a protein coding gene important in the mitochondrial transport of carnitine resulting in decreased rate of fatty acid beta-oxidation. DNA methylation in this gene has been recently reported to be associated with lipids and metabolic syndrome (42, 43). GGT is known to be involved in pathophysiology of cardio-metabolic outcomes including hepatic steatosis, type 2 diabetes and metabolic syndrome (44). The genes identified contributing mostly to the modulation of cholesterol and lipid metabolism could play a key role in the pathway from altered liver enzyme levels to the development of complex metabolic diseases.

We identified a protective association between cg06690548 (*SLC7A11*) and hepatic steatosis: DNA methylation levels of cg06690548 are inversely associated with steatosis of the liver compared to healthy individuals. Except for *SLC7A11,* the associations in the other loci did not remain independent after controlling for several cardio-metabolic risk factors. Over the recent years, it has become clear that epigenetic features might hinder a high predictive value for disease diagnosis, prognosis or risk stratification (45). Other studies are encouraged to investigate whether methylation differences could enable future large-scale screens for candidate epigenetic biomarkers important to predict future risk of hepatic steatosis, resulting in earlier detection of new cases and introducing new potential therapeutic interventions

The interpretation of our results should be made in light of strengths and limitation of the current study. The strengths of this study include the large sample size of healthy adults with available data on DNA methylation with internal replication, the use of complementary data such as expression levels and genetic variation, further integration with clinical outcome data and finally, validation with *in vitro* experiments for the top identified gene. Our study, however, has some limitations. We used whole blood samples for the quantification of DNA methylation in relation to liver enzymes and therefore, might have

missed CpGs important to other relevant tissues such as liver. While these identified CpGs represent only a subset of epigenetic markers associated with liver enzymes, that does not affect their validity. Additionally, fat accumulation for the diagnosis of hepatic steatosis was done using ultrasound, which is a qualitative technique and is less precise than quantitative imaging techniques such as MRI in diagnosis of liver fat, however, it is a widely accepted modality in large population-based studies. Methylation levels and liver enzymes were measured at the same time point, making it challenging to understand whether specific DNA methylation marks are affecting or being affected by liver enzymes or their determinants. Finally, although we adjusted for several cardio-metabolic risk factors, we cannot rule out residual confounding.

In conclusion, this is the first epigenome-wide association study of liver enzymes in whole blood. We identified four differentially methylated sites for liver enzymes including *SLC7A11, SLC43A1*, *SLC1A5* and *PHGDH* of which *SLC7A11* was independently associated with hepatic steatosis. We provide experimental evidence suggesting that *SLC7A11* play a role in lipid metabolism through regulating the expression of known lipid genes. While the current study sheds light on the epigenetic and genetic mechanisms associated with liver enzymes, future efforts in larger different population based studies could lead to identification of new differentially methylated sites.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

References

1. Pratt DS, Kaplan MM. Primary care: Evaluation of abnormal liver-enzyme results in asymptomatic patients. New England Journal of Medicine. 2000;342(17):1266-71.

2. Phillips ML, Boase S, Wahlroos S, Dugar M, Kow L, Stahl J, et al. Associates of change in liver fat content in the morbidly obese after laparoscopic gastric banding surgery. Diabetes Obes Metab. 2008;10(8):661-7.

3. Xu KC, Meng XY, Wu JW, Shen B, Shi YC, Wei Q. Diagnostic-Value of Serum Gamma-Glutamyl Transferase Isoenzyme for Hepatocellular-Carcinoma - a 10-Year Study. American Journal of Gastroenterology. 1992;87(8):991-5.

4. Kunutsor SK, Apekey TA, Walley J. Liver Aminotransferases and Risk of Incident Type 2 Diabetes: A Systematic Review and Meta-Analysis. American Journal of Epidemiology. 2013;178(2):159-71.

5. Kunutsor SK, Apekey TA, Khan H. Liver enzymes and risk of cardiovascular disease in the general population: A meta-analysis of prospective cohort studies. Atherosclerosis. 2014;236(1):7-17.

6. Kunutsor SK, Apekey TA, Seddoh D, Walley J. Liver enzymes and risk of all-cause mortality in general populations: a systematic review and meta-analysis. International Journal of Epidemiology. 2014;43(1):187-201.

7. Watkins PB. Idiosyncratic liver injury: Challenges and approaches. Toxicologic Pathology. 2005;33(1):1-5.

8. Danielsson J, Kangastupa P, Laatikainen T, Aalto M, Niemela O. Impacts of common factors of life style on serum liver enzymes. World J Gastroenterol. 2014;20(33):11743-52.

9. van Beek JHDA, Lubke GH, de Moor MHM, Willemsen G, de Geus EJC, Hottenga JJ, et al. Heritability of liver enzyme levels estimated from genome-wide SNP data. European Journal of Human Genetics. 2015;23(9):1223-8.

10. Chambers JC, Zhang WH, Sehmi J, Li XZ, Wass MN, Van der Harst P, et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. Nature Genetics. 2011;43(11):1131-U129.

11. Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, et al. Genome-wide association study of hematological and biochemical traits in a Japanese population. Nature Genetics. 2010;42(3):210-U25.

12. Slomko H, Heo HJ, Einstein FH. Minireview: Epigenetics of Obesity and Diabetes in Humans. Endocrinology. 2012;153(3):1025-30.

13. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74.

14. Hofman A, Brusselle GGO, Murad SD, van Duijn CM, Franco OH, Goedegebure A, et al. The Rotterdam Study: 2016 objectives and design update. European Journal of Epidemiology. 2015;30(8):661-708.

15. Munteanu M MD, Thabut D, et al. Intra-individual fasting versus postprandial variation of biochemical markers of liver fibrosis (FibroTest) and activity (ActiTest). Comp Hepatol. 2004;3(3).

16. Hamaguchi M, Kojima T, Itoh Y, Harano Y, Fujii K, Nakajima T, et al. The severity of ultrasonographic findings in nonalcoholic fatty liver disease reflects the metabolic syndrome and visceral fat accumulation. Am J Gastroenterol. 2007;102(12):2708-15.

