



Published in final edited form as:

Cell Metab. 2013 July 2; 18(1): 130–143. doi:10.1016/j.cmet.2013.06.013.

A Genome-Wide Association Study of the Human Metabolome in a Community-Based Cohort

Eugene P. Rhee^{1,2,*}, Jennifer E. Ho^{3,4,*}, Ming-Huei Chen^{4,5,6,*}, Dongxiao Shen⁷, Susan Cheng^{4,8}, Martin G. Larson^{4,9}, Anahita Ghorbani¹⁰, Xu Shi⁷, Iiro T. Helenius⁷, Christopher J. O'Donnell^{4,10,11}, Amanda L. Souza², Amy Deik², Kerry A. Pierce², Kevin Bullock², Geoffrey A. Walford^{12,13}, Ramachandran S. Vasam^{4,14}, Jose C. Florez^{12,13,15}, Clary Clish², J.-R. Joanna Yeh⁷, Thomas J. Wang^{16,17,†,‡}, and Robert E. Gerszten^{2,7,10,†,‡}

¹Nephrology Division, Massachusetts General Hospital (MGH), Boston, MA 02114, USA

²Metabolite Profiling Platform, Broad Institute, Cambridge, MA 02142, USA

³Cardiovascular Medicine Section, Department of Medicine, Boston University (BU) School of Medicine, Boston, MA 02118, USA

⁴Framingham Heart Study of the NHLBI and BU School of Medicine, Framingham, MA 01702, USA

⁵Department of Neurology, BU School of Medicine, Boston, MA 02118, USA

⁶Department of Biostatistics, BU School of Public Health, Boston, MA 02118, USA

⁷Cardiovascular Research Center, MGH, Boston, MA 20114, USA

⁸Division of Cardiovascular Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA

⁹Department of Mathematics and Statistics, BU, Boston, MA 02215, USA

¹⁰Cardiology Division, MGH, Boston, MA 02114, USA

¹¹Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA

¹²Diabetes Research Center (Diabetes Unit), MGH, Boston, MA 02114, USA

¹³Center for Human Genetic Research, MGH, Boston, MA 02114, USA

¹⁴Preventive Medicine and Epidemiology, BU School of Medicine, Boston, MA 02118, USA

¹⁵Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, USA

¹⁶Division of Cardiovascular Medicine, Vanderbilt University Medical Center, Nashville, TN 37232, USA

¹⁷Vanderbilt Heart & Vascular Institute, Nashville, TN 37232, USA

SUMMARY

†correspondence: thomas.j.wang@vanderbilt.edu (T.J.W.), (P) 615-936-1717, (F) 615-936-2029; rgerszten@partners.org (R.E.G.), (P) 617-724-8322, (F) 617-726-5806.

‡equal contribution

†co-senior author

Metabolite data have been uploaded on the database of Genotypes and Phenotypes (<http://www.ncbi.nlm.nih.gov/gap>), accession numbers: pht002234.v3.pt, pht002566.v2.p7, pht002343.v2.p7, pht002567.v2.p7, pht002565.v2.p7, and pht002894.v1.p7.

Because metabolites are hypothesized to play key roles as markers and effectors of cardio-metabolic diseases, recent studies have sought to annotate the genetic determinants of circulating metabolite levels. We report a genome-wide association study (GWAS) of 217 plasma metabolites, including >100 not measured in prior GWAS, in 2,076 participants of the Framingham Heart Study. For the majority of analytes, we find that estimated heritability explains >20% of inter-individual variation, and that variation attributable to heritable factors is greater than that attributable to clinical factors. Further, we identify 31 genetic loci associated with plasma metabolites, including 23 that have not previously been reported. Importantly, we include GWAS results for all surveyed metabolites, and demonstrate how this information highlights a role for *AGXT2* in cholesterol ester and triacylglycerol metabolism. Thus, our study outlines the relative contributions of inherited and clinical factors on the plasma metabolome and provides a resource for metabolism research.

INTRODUCTION

Recent studies have begun to integrate genomic and metabolomic data in human cohorts (Demirkan et al., 2012; Gieger et al., 2008; Hicks et al., 2009; Illig et al., 2010; Kettunen et al., 2012; Suhre et al., 2011a; Suhre et al., 2011b; Tukiainen et al., 2012). Because metabolites are hypothesized to play key roles as markers and effectors of cardiometabolic diseases, these efforts seek to both refine and expand our understanding of the causal determinants of circulating metabolite levels. Studies to date have been notable for the identification of loci at enzymes or transport proteins directly involved with a given metabolite's disposition (Suhre and Gieger, 2012). In turn, many of these loci have shown relatively large effect sizes on metabolite levels, as compared to findings in genome wide association studies (GWAS) for common diseases.

Whereas genetically informative deoxyribonucleic acid sequence is limited to four distinct chemical motifs, endogenous metabolites span a variety of compound classes with significant differences in size and polarity, across a wide range of concentrations. As a consequence, no single analytical method is able to accommodate the chemical diversity of the entire metabolome. Thus, GWAS of metabolite traits to date have employed various methodologies including nuclear magnetic resonance spectroscopy and mass spectrometry (MS), with the latter coupled to both gas and liquid chromatography (LC) (Rhee and Gerszten, 2012). Even with a given analytical tool, distinct methods have been required to survey polar versus lipid analytes. We have developed a LC-MS based metabolomics platform that measures a total of 217 analytes (113 polar analytes and 104 lipid analytes), including >100 not measured in prior GWAS.

In the current study, we performed metabolite profiling on plasma obtained from 2,076 individuals in the Framingham Heart Study (FHS). The family-based structure of this cohort, as well as its rich cardiometabolic phenotyping, presents a unique opportunity to study the relative contributions of heritable, environmental, and clinical factors influencing the plasma metabolome. For many metabolites, we confirm that a substantial fraction of metabolite variability is heritable (Shah et al., 2009; Kettunen et al., 2012), often exceeding the influence of measured clinical factors. Using GWAS, we also identify numerous locus-metabolite associations and demonstrate how these findings complement and extend prior association studies of complex human traits. Finally, we include a proof-of-principle demonstration of how the breadth of metabolite, genotype, and phenotype data we present in FHS can motivate functional studies to provide biological insight.

RESULTS

A total of 2,076 participants of the FHS Offspring Cohort, including 873 sibships, underwent metabolic profiling and genome-wide genotyping. The mean age was 55 years and 51% of participants were women (Table 1).

Relative contributions of heritable and clinical factors to metabolite levels

The estimated proportion of inter-individual metabolite variation attributable to heritable factors (including genome-wide significant loci) compared with clinical factors (age, sex, systolic blood pressure, anti-hypertensive medication use, body-mass index, diabetes, smoking status, and prevalent cardiovascular disease) is displayed in Figure 1A and Figure S1. Metabolites most influenced by clinical variables include the nicotine metabolite cotinine (70% of variation explained by clinical factors, 66% by smoking alone), and fructose/glucose/galactose (45%). Adjustment for renal function in secondary analyses did not appreciably change results (Table S1).

For the majority of metabolites, the proportion of variation attributable to heritable factors was greater than that of clinical factors: for 93% of metabolites assayed, measured clinical factors accounted for 20% or less of inter-individual variation. By contrast, estimated heritability explained greater than 20% of the inter-individual variation for 66% of metabolites. Amino acids and other polar analytes had the highest heritability estimates, including carnosine ($h^2=0.86$, $P=6.8\times 10^{-4}$), anthranilic acid ($h^2=0.84$, $P=3.2\times 10^{-14}$), and glutamate ($h^2=0.82$, $P=9.1\times 10^{-13}$), whereas heritability estimates for lipid analytes were lower, with the highest estimate for lysophosphatidylcholine (LPC) 22:6 ($h^2=0.46$, $P=2.0\times 10^{-7}$).

Heritability estimates for essential amino acids were lower than for non-essential amino acids: mean $h^2=0.29$, range 0.14–0.43, versus mean $h^2=0.53$, range 0.15–0.82, respectively; $P=0.01$ (Figure 1B). Similarly, none of the essential amino acids were associated with genetic loci at a genome-wide significant level, whereas 5 of the 10 non-essential amino acids monitored by our platform had genome-wide significant findings in our study. Conversely, clinical factors explained a greater proportion of variation for essential versus non-essential amino acids (mean R^2 for clinical model=0.17, range 0.04–0.34, versus mean $R^2=0.09$, range 0.03–0.16, respectively). These findings align with the relative contributions of endogenous (inherited) versus dietary (environmental) factors for these small molecules, and provide internal validation for the observed distribution of metabolite heritability.

GWAS identifies 31 genetic loci associated with plasma metabolite levels

Genome-wide associations are displayed in Table 2, and quantile-quantile and linkage disequilibrium-plots for these associations are displayed in Figures S2 and S3. Of 217 metabolites analyzed, 64 had at least one genome-wide significant locus. Conversely, 31 discrete loci were associated with at least one metabolite trait, and a number of loci were associated with multiple metabolites (Table 2). Our data both replicate previously identified associations, as well as identify numerous novel locus-metabolite associations (Figure 2). These novel findings include loci that span genes encoding proteins with a direct biochemical relationship with a given metabolite, loci previously associated with complex human disease traits, and loci with no prior significant associations in GWAS.

Confirmation of previously established locus-metabolite associations

Eight of the locus-metabolite associations identified in our study have been previously reported, and 7 of these 8 associations involve genes directly related to the transport or synthesis of a given metabolite (Figure 2). For example, we replicate prior associations

between loci at *SLC16A9*, which encodes a carnitine efflux transporter, and carnitine; *PRODH* (proline dehydrogenase), which encodes the enzyme that catalyzes the first step of proline catabolism, and proline; and *PHGDH* (phosphoglycerate dehydrogenase), which encodes the enzyme that catalyzes the first and rate-limiting step of serine biosynthesis, and serine (Suhre et al., 2011a). A locus at *SLC16A10*, which encodes a tyrosine and phenylalanine transporter, has previously been associated with the ratio of isoleucine/tyrosine (Suhre et al., 2011a) and the ratio of alanine/tyrosine (Kettunen et al., 2012); we show that this association holds true for tyrosine alone. We also identify an association between a locus at *AGXT2* (alanine-glyoxylate aminotransferase-2) and its enzyme substrate β -aminoisobutyric acid. Prior work has shown an association between this locus and urinary levels of β -aminoisobutyric acid (Suhre et al., 2011b).

Further, we confirm the association between glycine and a variant at *CPS1* (carbamoyl phosphate synthase 1), which encodes the enzyme that catalyzes the first committed step of the urea cycle. Although not a urea cycle intermediate, glycine can react with arginine (a urea cycle intermediate) to yield ornithine (a urea cycle intermediate) and guanidinoacetic acid. Methylation of guanidinoacetic acid yields creatine, which is ultimately metabolized to creatinine. Notably, we identify a novel association between the *CPS1* locus and creatine, whereas others have identified an association between *CPS1* common variants and creatinine (Kottgen et al., 2010). Thus, complementary metabolomic data sets are able to extend the network of locus-metabolite associations along defined biochemical pathways.

Finally, we replicate the association between the *FADS1-3* (fatty acid desaturase 1–3) locus and phosphatidylcholines (PCs) 36:4 and 38:4 (Gieger et al., 2008), as well as between a variant within *GCKR* (glucokinase regulator) and alanine (the *GCKR* variant was previously associated with alanine/glutamine) (Kettunen et al., 2012). Novel associations at these loci with triacylglycerol (TAG) traits are discussed further below.

Novel associations in directly related biological pathways

Among the numerous novel findings in our study, we first describe eight locus-metabolite associations with strong biological plausibility. In each case, the locus of interest includes a gene encoding a protein directly responsible for the metabolism or transport of the given metabolite. For three of the loci, mutations have been identified as the cause of human disease. For example, we identify an association between a variant at *UMPS* (uridine monophosphate synthase) and orotic acid. *UMPS* encodes an enzyme that combines orotic acid and ribose-5-phosphate to form uridine monophosphate, and mutations in this gene have been identified as the cause of hereditary orotic aciduria (OMIM: 258900). Similarly, we identify an association between a common variant at *AGA* (aspartylglucosaminidase) and asparagine. *AGA* encodes an enzyme that cleaves asparagine from N-acetylglucosamines as a final step in the lysosomal breakdown of glycoproteins, and mutations in this gene result in the lysosomal storage disease aspartylglucosaminuria (OMIM: 613228). Finally, we find an association between the *SERPINA7* locus and thyroxine. *SERPINA7* encodes thyroxine-binding globulin, and mutations in this gene result in various degrees of thyroxine-binding globulin deficiency (OMIM: 314200).

In addition to these findings with established human disease correlates, we identify five other locus-metabolite associations with strong biochemical underpinnings. We find an association between a variant at *DMDGH* (dimethylglycine dehydrogenase) and its enzyme substrate dimethylglycine; at *GMPR* (guanosine monophosphate reductase) and the purine nucleoside xanthosine; at *SLC6A13*, which encodes a transporter with known specificity for γ -aminoisobutyrate (GABA), and the GABA-isomer β -aminoisobutyric acid; at *APOA1/C3/A4/A5* (apolipoprotein A1/C3/A4/A5) and various TAGs and diacylglycerols (DAGs); and at *DDAH1* (dimethylarginine dimethylaminohydrolase 1) and NG-monomethylarginine

(NMMA). Prior studies have shown that DDAH1 is responsible for the degradation of the dimethylarginines NMMA and asymmetric dimethylarginine (ADMA), but not symmetric dimethylarginine (SDMA) (Hu et al., 2011), and *DDAH1* polymorphisms have previously been associated with ADMA levels (Abhary et al., 2010). In our data, the top SNP (rs18582) at *DDAH1* had a modest association with ADMA ($P=5.6\times 10^{-6}$), but no association with SDMA ($P=0.15$).