17. Sandoval J, Heyn HA, Moran S, Serra-Musach J, Pujana MA, Bibikova M, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics. 2011;6(6):692-702.

18. Touleimat NaJT. Complete pipeline for Infinium((R)) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. . Epigenomics. 2012;4(3):325-41.

19. Pidsley R, Wong CCY, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. Bmc Genomics. 2013;14.

20. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. Bmc Bioinformatics. 2012;13.

21. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30(10):1363-9.

22. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190-1.

23. Mangravite LM, Thorn CF, Krauss RM. Clinical implications of pharmacogenomics of statin treatment. Pharmacogenomics J. 2006;6(6):360-74.

24. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, et al. Pharmacogenomics knowledge for personalized medicine. Clin Pharmacol Ther. 2012;92(4):414-7.

25. Dominici S, Paolicchi A, Corti A, Maellaro E, Pompella A. Prooxidant reactions promoted by soluble and cell-bound gamma-glutamyltransferase activity. Methods Enzymol. 2005;401:484-501.

26. Braun KV, Dhana K, de Vries PS, Voortman T, van Meurs JB, Uitterlinden AG, et al. Epigenome-wide association study (EWAS) on lipids: the Rotterdam Study. Clin Epigenetics. 2017;9:15.

27. Sayols-Baixeras S, Subirana I, Lluis-Ganella C, Civeira F, Roquer J, Do AN, et al. Identification and validation of seven new loci showing differential DNA methylation related to serum lipid profile: an epigenome-wide approach. The REGICOR study. Hum Mol Genet. 2016.

28. Kulkarni H, Kos MZ, Neary J, Dyer TD, Kent JW, Jr., Goring HH, et al. Novel epigenetic determinants of type 2 diabetes in Mexican-American families. Hum Mol Genet. 2015;24(18):5330-44.

29. Mardinoglu A, Agren R, Kampf C, Asplund A, Uhlen M, Nielsen J. Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. Nat Commun. 2014;5:3083.

30. Sim WC, Yin HQ, Choi HS, Choi YJ, Kwak HC, Kim SK, et al. L-serine supplementation attenuates alcoholic fatty liver by enhancing homocysteine metabolism in mice and rats. J Nutr. 2015;145(2):260-7.

31. Aslibekyan S, Demerath EW, Mendelson M, Zhi D, Guan W, Liang L, et al. Epigenome-wide study identifies novel methylation loci associated with body mass index and waist circumference. Obesity (Silver Spring). 2015;23(7):1493-501.

32. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature. 2017;541(7635):81-6.

33. Koepsell H, Endou H. The SLC22 drug transporter family. Pflugers Arch. 2004;447(5):666-76.

34. Motohashi H, Inui K. Multidrug and toxin extrusion family SLC47: physiological, pharmacokinetic and toxicokinetic importance of MATE1 and MATE2-K. Mol Aspects Med. 2013;34(2-3):661-8.

35. Altman BJ, Stine ZE, Dang CV. From Krebs to clinic: glutamine metabolism to cancer therapy. Nat Rev Cancer. 2016;16(10):619-34.

36. Todorova VK, Kaufmann Y, Luo S, Klimberg VS. Tamoxifen and raloxifene suppress the proliferation of estrogen receptor-negative cells through inhibition of glutamine uptake. Cancer Chemother Pharmacol. 2011;67(2):285-91.

37. Petersen AK, Zeilinger S, Kastenmuller G, Romisch-Margl W, Brugger M, Peters A, et al. Epigenetics meets metabolomics: an epigenome-wide association study with blood serum metabolic traits. Hum Mol Genet. 2014;23(2):534-45.

38. Azziz R, Potter HD, Bradley EL, Jr., Boots LR. delta 5-Androstene-3 beta,17 beta-diol in healthy eumenorrheic women: relationship to body mass and hormonal profile. Fertil Steril. 1994;62(2):321-6.

39. Hidalgo B, Irvin MR, Sha J, Zhi D, Aslibekyan S, Absher D, et al. Epigenome-wide association study of fasting measures of glucose, insulin, and HOMA-IR in the Genetics of Lipid Lowering Drugs and Diet Network study. Diabetes. 2014;63(2):801-7.

40. Chambers JC, Loh M, Lehne B, Drong A, Kriebel J, Motta V, et al. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. Lancet Diabetes Endocrinol. 2015;3(7):526-34.

41. Sag D, Cekic C, Wu R, Linden J, Hedrick CC. The cholesterol transporter ABCG1 links cholesterol homeostasis and tumour immunity. Nat Commun. 2015;6:6354.

42. Irvin MR, Zhi D, Joehanes R, Mendelson M, Aslibekyan S, Claas SA, et al. Epigenome-wide association study of fasting blood lipids in the Genetics of Lipid-lowering Drugs and Diet Network study. Circulation. 2014;130(7):565-72.

43. Das M SJ, Hidalgo B, Aslibekyan S, Do AN, Zhi D, Sun D, Zhang T, Li S, Chen W, Srinivasan SR, Tiwari HK, Absher D, Ordovas JM, Berenson GS, Arnett DK, Irvin MR. Association of DNA Methylation at CPT1A Locus with Metabolic Syndrome in the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) Study. Plos One. 2016;11(1).

44. Kunutsor SK, Abbasi A, Adler AI. Gamma-glutamyl transferase and risk of type II diabetes: an updated systematic review and dose-response meta-analysis. Ann Epidemiol. 2014;24(11):809-16.

45. Lee J, Kim Y, Friso S, Choi SW. Epigenetics in non-alcoholic fatty liver disease. Mol Aspects Med. 2017;54:78-88.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Figure Legend

Figure 1. Illustration of overall study flow.

Figure 2 A). Manhattan plot of epigenome-wide results testing for association between and the methylation status of GGT levels. The x-axis displays the chromosome on which the CpG is located, and the y-axis displays $-\log_{10}$ (*P*-value). The red horizontal line represents the Bonferroni-adjusted p value threshold.