Other novel locus-metabolite associations

Several of the novel locus-metabolite associations identified in our study include loci without a clear biochemical relationship with the given metabolite. In several cases, however, these loci have been associated with human disease or complex disease phenotypes. For example, mutations in *SLC7A9* cause cystinuria type B (OMIM: 220100) (Feliubadalo et al., 1999) and common variants in *SLC7A9* have been associated with chronic kidney disease (CKD) (Kottgen et al., 2010). We report an association between the *SLC7A9* locus and NMMA, with the minor allele at this locus associated with a lower risk of CKD and lower plasma levels of NMMA, raising the question of whether NMMA is a potential biomarker or effector of CKD progression. Among the 2,076 individuals in the current study, 123 with normal kidney function at the time of metabolite profiling developed new-onset CKD in the subsequent 8 years – interestingly, higher plasma levels of NMMA were significantly associated with the risk of developing future CKD (OR per SD 1.32, $P=0.003$) (Rhee et al., 2013).

Further examples of locus-metabolite associations identified in our study with potential links to human disease include an association between the *HPS1* locus (Hermansky-Pudlak syndrome 1, OMIM: 203300) and ADMA. Similarly, loci at *SYNE2* (spectrin repeat containing, nuclear envelope 2), associated with sphingomyelin (SM) 14:0 in our data, has been associated with atrial fibrillation (Ellinor et al., 2012); at *DGKB* (diacylglycerol kinase), associated with indole propionate, has been linked to fasting glucose (Dupuis et al., 2010); at *NTAN1* (N-terminal asparagine amidase), associated with cholesterol ester (CE) 20:3, has been associated with bone mineral density (Estrada et al., 2012); at *LIPC*, associated with lysophosphatidylethanolamine (LPE) 16:0, has been associated with macular degeneration and the metabolic syndrome (Neale et al., 2010; Kristiansson et al., 2012); at *SLCO1B1* (solute carrier organic anion transporter family member 1B1), associated with LPE 20:4, has been associated with statin-induced myopathy (Link et al., 2008); and at *PDE4D* (phosphodiesterase 4D), associated with SM 24:1, has previously been linked to stroke (Song et al., 2006), although this association was not validated in a larger study (Bellenguez et al., 2012).

Although we catalog each of the loci that have been associated with human disease and have at least one genome-wide significant metabolite association in our study, metabolite associations that do not reach this threshold may also provide pathophysiologic insights. As an example, we examined metabolite associations with variants spanning *KCNQ1* (potassium voltage-gated channel, KQT-like subfamily, member 1); common variants in *KCNQ1* have previously been associated with type 2 diabetes, with the hypothesis that the encoded channel may modulate pancreatic insulin secretion (Unoki et al., 2008; Yasuda et al., 2008). In our study, we note an association between rs384037 in *KCNQ1* and triiodothyronine levels ($P=5.3\times 10^{-5}$). Furthermore, across the 2,076 individuals in the current study, triiodothyronine levels are strongly associated with metabolic traits, including plasma insulin levels ($P=1.5\times 10^{-20}$), plasma triglyceride levels ($P=5.3\times 10^{-12}$), and body-mass index ($P=4.9\times 10^{-8}$). Thus, our data in FHS raise the possibility that the association between common variants in *KCNQ1* and type 2 diabetes may also be mediated by the gene's role in modulating plasma triiodothyronine levels.

Lipid profiling demonstrates heterogeneous effects of loci associated with total triglycerides

Although prior GWAS have identified numerous loci associated with total triglyceride levels, our study is the first to incorporate comprehensive TAG profiling. As various combinations of acyl chains may be esterified to a glycerol backbone, bulk triglycerides are actually composed of dozens of distinct TAG molecules. In the current study, we identify significant TAG associations for three loci previously associated with total triglycerides – *GCKR*, and the *FADS1-3* and *APOA1/C3/A4/A5* gene clusters. A fourth region at chromosome 7p12.1 (rs6593086) found to be associated with TAG traits has not been previously associated with total triglyceride levels. To test the hypothesis that the higher resolution phenotyping enabled by our platform sheds insight on these associations, we examined each locus's association with all of the TAGs monitored by our platform.

For leading SNPs at *GCKR*, *FADS1-3*, *APOA1/C3/A4/A5*, and rs6593086, Figure 3 depicts the beta coefficient and P-value for association across the 46 TAGs measured in FHS. As suggested by the four significant findings (TAGs 48:2, 48:3, 50:3, 50:4), the *GCKR* locus demonstrated a stronger association with TAGs of relatively lower carbon content (Figure 3A). A comprehensive view of PCs shows a similar pattern of association (including significant associations with PC 32:2 and PC34:3) (Figure 4A). Notably, the top SNP associated with these metabolic traits was rs1260326, a missense variant (L446P) that in functional studies has been established as the likely causal variant explaining the association with fasting bulk triglyceride and glucose levels (Beer et al., 2009; Orho-Melander et al., 2008).

In contrast to the *GCKR* locus, the *FADS1-3* locus had stronger associations with TAGs of relatively higher carbon and double bond content (Figure 3B), including significant associations with TAGs 54:4, 58:10, and 58:11. These data extend prior work that has demonstrated a similar pattern of association between this locus and plasma phospholipid carbon and double bond content (Gieger et al., 2008), and is corroborated by our own PC data (Figure 4B). The leading SNP in the *APOA1/C3/A4/A5* gene cluster was associated with an intermediate TAG phenotype relative to *GCKR* and *FADS1-3*, demonstrating stronger associations for TAGs with intermediate carbon content (Figure 3C, i.e., TAGs with 50 to 56 carbons). Figure 3D demonstrates a striking pattern of TAG associations for rs6593086, a SNP that is located > 50 kb from the closest coding gene (*POM121L12*). As with the *FADS1-3* locus, this SNP had stronger associations with TAGs of relatively greater carbon and double bond content, although the significant associations were non-overlapping. Further, rs6593086 had a consistent direction of association across the majority of TAGs, whereas the direction of effect for the *FADS1-3* locus differed at the extremes of TAG carbon content.

Genome wide association data across all surveyed metabolites as a resource for metabolism research

Although the novelty of our TAG data set motivates interest in select TAG association patterns, further interrogation of the breadth of our data will provide other insights as well. To that end, we include GWAS data for each of the 217 metabolites surveyed by our platform, including all loci with $P < 1.0 \times 10^{-3}$ (Table S2), as well as comprehensive metabolite data for each locus with at least one genome-wide significant association (Table S3). With these resources, we believe that independent investigators will be able to rapidly interrogate the genetic underpinnings of metabolites of interest, including biologically meaningful associations that do not meet the genome-wide significance threshold (but potentially of high statistical significance in a focused interrogation). Conversely, investigators focused on a particular gene highlighted by our study will be able to query

which metabolites are downstream of common genetic variation in that gene. Finally, access to the breadth of our metabolite GWAS data will complement the publicly available metabolite data that we have already uploaded on the database of Genotypes and Phenotypes (<http://www.ncbi.nlm.nih.gov/gap>) for all 2,076 individuals in FHS profiled in the current study.

To provide proof of principle of this approach, we examined our data in the FHS on β -aminoisobutyric acid. In cross-sectional analyses, plasma β -aminoisobutyric acid levels have a negative correlation with serum triglyceride levels in the FHS (a 1-SD decrement in log β -aminoisobutyric acid is associated with a 1.07 mg/dL increase in serum triglycerides, $P=2.3\times 10^{-21}$). In the current study, we identify a striking association between the *AGXT2* locus and plasma β -aminoisobutyric acid levels ($P=5.8\times 10^{-83}$), with the top SNP (rs37370) in *AGXT2* accounting for 36% of its estimated heritability. In light of the cross-sectional association between β -aminoisobutyric acid levels and total triglycerides, we were interested to note that rs37370 also had many nominal associations with plasma TAGs and CEs (Table S4), with the direction of association opposite for TAGs versus CEs, suggesting that genes responsible for β -aminoisobutyric acid metabolism may have a causal and opposing impact on plasma TAG and CE levels.

In order to test this potential link between β -aminoisobutyric acid and lipid homeostasis, we knocked down *agxt2* in zebrafish using morpholino antisense oligonucleotides (Figures S4A and S4B). In some fish, this knockdown resulted in no overt phenotype, whereas in others it resulted in a prominent defect in yolk sack extension and mild pericardial edema (Figure 5A). When compared to control fish, *agxt2* knockdown fish with normal phenotype (AGXT2 I) and abnormal phenotype (AGXT2 II) had lower *agxt2* mRNA (Figure 5B) and β -aminoisobutyric acid levels (Figure 5C); further, AGXT2 II fish had a trend for lower *agxt2* mRNA and β -aminoisobutyric acid levels compared to AGXT2 I fish. Lipid profiling demonstrated a broad, opposing, and dose-dependent impact of *agxt2* knockdown on fish CE and TAG levels (Figures 5D and 5E) that aligns with the opposing directionality of association between the *AGXT2* locus and TAG and CE levels in humans (Figure 5F).

To test whether the association between β -aminoisobutyric acid and lipid metabolism is specific to *agxt2*, we used morpholino antisense oligonucleotides to knock down *abat* (Figure S4C), which encodes an enzyme that catalyzes an alternative pathway for β -aminoisobutyric acid metabolism (Figure S4A). As with *agxt2* knockdown, *abat* knockdown resulted in an abnormal phenotype notable for a defect in yolk sack extension and pericardial edema (Figure S4D). Furthermore, *abat* knockdown in fish resulted in decreased β -aminoisobutyric acid levels compared to control fish (Figure S4E), and recapitulated the decrease in CE levels and the increase in TAG levels seen in *agxt2* knockdown fish (Figures S4F and S4G).

Because cholesterol esterification is confined to a defined set of enzymes, encoded by *lcat*, *soat1*, and *soat2*, we next assessed the effect of *agxt2* and *abat* knockdown in zebrafish on the expression of these genes (Figure S5). We found that both *agxt2* and *abat* knockdown resulted in decreased expression of *lcat* and *soat2*, consistent with the lower CE levels identified by LC-MS. Notably, humans with inherited *LCAT* deficiency develop both marked reductions in circulating CE levels as well as hypertriglyceridemia (Frohlich J et al., 1988). *Lcat* ablation in mice similarly results in hypertriglyceridemia, whereas transgenic *Lcat* overexpression results in lower plasma triglyceride levels (Ng D et al., 1997; Francone OL et al. 1995). Although the exact molecular pathways linking CE and TAG metabolism in these contexts have not been fully elucidated, Ng et al have shown that *Lcat* deficiency in mice results in 1.) increased triglyceride production, with increased expression of *Srebp-1*, *Fas*, and *Acc-1*, 2.) decreased triglyceride catabolism, with impaired lipase activity, and 3.)

increased expression of *Hmgcr* and decreased expression of *Soat2* (Ng D et al., 2004; Song H et al., 2006). In addition to lowering the expression of *lcat* and *soat2*, we found that both *agxt2* and *abat* knockdown in zebrafish resulted in increased expression of *srebp-1* and *hmgcr*, as well as decreased expression of *lipc* (Figure S5). Taken together, these data extend the results of gene-metabolite-phenotype data in FHS and highlight a functional link between β -aminoisobutyric acid, CE, and TAG metabolism in zebrafish.

DISCUSSION

Our platform surveys >100 metabolites not screened in prior GWAS, and extends recent efforts to annotate the common genetic determinants of circulating metabolite levels. Previously unmeasured metabolites include several distinct classes of lipid analytes for which we report numerous locus-metabolite associations, many in loci previously associated with human disease. Furthermore, using the rigorous characterization of clinical factors and family-based relationships of FHS participants, we delineate the relative contributions of inherited, environmental, and clinical factors on the metabolome. For select loci, we show that a broad view of metabolite associations provides insight on gene function, in some cases confirming known biochemical functions of the gene product (e.g., *FADS1-3*) and in others highlighting unanticipated metabolic roles (e.g., *AGXT2*).

For the majority of analytes, variation attributable to heritable factors is greater than that attributable to clinical factors, with the notable exception of the tobacco metabolite cotinine. In fact, heritability estimates for many metabolites are considerably higher than for traditional biomarkers, such as B-type natriuretic peptide ($h^2=0.35$) (Wang et al., 2003) or C-reactive protein ($h^2=0.30$) (Schnabel et al., 2009). In some cases, this highlights metabolites that serve as proximal reporters of underlying gene function. For example, the top SNP (rs37370) in *AGXT2* accounts for approximately a third of the estimated heritability for its enzyme substrate β -aminoisobutyric acid. The top SNPs for glycine (rs7422339, *CPS1*) and PCs 36:4 and 38:4 (rs102275, *FADS1-3*) account for nearly all of their heritability (Figure 1). For most metabolites, however, either no genome-wide significant association was identified or the top genome wide significant SNP explained only a small fraction of overall heritability. To what extent the unexplained heritability for these metabolites is attributable to common polymorphisms with sub-genome wide associations, the effect of rare variants or copy number variants not captured by SNPs in GWAS arrays, or other factors (including shared environmental factors) remains undetermined.

For select loci associated with human disease, e.g. *UMTS* and hereditary orotic aciduria, the locus-metabolite association identified in our study reflects the gene product's enzymatic function. By contrast, several loci with previously established disease associations have no enzymatic or transport function directly related to the associated metabolite. In these cases, the locus-metabolite association identified in our study may provide information on the pathophysiologic link between a given locus and disease (Adamski, 2012; Suhre and Gieger, 2012). For example, the *SLC7A9* locus, associated with NMMA in our study, encodes an amino acid transporter in the kidney with specificity for dibasic amino acids including cystine and arginine (Mora et al., 1996). Common variants in *SLC7A9* have been associated with CKD (Kottgen et al., 2010). However, CKD is not characterized by cystinuria or cystine stones, as with the Mendelian disorder attributable to *SLC7A9* mutations. Our data highlight plasma NMMA, a methylarginine that inhibits NO synthase (Vallance et al., 1992), as a potential intermediary between common variation at this locus and renal disease. Indeed, we find that elevated plasma levels of NMMA are associated with an increased risk of future CKD among individuals with normal kidney function at baseline. Thus, our data raise the hypothesis that NMMA could be both a biomarker and effector of CKD risk.