Figure 2 B). Manhattan plot of epigenome-wide results testing for association between and the methylation status of ALT levels. The x-axis displays the chromosome on which the CpG is located, and the y-axis displays $-\log_{10}$ (*P*-value). The red horizontal line represents the Bonferroni-adjusted p value threshold.

Figure 3 A). Expression levels of *SLC7A11* in 9 different human liver cell lines. The expression level of SLC7A11 was quantified by qRT-PCR. The Ct values of SLC7A11 are shown in the Y axes of the figure (lower Ct value indicates higher level of gene expression). The results were presented as mean ± SEM (n=3).

Figure 3 B). The relative expression levels of *SLC7A11* against *GAPDH* (reference gene) are shown (n=3). HepaRG cells have the highest level of *SLC7A11* expression relative to GAPDH.

Figure 3 C). The expression levels of *SLC7A11* and 9 lipid-associated genes in PLC/PRF/5 and HepaRG cell lines are shown. The relative gene expression levels were quantified by qRT-PCR (n=3). This figure shows that compared with PLC/PRF/5 cells, HepaRG cells express much higher levels of the lipid-associated genes, indicating a positive correlation between *SLC7A11* and lipid-associated genes, in particular LPL. *GAPDH* serves as a reference gene, and the gene expression levels in PLC/PRF/5 set as 1.

Figure 3 D). Knockdown of SLC7A11 gene by lentiviral shRNA vectors in HepaRG cells results in significant decrease in expression of 7 out of the 9 lipid-associated genes. The experiments were repeated 5 times and the results were presented as mean ± SEM. *GAPDH* serves as a reference gene, and the gene expression levels in CTR cells set as 1. Comparisons between groups were performed with Mann-Whitney test. Differences were considered significant at *P*-value < 0.05.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 1. Characteristics of Rotterdam Study participants

Variable	Discovery (N= 731)	Replication (N= 719)
Age (years)*	59.6 (46.4 - 89.3)	67.5 (51.5 - 79.9)
Female, N (%)	395 (54)	437 (57)
AST (U/L)	24.9 (9 - 596)	25.4 (10 - 91)
ALT(U/L)	27.8 (20 - 90)	20.9 (1 - 121)
GGT(U/L)	32.2 (2 - 197)	30.2 (4 - 188)
BMI (kg/m^2)	27.4(10 - 50.2)	27.7 (17.6 - 47.6)
Smoking history (%)		
Current smoking	195 (27)	77 (11)
Former smoker	319 (44)	393 (55)
Never smoker	210 (29)	246 (34)
Alcohol consumption (%)		
Never	69 (10)	99 (14)
1 time per month	97 (13.3)	95 (13)
2-4 times per month	97 (13.3)	95 (13)
2-3 times per week	171 (23.3)	133 (18)
4 or more times per week	295 (40)	294 (41)
Hepatic Steatosis	-	280 (40)
Coffee intake (g/day)	-	406 (0 - 1276)
Fasting glucose (mmol/l)	5.4 ± 1.21	5.4 ± 1.15
HDL-cholesterol (mmol/l)	1.4 ± 0.40	1.5 ± 0.44
Triglycerides (mmol/l)	1.3 (0.9 - 1.8)	1.4 ± 0.85
Hypertensive medication	315 (50)	310 (43)

Abbreviations: ALT, Alanine aminotransferases; AST, aspartate aminotransferase; BMI, body mass index; GGT, gamma-glutamyl transferase; N, number. Data are mean ± SD or median (interquartile range) or N (%).

ACCEPTED MANUSCRIPT

Table 2. Epigenome-wide association between DNA methylation and liver enzymes (GGT, ALT and AST)

				Discovery				Replication			
	CpG	Chr	Gene	Position	Effect	SE	*P*-value	N	Effect	SE	*P*-value	N
GGT												
	cg06690548*	4	*SLC7A11*	139162808	-0.011	0.001	1.5×10^{-14}	709	-0.015	0.002	**3.9×10^{-9}**	699
	cg02711608	19	*SLC1A5*	47287964	-0.015	0.002	7.6×10^{-12}	722	-0.004	0.002	0.045	714
	cg11376147*	11	*SLC43A1*	57261198	-0.011	0.001	3.2×10^{-10}	722	-0.004	0.001	**0.004**	714
	cg22304262*	19	*SLC1A5*	47287778	-0.011	0.002	4.8×10^{-9}	722	-0.008	0.002	**0.001**	713
	cg14476101*	1	*PHGDH*	120255992	-0.017	0.003	1.1×10^{-8}	722	-0.017	0.003	**6.7×10^{-6}**	714
	cg04095776	6	*PSORS1C1*	31106941	-0.009	0.001	4.3×10^{-8}	722	0.002	0.002	0.31	713
	cg11024682	17	*SREBF1*	17730094	0.009	0.001	5.7×10^{-8}	722	0.003	0.001	0.09	713
	cg03497652	16	*ANKS3*	4751569	0.016	0.003	7.5×10^{-8}	722	0.008	0.003	0.007	714
ALT												
	cg06690548	4	*SLC7A11*	139162808	-0.013	0.0023	1.8×10^{-8}	705	-0.012	0.0035	0.0004	703

Abbreviations: **Chr, chromosome; SE, standard error; N, number of participants. *CpG with association confirmed by replication dataset; significance threshold: 1.1×10^{-7} (discovery cohort), 0.0063 (replication cohort). Effect represent the change in methylation proportion per unit change in liver enzyme level (log(U/L)).**

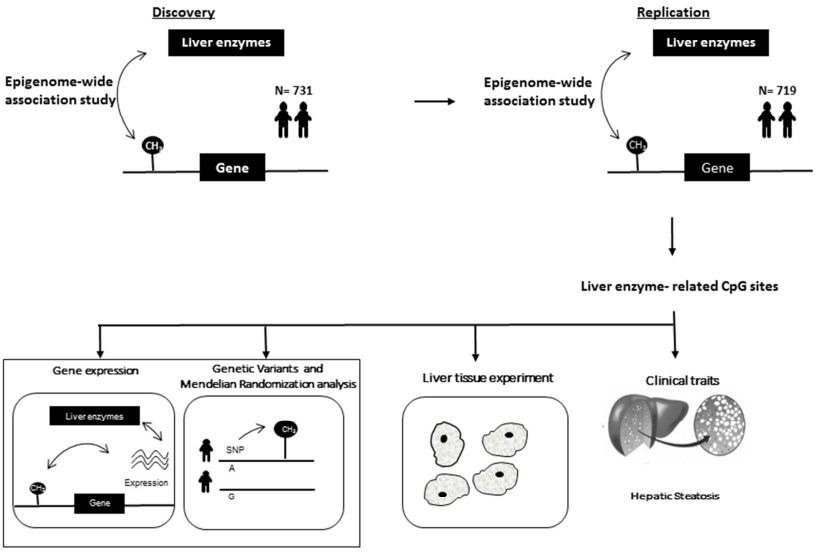
ACCEPTED MANUSCRIPT

Table 3. Association of differential DNA methylation at the top CpGs with hepatic steatosis.

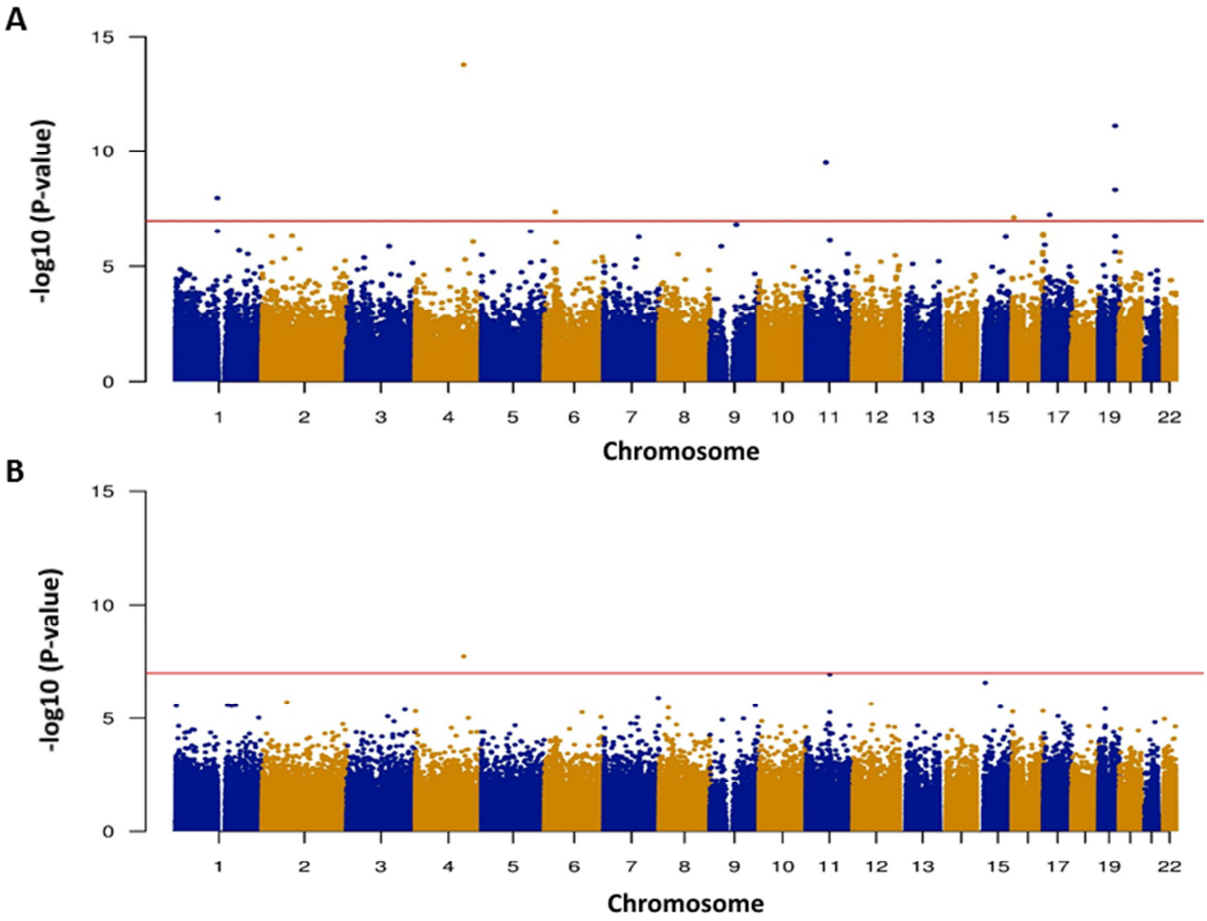
			Hepatic Steatosis (n= 280)					
			Model 1		Model 2		Model 3	
Gene	**CpG**	**Chr**	**OR (95% CI)**	***P*-value**	**OR (95% CI)**	***P*-value**	**OR (95% CI)**	***P*-value**
SLC7A11	cg06690548	4	0.66 (0.49 - 0.82)	**1.09×10^{-6}**	0.68 (0.55 - 0.85)	**7.7×10^{-4}**	0.69 (0.55 - 0.87)	**2.2×10^{-3}**
SLC43A1	cg11376147	11	0.78 (0.62 - 0.95)	**4.38×10^{-3}**	0.81 (0.66 - 1.01)	0.06	0.81 (0.65 - 1.00)	0.05
PHGDH	cg14476101	1	0.75 (0.58 - 0.91)	**6.87×10^{-4}**	0.80 (0.65 - 0.99)	0.04	0.78 (0.62 – 0.97)	0.02
SLC1A5	cg22304262	19	0.72 (0.55 - 0.88)	**9.58×10^{-5}**	0.77 (0.62 - 0.94)	0.01	0.80 (0.65 - 1.00)	0.05