Because a narrow focus on only genome-wide significant associations is likely to overlook biologically meaningful findings, we also highlight a sub-genome-wide significant association between *KCNQ1*, previously associated with type 2 diabetes, and triiodothyronine levels. Notably, recent studies demonstrate an essential role for the KCNQ1 channel in thyroid I⁻ uptake (Purtell K et al., 2012). Further, *Kcnq1* ablation in mice has been shown to result in hypothyroidism (Frohlich H et al., 2011). In cross-sectional analyses, we find that plasma triiodothyronine levels are strongly correlated with metabolic parameters in FHS, including plasma insulin, plasma triglycerides, and BMI. Thus, in addition to raising an intriguing link between potassium channel function in the thyroid and type 2 diabetes risk, these findings reinforce the potential value of the full breadth and depth of our data.

We also show that a more granular lipidomic analysis informs findings from prior GWAS of bulk lipid measures. More specifically, we examine the association across all TAGs for select loci and identify distinct TAG signatures that highlight underlying gene function. The *FADS1-3* locus had a strong association with TAGs of high double bond content, as would be expected given the critical role the fatty acid desaturases encoded at this locus play in the synthesis of ω -3 and ω -6 polyunsaturated fatty acids. By contrast, the *GCKR* locus had a stronger association for TAGs of lower carbon and double bond content. We and others have previously shown that these TAGs are tightly correlated with insulin resistance, whereas TAGs with relatively higher carbon (and double bond) content are associated with increased insulin sensitivity (Kotronen et al., 2009; Rhee et al., 2011). The *APOA1/C3/A4/A5* locus was associated with TAGs of intermediate carbon content. Unlike TAGs at the lower and higher extremes of carbon content, these TAGs are comprised primarily of the abundant fatty acids palmitic acid, palmitoleic acid, stearic acid, and oleic acid, and as a result, make the largest quantitative contribution to total plasma triglyceride levels (i.e. plasma lipoprotein triglyceride content) (Rhee et al., 2011). Finally, rs6593086 demonstrated a striking pattern of TAG association – although no coding gene resides within 50kb of this SNP, our data predicts a role for this genomic region in the regulation of fatty acid desaturation and/or elongation.

In order to further demonstrate how a metabolomic approach is able to elucidate new biology, we tested the effects of *agxt2* and *abat* knockdown in zebrafish. Because cross-sectional analyses highlighted a strong correlation between β -aminoisobutyric acid levels and total triglycerides in the FHS, and because the *AGXT2* locus (a powerful determinant of β -aminoisobutyric acid levels) also had nominal associations with plasma TAGs and CEs, we hypothesized that modulating *agxt2* would impact β -aminoisobutyric acid levels and modulate lipid homeostasis in fish. Indeed, we noted a dose-dependent increase in TAGs and decrease in CEs following gene knockdown. Similar experiments using *abat* knockdown resulted in a similar decrement in β -aminoisobutyric acid levels, and recapitulated the increase in TAGs and decrease in CEs observed with *agxt2* knockdown. Both *agxt2* and *abat* knockdown resulted in decreased *lcat* and *soat2* expression, increased *srebp-1* expression, and decreased *lipc* expression, all findings that would be expected to result in the decreased CE and increased TAG levels observed in these fish. Although further studies are needed to understand how modulating β -aminoisobutyric acid, a catabolite of thymine and valine, impacts the expression of *lcat*, *srebp-1*, and other fundamental effectors of lipid metabolism, our data provide an example of how the breadth of gene, metabolite, and phenotype data we present in the FHS can provide a springboard for research in metabolism.

Because our platform measures numerous metabolites not surveyed in prior metabolomics GWAS, additional profiling in other cohorts will be required to validate our findings. However, we note that our study does recapitulate 8 locus-metabolite associations reported

in prior metabolomics GWAS. Further, 8 of our novel findings highlight loci encoding enzymes or transporters directly linked to the associated metabolite, providing internal validation by way of biological plausibility. Nevertheless, future efforts will seek to provide external replication for several of the associations identified in this study. As the FHS includes middle-aged to older individuals of predominantly European ancestry, future studies should also include more diverse populations in order to test the generalizability of our findings.

In summary, we performed a metabolomics GWAS on plasma obtained from 2,076 individuals in a large, community-based cohort. This study corroborates prior data demonstrating the relatively high heritability estimates for circulating metabolites, and is the first to compare these measures against the contribution of clinical covariates. Further, we report 31 loci with significant metabolite associations, and show how examining association patterns across the metabolome for select loci may begin to provide insights in metabolism. Importantly, we include GWAS data on all metabolites surveyed, as well as full metabolite data for each locus with at least one significant association, as a resource for the scientific community. We anticipate that future efforts to understand and synthesize sub-genome wide significant findings, and to understand patterns of metabolite associations, will further elucidate the genetic determinants of the plasma metabolome and provide new insights about gene function and disease pathogenesis in the context of human metabolism.

EXPERIMENTAL PROCEDURES

Study Sample

The Framingham Heart Study (FHS) Offspring cohort is a prospective, observational, community-based cohort (Kannel et al., 1979). The children of FHS participants and their spouses were recruited in 1971 and have been followed with serial examinations. A total of 2,076 participants of European descent who attended the fifth examination (1991–1995) and underwent metabolic profiling and genome-wide genotyping were included in this analysis. All participants provided informed consent and the study protocol was approved by the Boston University Medical Center IRB.

Clinical Assessment

Participants underwent a comprehensive medical history, physical examination, and anthropometry at the fifth examination. Diabetes mellitus was defined as a fasting glucose ≥ 126 mg/dL, non-fasting blood glucose ≥ 200 mg/dL, or the use of insulin or oral antidiabetic medications. Participants who had smoked regularly during the prior year were considered current smokers. Fasting total and high-density lipoprotein cholesterol levels were obtained. Estimated glomerular filtration rate (eGFR) was calculated using the Modification of Diet in Renal Disease equation (Levey et al., 1999). Previous cardiovascular events were adjudicated by a 3-physician panel after review of medical records, and included a history of coronary heart disease, heart failure, and stroke.

Genome-wide Genotyping and Imputation

Genotyping methods have previously been described (Wilk et al., 2009). In brief, genome-wide genotyping was conducted using the Affymetrix 500K mapping array and the Affymetrix 50K gene-focused MIP array. Genotypes were called using Chiamo (<http://www.stats.ox.ac.uk/~marchini/software/gwas/chiamo.html>). Imputation of 2.5 million SNPs was then performed (HapMap CEU population, release 22, build 36; <http://hapmap.org>) using a hidden Markov model that was implemented in MACH (version 1.0.15) (Li et al., 2010).

Metabolite profiling

Blood samples were collected after an overnight fast, immediately centrifuged, and stored at -80°C until assayed. Amino acids, amino acid derivatives, urea cycle intermediates, nucleotides and other positively charged polar metabolites were profiled as previously described using 10 μL of plasma (Wang et al., 2011). Lipids were profiled as previously described using 10 μL of plasma (Rhee et al., 2011). For each lipid analyte, the first number denotes the total number of carbons in the lipid acyl chain(s), and the second number (after the colon) denotes the total number of double bonds in the lipid acyl chain(s). For organic acids, sugars, bile acids, and other negatively charged polar metabolites, 30 μL of plasma were used and MS data were acquired using ESI and MRM in the negative ion mode (details provided in the Supplemental Experimental Procedures). Table S5 lists the 217 metabolites measured by our platform and specifies overlap with prior metabolomics-GWAS.

Statistical Analysis

The percent inter-individual variability in log-transformed metabolite concentrations accounted for by measured clinical factors (R^2) was assessed using multivariable linear regression models, adjusted for age, sex, systolic blood pressure, anti-hypertensive medication use, body-mass index, diabetes, smoking status, and prevalent cardiovascular disease. In secondary analyses, we further adjusted for eGFR. Analyses were performed using SAS, version 9.1.3 (SAS Institute, Cary, NC).

Due to right-skewed distributions of metabolite levels and differences in scaling, genetic analyses were conducted using normalized residuals of metabolite levels, adjusted for age and sex. Heritability of each metabolite was estimated using variance-component models and Sequential Oligogenic Linkage Analysis Routines (SOLAR) (Almasy and Blangero, 1998). The association of genetic variants and metabolite concentrations was tested using linear mixed effects models to accommodate pedigree data under an additive genetic model. Genome-wide association analyses were performed using R (Chen and Yang, 2010), and implemented using the *lmekin* function in the kinship package (Therneau et al., 2003). Population stratification was accounted for by adjusting for PC1 if $P < 0.0001$, and the final genomic control parameter lambda was < 1.03 for all analyses. Results were filtered for minor allele frequency $> 5\%$ and imputation ratio of > 0.80 . Results were considered genome-wide significant at $P < 5.0 \times 10^{-8}$.

Zebrafish studies

All zebrafish studies were approved by the Massachusetts General Hospital Subcommittee on Research Animal Care. The splice blocking *agxt2* and *abat* morpholino oligonucleotides (MOs) were designed to target the intron1-exon2 junction and the intron4-exon5 junction (Figure S4) in *agxt2* and *abat*, respectively (Gene Tools, LLC), and were injected into one- to two-cell stage zebrafish embryos using an Eppendorf Femtojet microinjector. 48 hours post fertilization embryos were utilized for β -aminoisobutyric acid measurement and 5 days post fertilization larvae were used for CE and TAG related gene expression and lipid profiling. Metabolite and gene expression data were compared across groups using two-tailed t-tests.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH contracts N01-HC-25195, R01-DK-HL-081572 (R.E.G. and T.J.W.), R01-HL-098280 (R.E.G.), R-01-HL-093328 (R.S.V.), U01-HL-107440 (R.E.G.), K23-HL-116780 (J.E.H.), K08-DK-090142 (E.P.R.), the Leducq Foundation (to R.E.G.), and the American Heart Association 12IRG9130006 (J.-R.J.Y and I.T.H) and 0940109N (R.E.G.). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center.

REFERENCES

- Abhary S, Burdon KP, Kuot A, Javadiyan S, Whiting MJ, Kasmeridis N, Petrovsky N, Craig JE. Sequence variation in DDAH1 and DDAH2 genes is strongly and additively associated with serum ADMA concentrations in individuals with type 2 diabetes. *PLoS One*. 2010; 5:e9462. [PubMed: 20209122]
- Adamski J. Genome-wide association studies with metabolomics. *Genome Med*. 2012; 4:34. [PubMed: 22546499]
- Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet*. 1998; 62:1198–1211. [PubMed: 9545414]
- Beer NL, Tribble ND, McCulloch LJ, Roos C, Johnson PR, Orho-Melander M, Gloyn AL. The P446L variant in GCKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. *Hum. Mol. Genet*. 2009; 18:4081–4088. [PubMed: 19643913]
- Bellenguez C, Bevan S, Gschwendtner A, Spencer CC, Burgess AI, Pirinen M, Jackson CA, Traylor M, Strange A, Su Z, et al. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat. Genet*. 2012; 44:328–333. [PubMed: 22306652]
- Chambers JC, Zhang W, Lord GM, van der Harst P, Lawlor DA, Sehmi JS, Gale DP, Wass MN, Ahmadi KR, Bakker SJ, et al. Genetic loci influencing kidney function and chronic kidney disease. *Nat. Genet*. 2010; 42:373–375. [PubMed: 20383145]
- Chen MH, Yang Q. GWAFA: an R package for genome-wide association analyses with family data. *Bioinformatics*. 2010; 26:580–581. [PubMed: 20040588]
- Danik JS, Pare G, Chasman DI, Zee RY, Kwiatkowski DJ, Parker A, Miletich JP, Ridker PM. Novel loci, including those related to Crohn disease, psoriasis, and inflammation, identified in a genome-wide association study of fibrinogen in 17 686 women: the Women's Genome Health Study. *Circ. Cardiovasc. Genet*. 2009; 2:134–141. [PubMed: 20031577]
- Dehghan A, Dupuis J, Barbalic M, Bis JC, Eiriksdottir G, Lu C, Pellikka N, Wallaschofski H, Kettunen J, Henneman P, et al. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation*. 2011; 123:731–738. [PubMed: 21300955]
- Demirkan A, van Duijn CM, Ugocsai P, Isaacs A, Pramstaller PP, Liebisch G, Wilson JF, Johansson A, Rudan I, Aulchenko YS, et al. Genome-wide association study identifies novel loci associated with circulating phospho- and sphingolipid concentrations. *PLoS Genet*. 2012; 8:e1002490. [PubMed: 22359512]
- Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet*. 2010; 42:105–116. [PubMed: 20081858]
- Eijgelsheim M, Newton-Cheh C, Sotoodehnia N, de Bakker PI, Muller M, Morrison AC, Smith AV, Isaacs A, Sanna S, Dorr M, et al. Genome-wide association analysis identifies multiple loci related to resting heart rate. *Hum. Mol. Genet*. 2010; 19:3885–3894. [PubMed: 20639392]
- Ellinor PT, Lunetta KL, Albert CM, Glazer NL, Ritchie MD, Smith AV, Arking DE, Muller-Nurasyid M, Krijthe BP, Lubitz SA, et al. Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nat. Genet*. 2012; 44:670–675. [PubMed: 22544366]
- Estrada K, Stykarsdottir U, Evangelou E, Hsu YH, Duncan EL, Ntzani EE, Oei L, Albagha OM, Amin N, Kemp JP, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat. Genet*. 2012; 44:491–501. [PubMed: 22504420]