Abbreviations: Chr, chromosome; OR 95% CI, odds ratio 95% confidence interval; n, number of cases; SE, standard error; *significant: level of significance *P*-value < 0.012
Model 1 adjusted for age, sex, granulocytes, lymphocytes, monocytes, cohort.
Model 2 adjusted as model 1 plus alcohol consumption, smoking history, coffee, BMI, hypertension, triglycerides, HDL, glucose levels.
Model 3 adjusted as model 2 plus liver enzyme levels

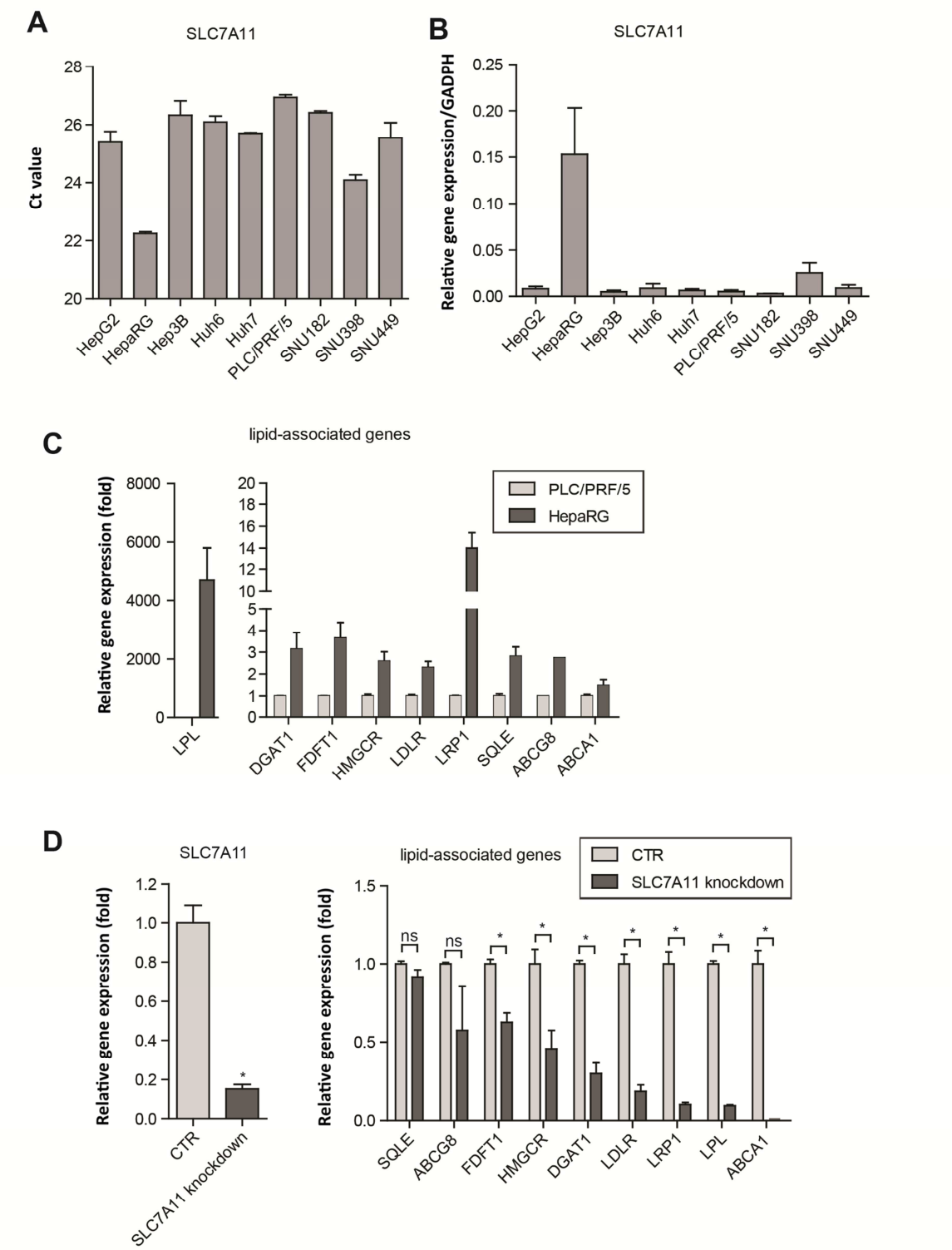
ACCEPTED MANUSCRIPT



ACCEPTED MANUSCRIPT



ACCEPTED MANUSCRIPT


A
SLC7A11
Ct value
28
26
24
22
20
HepG2
HepaRG
Hep3B
Huh6
Huh7
PLC/PRF/5
SNU182
SNU398
SNU449
B
SLC7A11
Relative gene expression/GADPH
0.25
0.20
0.15
0.10
0.05
0.00
HepG2
HepaRG
Hep3B
Huh6
Huh7
PLC/PRF/5
SNU182
SNU398
SNU449
C
lipid-associated genes
Relative gene expression (fold)
8000
6000
4000
2000
0
20
18
16
14
12
10
5
4
3
2
1
0
PLC/PRF/5
HepaRG
LPL
DGAT1
FDFT1
HMGCR
LDLR
LRP1
SQLE
ABCG8
ABCA1
D
SLC7A11
Relative gene expression (fold)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
*
CTR
SLC7A11 knockdown
lipid-associated genes
CTR
SLC7A11 knockdown
Relative gene expression (fold)
1.5
1.0
0.5
0.0
ns
ns
*
*
*
*
*
*
*
SQLE
ABCG8
FDFT1
HMGCR
DGAT1
LDLR
LRP1
LPL
ABCA1

ACCEPTED MANUSCRIPT

Supplementary material

ACCEPTED MANUSCRIPT

Methods

Covariates

Body mass index (BMI) was calculated by dividing weight in kilograms by height in meters squared. Waist circumference was measured in centimeters. Information on current and past smoking behavior and medication use was acquired from questionnaires. Alcohol intake was assessed in grams of ethanol per day from food frequency questionnaires. Participants were asked for the average daily consumption of alcohol. Blood lipids and glucose levels were measured using automatic enzymatic procedures (Roche Diagnostics GmbH, Mannheim, DE). The average of two blood pressure measurements, obtained at a single visit in upright position after a minimum of 5 minutes' rest, was used for analysis.

Mendelian Randomization approach

For each top CpG site, a *cis* SNP (1Mb distance) was identified influencing DNA methylation levels (1). We used the strength of association between the SNP and CpG together with the strength of association between CpG and liver enzyme to estimate the strength of association between SNP and liver enzyme (**Supplementary Figure 2A**). The predicted effect sizes and standard errors ($Beta_{predicted}$ and $SE_{predicted}$, respectively) were calculated as follows: $Beta_{predicted} = Beta_{CpG\sim SNP} \times Beta_{LiverEnzyme\sim CpG}$;

$$SE_{predicted} = \sqrt{SE^2_{CpG\sim SNP} \times SE^2_{LiverEnzyme\sim CpG} + SE^2_{CpG\sim SNP} \times Beta^2_{LiverEnzyme\sim CpG} + SE^2_{LiverEnzyme\sim CpG} \times Beta^2_{CpG\sim SNP}}$$

The expected strength of associations were compared to the observed associations using a test of interaction (2). Consistency between the expected and observed estimates from these analyses supports a causal relation and refutes residual confounding or reverse causation. To explore whether liver enzymes have a causal effect on CpGs, we applied the same principles to conduct a Mendelian Randomization study (**Supplementary Figure 2B**) using a genetic risk score composed of top SNPs previously reported in the largest genome wide association study on liver enzymes (3).

Results

Enrichment analysis

We also examined whether DNA methylation of genes related to liver enzymes according to the largest

available genome-wide association study (GWAS) were enriched for associations with GGT and ALT. AST levels were not included in this GWAS study. For each GWAS replicated loci, we selected genes that the top variant or one of its proxies (R^2> 0.8) were located in as exonic or intronic variants, and genes whose expression levels were associated with the top variant or one of its proxies (eQTL). Association between expression levels and genetic variants were examined using a publically available dataset based on whole blood (http:// genenetwork.nl/bloodeqtlbrowser/), and associations with a false discovery rate of less than 5% were considered significant. For GGT, a total of 23 SNPs were selected because they were in-gene and 22 were identified with an eQTL effect, leading to a set of 42 unique GGT- related genes. For ALT, we found two in-gene variants and four eQTL variants resulting in five unique genes. The methylation probes within and near these genes were included in the analysis. Further, we examined whether these genes were enriched for association with liver enzymes using Fisher's combined probability test.

References

1. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature. 2017;541(7635):81-6.
2. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. BMJ. 2003;326(7382):219.
3. Chambers JC, Zhang W, Sehmi J, Li X, Wass MN, Van der Harst P, et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. Nat Genet. 2011;43(11):1131-8.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Supplementary Table 1. Quantitative real-time polymerase chain reaction (qRT-PCR) primer sequences.

Gene name	Sequences (*5' - 3'*)
DGAT1-F	GCTTCAGCAACTACCGTGGCAT
DGAT1-R	CCTTCAGGAACAGAGAAACCACC
FDFT1-F	TGTGACCTCTGAACAGGAGTGG
FDFT1-R	GCCCATAGAGTTGGCACGTTCT
HMGCR-F	GACGTGAACCTATGCTGGTCAG
HMGCR-R	GGTATCTGTTTCAGCCACTAAGG
LDLR-F	GAATCTACTGGTCTGACCTGTCC
LDLR-R	GGTCCAGTAGATGTTGCTGTGG
LRP1-F	CAACGGCATCTCAGTGGACTAC
LRP1-R	TGTTGCTGGACAGAACCACCTC
SQLE-F	CTCCAAGTTCAGGAAAAGCCTGG
SQLE-R	GAGAACTGGACTCGGGTTAGCT
LPL-F	CTGCTGGCATTGCAGGAAGTCT
LPL-R	CATCAGGAGAAAGACGACTCGG
ABCA1-F	CAGGCTACTACCTGACCTTGGT
ABCA1-R	CTGCTCTGAGAAACACTGTCCTC
ABCG8-F	TGTCCTCGCTACAGCAATCCTG
ABCG8-R	AGAAACAGGGCTGCGAGTGACT
SLC7A11-F	TCCTGCTTTGGCTCCATGAACG
SLC7A11-R	AGAGGAGTGTGCTTGCGGACAT

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Supplementary Table 2. Top association from the meta-analysis of the discovery and replication population.

			Model 1			Model 2		
	CpG	**Gene**	**Effect**	**SE**	***P*-value**	**Effect**	**SE**	***P*-value**
GGT	cg06690548	*SLC7A11*	-0.012	0.001	1.2×10^{-22}	-0.01	0.001	5.5×10^{-16}
	cg14476101	*PHGDH*	-0.017	0.002	1.8×10^{-13}	-0.014	0.003	3.1×10^{-9}
	cg22304262	*SLC1A5*	-0.01	0.002	2.1×10^{-11}	-0.008	0.002	1.2×10^{-7}
	cg16246545	*PHGDH*	-0.014	0.002	2.2×10^{-11}	-0.012	0.002	1.2×10^{-8}
	cg11376147	*SLC43A1*	-0.008	0.001	1.7×10^{-10}	-0.006	0.001	7.7×10^{-7}
	cg02711608	*SLC1A5*	-0.009	0.006	4.1×10^{-10}	-0.007	0.002	5.1×10^{-7}
	cg01881899	*ABCG1*	0.006	0.001	6.0×10^{-9}	0.005	0.001	1.1×10^{-5}
	cg07626482	*SLC1A5*	-0.008	0.001	7.8×10^{-9}	-0.007	0.001	4.9×10^{-7}
	cg03497652	*ANKS3*	0.012	0.002	8.3×10^{-9}	0.011	0.002	1.1×10^{-6}
	cg00574958	*CPT1A*	-0.006	0.001	1.6×10^{-8}	-0.003	0.001	0.0005
ALT								
	cg06690548	*SLC7A11*	-0.012	0.002	1.9×10^{-11}	-0.01	0.002	1.6×10^{-7}
AST								
	cg23734418	NA	-0.031	0.006	6.2×10^{-8}	-0.031	0.006	7.5×10^{-8}

Abbreviations: SE, standard error; NA, not annotated

Model 1, Adjusted for age, sex, houseman estimated white blood cell proportions and batch effects.