- Feliubadalo L, Font M, Purroy J, Rousaud F, Estivill X, Nunes V, Golomb E, Centola M, Aksentijevich I, Kreiss Y, et al. Non-type I cystinuria caused by mutations in SLC7A9, encoding a subunit (bo,+AT) of rBAT. *Nat. Genet.* 1999; 23:52–57. [PubMed: 10471498]
- Francone OL, Gong EL, Ng DS, Fielding CJ, Rubin EM. Expression of human lecithin-cholesterol acyltransferase in transgenic mice. Effect of human apolipoprotein AI and human apolipoprotein all on plasma lipoprotein cholesterol metabolism. *J. Clin. Invest.* 1995; 96:1440–8. [PubMed: 7657816]
- Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, Lees CW, Balschun T, Lee J, Roberts R, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet.* 2010; 42:1118–1125. [PubMed: 21102463]
- Frohlich H, Boini KM, Seebohm G, Strutz-Seebohm N, Ureche ON, Foller M, Eichenmuller M, Shumilina E, Pathare G, Singh AK, et al. Hypothyroidism of gene-targeted mice lacking Kcnq1. *Pflugers Arch.* 2011; 461:45–52. [PubMed: 20978783]
- Frohlich J, McLeod R, Pritchard PH, Fesmire J, McConathy W. Plasma lipoprotein abnormalities in heterozygotes for familial lecithin:cholesterol acyltransferase deficiency. *Metabolism.* 1988; 37:3–8. [PubMed: 3121980]
- Gieger C, Geistlinger L, Altmaier E, de Angelis MH, Kronenberg F, Meitinger T, Mewes H-W, Wichmann H-E, Weinberger KM, Adamski J, et al. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet.* 2008; 4:e1000282. [PubMed: 19043545]
- Gieger C, Radhakrishnan A, Cvejic A, Tang W, Porcu E, Pistis G, Serbanovic-Canic J, Elling U, Goodall AH, Labrune Y, et al. New gene functions in megakaryopoiesis and platelet formation. *Nature.* 2011; 480:201–208. [PubMed: 22139419]
- Gottlieb DJ, O'Connor GT, Wilk JB. Genome-wide association of sleep and circadian phenotypes. *BMC Med. Genet.* 2007; 8(Suppl 1):S9. [PubMed: 17903308]
- Hicks AA, Pramstaller PP, Johansson A, Vitart V, Rudan I, Ugozai P, Aulchenko Y, Franklin CS, Liebisch G, Erdmann J, et al. Genetic determinants of circulating sphingolipid concentrations in European populations. *PLoS Genet.* 2009; 5:e1000672. [PubMed: 19798445]
- Himes BE, Hunninghake GM, Baurley JW, Rafaels NM, Sleiman P, Strachan DP, Wilk JB, Willis-Owen SA, Klanderman B, Lasky-Su J, et al. Genome-wide association analysis identifies PDE4D as an asthma-susceptibility gene. *Am. J. Hum. Genet.* 2009; 84:581–593. [PubMed: 19426955]
- Hu X, Atzler D, Xu X, Zhang P, Guo H, Lu Z, Fassett J, Schwedhelm E, Boger RH, Bache RJ, et al. Dimethylarginine dimethylaminohydrolase-1 is the critical enzyme for degrading the cardiovascular risk factor asymmetrical dimethylarginine. *Arterioscler. Thromb. Vasc. Biol.* 2011; 31:1540–1546. [PubMed: 21493890]
- Illig T, Gieger C, Zhai G, Römisch-Margl W, Wang-Sattler R, Prehn C, Altmaier E, Kastenmüller G, Kato BS, Mewes H-W, et al. A genome-wide perspective of genetic variation in human metabolism. *Nat. Genet.* 2010; 42:137–141. [PubMed: 20037589]
- Imboden M, Bouzigon E, Curjuric I, Ramasamy A, Kumar A, Hancock DB, Wilk JB, Vonk JM, Thun GA, Siroux V, et al. Genome-wide association study of lung function decline in adults with and without asthma. *J. Allergy Clin. Immunol.* 2012; 129:1218–1228. [PubMed: 22424883]
- Johnson AD, Kavousi M, Smith AV, Chen MH, Dehghan A, Aspelund T, Lin JP, van Duijn CM, Harris TB, Cupples LA, et al. Genome-wide association meta-analysis for total serum bilirubin levels. *Hum. Mol. Genet.* 2009; 18:2700–2710. [PubMed: 19414484]
- Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, Nakamura Y, Kamatani N. Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat. Genet.* 2010; 42:210–215. [PubMed: 20139978]
- Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. *Am. J. Epidemiol.* 1979; 110:281–290. [PubMed: 474565]
- Kettunen J, Tukiainen T, Sarin AP, Ortega-Alonso A, Tikkanen E, Lyytikäinen LP, Kangas AJ, Soininen P, Wurtz P, Silander K, et al. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat. Genet.* 2012; 44:269–276. [PubMed: 22286219]

- Kim YJ, Go MJ, Hu C, Hong CB, Kim YK, Lee JY, Hwang JY, Oh JH, Kim DJ, Kim NH, et al. Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. *Nat. Genet.* 2011; 43:990–995. [PubMed: 21909109]
- Kolz M, Johnson T, Sanna S, Teumer A, Vitart V, Perola M, Mangino M, Albrecht E, Wallace C, Farrall M, et al. Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. *PLoS Genet.* 2009; 5:e1000504. [PubMed: 19503597]
- Kotronen A, Velagapudi VR, Yetukuri L, Westerbacka J, Bergholm R, Ekroos K, Makkonen J, Taskinen M-R, Oresic M, Yki-Järvinen H. Serum saturated fatty acids containing triacylglycerols are better markers of insulin resistance than total serum triacylglycerol concentrations. *Diabetologia.* 2009; 52:684–690. [PubMed: 19214471]
- Kottgen A, Pattaro C, Boger CA, Fuchsberger C, Olden M, Glazer NL, Parsa A, Gao X, Yang Q, Smith AV, et al. New loci associated with kidney function and chronic kidney disease. *Nat. Genet.* 2010; 42:376–384. [PubMed: 20383146]
- Kristiansson K, Perola M, Tikkanen E, Kettunen J, Surakka I, Havulinna AS, Stancakova A, Barnes C, Widen E, Kajantie E, et al. Genome-wide screen for metabolic syndrome susceptibility Loci reveals strong lipid gene contribution but no evidence for common genetic basis for clustering of metabolic syndrome traits. *Circ. Cardiovasc. Genet.* 2012; 5:242–249. [PubMed: 22399527]
- Lange LA, Croteau-Chonka DC, Marville AF, Qin L, Gaulton KJ, Kuzawa CW, McDade TW, Wang Y, Li Y, Levy S, et al. Genome-wide association study of homocysteine levels in Filipinos provides evidence for CPS1 in women and a stronger MTHFR effect in young adults. *Hum. Mol. Genet.* 2010; 19:2050–2058. [PubMed: 20154341]
- Lemaitre RN, Tanaka T, Tang W, Manichaikul A, Foy M, Kabagambe EK, Nettleton JA, King IB, Weng LC, Bhattacharya S, et al. Genetic loci associated with plasma phospholipid n-3 fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium. *PLoS Genet.* 2011; 7:e1002193. [PubMed: 21829377]
- Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. *Ann. Intern. Med.* 1999; 130:461–470. [PubMed: 10075613]
- Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* 2010; 34:816–34. [PubMed: 21058334]
- Link E, Parish S, Armitage J, Bowman L, Heath S, Matsuda F, Gut I, Lathrop M, Collins R. SLCO1B1 variants and statin-induced myopathy--a genomewide study. *N. Engl. J. Med.* 2008; 359:789–799. [PubMed: 18650507]
- Mora C, Chillaron J, Calonge MJ, Forgo J, Testar X, Nunes V, Murer H, Zorzano A, Palacin M. The rBAT gene is responsible for L-cystine uptake via the b0,(+)-like amino acid transport system in a “renal proximal tubular” cell line (OK cells). *J. Biol. Chem.* 1996; 271:10569–10576. [PubMed: 8631857]
- Neale BM, Fagerness J, Reynolds R, Sobrin L, Parker M, Raychaudhuri S, Tan PL, Oh EC, Merriam JE, Souied E, et al. Genome-wide association study of advanced age-related macular degeneration identifies a role of the hepatic lipase gene (LIPC). *Proc. Natl. Acad. Sci. USA.* 2010; 107:7395–7400. [PubMed: 20385826]
- Ng DS, Francone OL, Forte TM, Zhang J, Haghpassand M, Rubin EM. Disruption of the murine lecithin:cholesterol acyltransferase gene causes impairment of adrenal lipid delivery and up-regulation of scavenger receptor class B type I. *J. Biol. Chem.* 1997; 272:15777–81. [PubMed: 9188474]
- Ng DS, Xie C, Maguire GF, Zhu X, Ugwu F, Lam E, Connelly PW. Hypertriglyceridemia in lecithin-cholesterol acyltransferase-deficient mice is associated with hepatic overproduction of triglycerides, increased lipogenesis, and improved glucose tolerance. *J. Biol. Chem.* 2004; 279:7636–42. [PubMed: 14668345]
- Orho-Melander M, Melander O, Guiducci C, Perez-Martinez P, Corella D, Roos C, Tewhey R, Rieder MJ, Hall J, Abecasis G, et al. Common missense variant in the glucokinase regulatory protein gene is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations. *Diabetes.* 2008; 57:3112–3121. [PubMed: 18678614]