Model 2, model 1 + BMI, smoking history, alcohol consumption.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Supplementary Table 3. Association between the DNA methylation of replicated liver enzymes with genetic variants of near-by (cis- meQTLs) and distant genes (Trans-meQTLs) and expression levels of nearby genes (cis-eQTMs)

				cis- meQTLs			Cis-eQTMs			Trans-meQTLs			
	Gene	Chr	CpG position	SNP	SNP Position	P-value	Probe	Probe position	P-value	SNP	Chr	SNP Chr. position	P-value
cg06690548	SLC7A11	4			No data			No data				No data	
cg14476101	PHGDH	1	120255968	rs11583993	120255370	3.85×10^{-228}	ENSG00092621	120202421	2.05×10^{-55}			No data	
				rs41276626	120262112	3.14×10^{-14}							
				rs34291690	120099137	9.2×10^{-8}							
cg22304262	SLC1A5	19	47287802	rs8105903	47288150	3.27×10^{-310}		No data		rs11684404	2	88924622	8.8×10^{-14}
				rs874460	47176752	3.6×10^{-13}				rs7571971	2	88895351	1.6×10^{-8}
				rs73940786	47340168	4.2×10^{-6}							
cg11376147	SLC43A1	11	57261174	rs2511984	57278733	1.7×10^{-7}		No data				No data	

Abbreviations: Chr, chromosome; SE, standard error.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Supplementary Table 4. Association of single nucleotide polymorphisms (SNP) identified by the integrative analysis (Supplementary Table 3*) with GGT levels in the Rotterdam Study.

SNP	Effect	SE	*P*-value
rs8105903	0.01	0.01	0.2
rs11583993	-0.02	0.01	0.05
rs73940786	0.02	0.01	0.1
rs41276626	0.03	0.01	0.008
rs34291690	0.01	0.01	0.2

Abbreviations: SE, standard error; SNP, single nucleotide polymorphisms. *P*-value corrected for 5 SNPs < 0.01. *The association of the remaining SNPs was not identifiable in the Rotterdam Study.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Supplementary Table 5. Hypothesis: Methylation as cause of GGT. Genetic associations to investigate the potential relationships between GGT and DNA methyaltion in blood.

				Effect of SNP on methylation (outcome variable) from Wahl et al. (n=4034)			**Relationship between methylation and GGT (outcome variable) in the Rottedam Study meta-analysis (n= 1450)**			**Predicted effect of SNP on GGT via methylation**		**Observed effect of SNP on GGT in the Rotterdam Study (n=8383)**		
CpG	SNP	Chr	Nearest gene	$Effect_{SNP\text{-}CpG}$	$SE_{SNP\text{-}CpG}$	$P_{SNP\text{-}CpG}$	$Effect_{CpG\text{-}GGT}$	$SE_{CpG\text{-}GGT}$	$P_{CpG\text{-}GGT}$	$Effect_{SNP\text{-}GGT}$ (predicted)	$SE_{SNP\text{-}GGT}$ (predicted)	$Effect_{SNP\text{-}GGT}$ (observed)	$SE_{SNP\text{-}GGT}$ (observed)	$P_{SNP\text{-}GGT}$ (observed)
cg06690548	rs6835443	4	*SLC7A11*	-0.0038	0.0009	5.0×10^{-5}	-4.25	0.46	0.0001	**0.013**	0.003	**0.007**	0.003	0.04
cg11376147	rs2729371	11	*SLC43A1*	-0.002	0.0004	2.0×10^{-6}	-3.47	0.51	0.0001	**0.007**	0.001	**-0.007**	0.004	0.12
cg14476101	rs1886736	1	*PHGDH*	-0.0479	0.0017	3.3×10^{-183}	-1.89	0.26	0.0001	**0.097**	0.014	**-0.01**	0.006	0.09

Abbreviations: Chr, chromosome; GRS, genetic risk score; SE, standard error; P, P value.

ACCEPTED MANUSCRIPT

Supplementary Table 6. Hypothesis: Methyaltion consequence of GGT. Genetic association to investigate the potential relationships between GGT and DNA methylation.

				Relationship between GRS and GGT (outcome variable) in the Rotterdam Study (n=8383)			**Relationship between GGT and methylation (outcome variable) in the Rotterdam Study (n=1450)**			**Predicted effect of GRS on methylation via GGT**		**Directly observed effect of GRS on methylation in the Rotterdam Study (n=1450)**		
CpG	Chr	Posisition	Nearest gene	Effect GRS-GGT	SE GRS-GGT	P GRS-GGT	Effect GGT-CPG	SE GGT-CPG	P GGT-CPG	Effect GRS-GGT (predicted)	SE GRS-GGT (predicted)	Effect GRS-GGT (observed)	SE GRS-GGT (observed)	P GRS-GGT (observed)
cg06690548	4	139,162,808	*SLC7A11*	0.005	0.0004	2.00×10^{-16}	-0.012	0.001	1.2×10^{-22}	$\mathbf{-6.0 \times 10^{-5}}$	0.00001	1.00×10^{-4}	0.0001	0.44
cg11376147	11	57,261,198	*SLC43A1*	0.005	0.0004	2.00×10^{-16}	-0.008	0.001	1.7×10^{-10}	$\mathbf{-4.0 \times 10^{-5}}$	0.00001	-1.00×10^{-4}	0.0001	0.26
cg14476101	1	120,255,992	*PHGDH*	0.005	0.0004	2.00×10^{-16}	-0.017	0.002	1.8×10^{-13}	$\mathbf{-8.5 \times 10^{-5}}$	0.00002	-4.00×10^{-4}	0.0002	0.06

Abbreviations: Chr, chromosome; GRS, genetic risk score; SE, standard error; P, P value. Predicted effects were calculated by multiplying the effect of GRS-GGT with the effect GGT-methylation.

ACCEPTED MANUSCRIPT

Supplementary Figure Legend

Supplementary Figure 1. Schematic overview of the interrelation of genes whose DNA methylation levels are associated with GGT levels. *PGHDH* methylation (cg14476101) was inversely associated with GGT levels (A) and positively associated with *PGHDH mRNA levels* (Probe ENSG0000009262) (B). Adjustment analysis for a genetic variant (rs41276626) located near by the gene, both related to cg14476101site (C) and GGT levels (D) in the Rotterdam study, indicated that the association between *PGHDH* methylation and GGT levels is independent of the genetic variation.

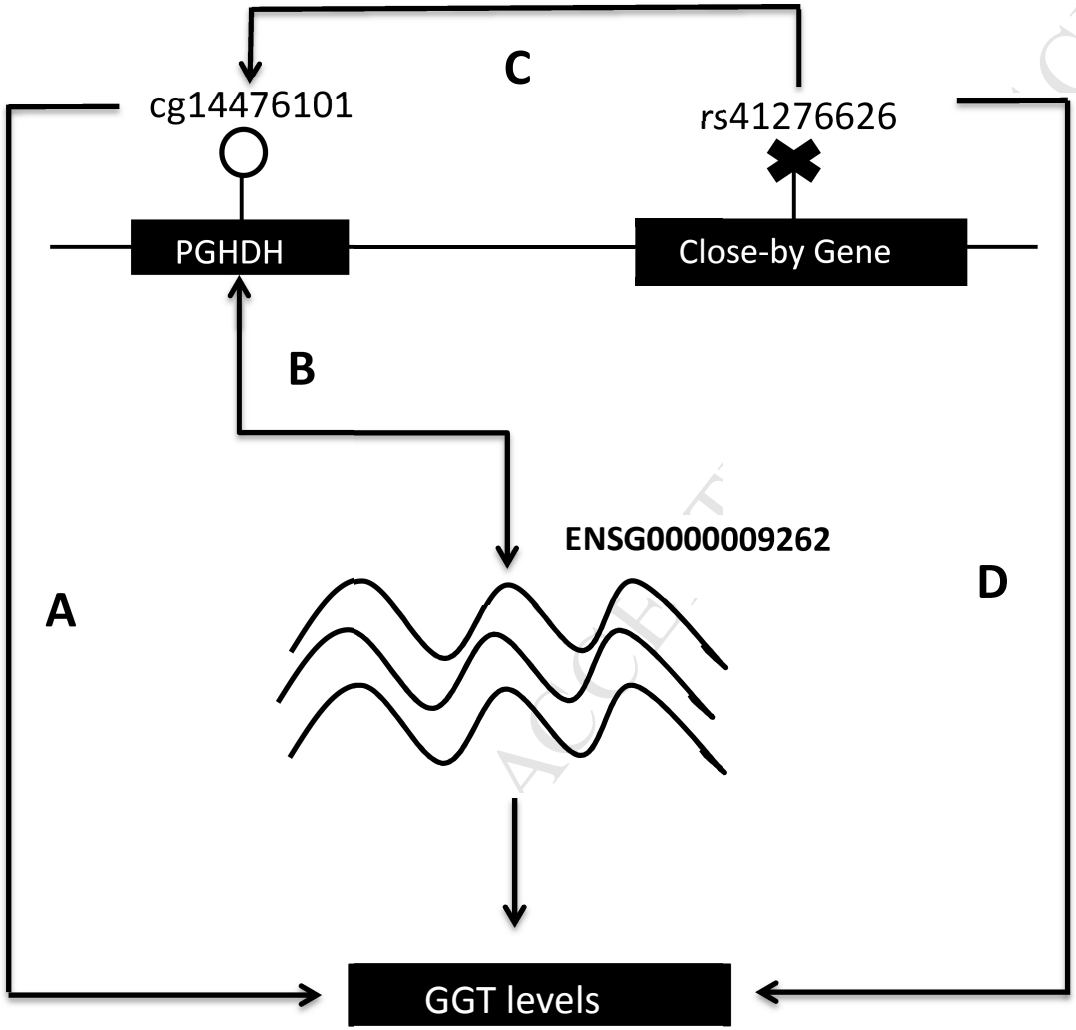
Supplementary Figure 2A. Graphical representation of the hypothesis: Methylation cause of liver enzymes. For each CpG, cis SNP (1Mb) was identified closely related with DNA methylation levels of the respective gene.

Supplementary Figure 2B. Graphical representation of the hypothesis: Methylation consequence of GGT. Genetic risk score (GRS) was calculated as a weighted average of 26 GGT-associated SNPs identified from the largest GWAS meta-analysis in liver enzymes.

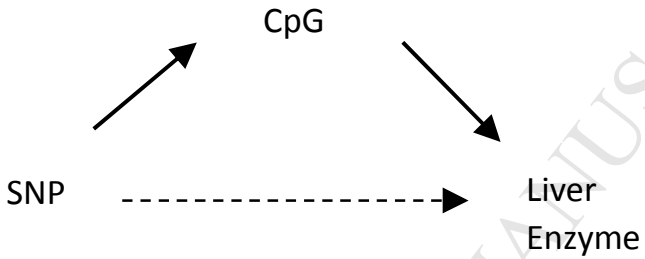
ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Supplementary Figure 1.



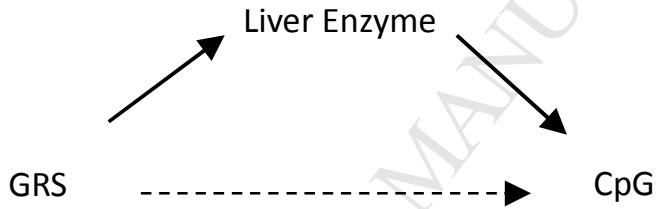
ACCEPTED MANUSCRIPT



Supplementary Figure 2A. Graphical representation of the hypothesis: Methylation cause of liver enzyme.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT



Supplementary Figure 2B. Graphical representation of the hypothesis: Methylation consequence of liver enzyme.

ACCEPTED MANUSCRIPT