- Purtell K, Paroder-Belenitsky M, Reyna-Neyra A, Nicola JP, Koba W, Fine E, Carrasco N, Abbott GW. The KCNQ1-KCNE2 K⁺ channel is required for adequate thyroid I uptake. *FASEB J.* 2012; 26:3252–9. [PubMed: 22549510]
- Rhee EP, Cheng S, Larson MG, Walford GA, Lewis GD, McCabe E, Yang E, Farrell L, Fox CS, O'Donnell CJ, et al. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. *J. Clin. Invest.* 2011; 121:1402–1411. [PubMed: 21403394]
- Rhee EP, Gerszten RE. Metabolomics and cardiovascular biomarker discovery. *Clin. Chem.* 2012; 58:139–147. [PubMed: 22110018]
- Rhee EP, Clish CB, Ghorbani A, Larson MG, Elmariah S, McCabe E, Yang Q, Cheng S, Pierce K, Deik A, et al. A Combined Epidemiologic and Metabolomic Approach Improves CKD Prediction. *J. Am. Soc. Nephrol.* 2013 epub ahead of print.
- Ridker PM, Pare G, Parker A, Zee RY, Danik JS, Buring JE, Kwiatkowski D, Cook NR, Miletich JP, Chasman DI. Loci related to metabolic-syndrome pathways including LEPR, HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: the Women's Genome Health Study. *Am. J. Hum. Genet.* 2008; 82:1185–1192. [PubMed: 18439548]
- Sanna S, Busonero F, Maschio A, McArdle PF, Usala G, Dei M, Lai S, Mulas A, Piras MG, Perseu L, et al. Common variants in the SLCO1B3 locus are associated with bilirubin levels and unconjugated hyperbilirubinemia. *Hum. Mol. Genet.* 2009; 18:2711–2718. [PubMed: 19419973]
- Saxena R, Hivert MF, Langenberg C, Tanaka T, Pankow JS, Vollenweider P, Lyssenko V, Bouatia-Naji N, Dupuis J, Jackson AU, et al. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat. Genet.* 2010; 42:142–148. [PubMed: 20081857]
- Schnabel RB, Lunetta KL, Larson MG, Dupuis J, Lipinska I, Rong J, Chen MH, Zhao Z, Yamamoto JF, Meigs JB, et al. The relation of genetic and environmental factors to systemic inflammatory biomarker concentrations. *Circ. Cardiovasc. Genet.* 2009; 2:229–237. [PubMed: 20031590]
- Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, Preuss M, Stewart AF, Barbalic M, Gieger C, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* 2011; 43:333–338. [PubMed: 21378990]
- Shah SH, Hauser ER, Bain JR, Muehlbauer MJ, Haynes C, Stevens RD, Wenner BR, Dowdy ZE, Granger CB, Ginsburg GS, et al. High heritability of metabolomic profiles in families burdened with premature cardiovascular disease. *Mol. Syst. Biol.* 2009; 5
- Song H, Zhu L, Picardo CM, Maguire G, Leung V, Connelly PW, Ng DS. Coordinated alteration of hepatic gene expression in fatty acid and triglyceride synthesis in LCAT-null mice is associated with altered PUFA metabolism. *Am. J. Physiol. Endocrinol. Metab.* 2006; 290:E17–E25. [PubMed: 16105858]
- Song Q, Cole JW, O'Connell JR, Stine OC, Gallagher M, Giles WH, Mitchell BD, Wozniak MA, Stern BJ, Sorkin JD, et al. Phosphodiesterase 4D polymorphisms and the risk of cerebral infarction in a biracial population: the Stroke Prevention in Young Women Study. *Hum. Mol. Genet.* 2006; 15:2468–2478. [PubMed: 16835261]
- Suhre K, Gieger C. Genetic variation in metabolic phenotypes: study designs and applications. *Nat. Rev. Genet.* 2012; 13:759–769. [PubMed: 23032255]
- Suhre K, Shin S-Y, Petersen A-K, Mohny RP, Meredith D, Wägele B, Altmaier E, Deloukas P, Erdmann J, et al. CARDIoGRAM. Human metabolic individuality in biomedical and pharmaceutical research. *Nature.* 2011a; 477:54–60. [PubMed: 21886157]
- Suhre K, Wallaschofski H, Raffler J, Friedrich N, Haring R, Michael K, Wasner C, Krebs A, Kronenberg F, Chang D, et al. A genome-wide association study of metabolic traits in human urine. *Nat. Genet.* 2011b; 43:565–569. [PubMed: 21572414]
- Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature.* 2010; 466:707–713. [PubMed: 20686565]
- Therneau TM, Grambsch PM, Pankratz VS. Penalized survival models and frailty. *J. Comput. Graph. Stat.* 2003; 12:156–175.
- Tukiainen T, Kettunen J, Kangas AJ, Lyytikäinen LP, Soininen P, Sarin AP, Tikkanen E, O'Reilly PF, Savolainen MJ, Kaski K, et al. Detailed metabolic and genetic characterization reveals new

- associations for 30 known lipid loci. *Hum. Mol. Genet.* 2012; 21:1444–1455. [PubMed: 22156771]
- Unoki H, Takahashi A, Kawaguchi T, Hara K, Horikoshi M, Andersen G, Ng DP, Holmkvist J, Borch-Johnsen K, Jorgensen T, et al. SNPs in *KCNQ1* are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat. Genet.* 2008; 40:1098–102. [PubMed: 18711366]
- Vallance P, Palmer RM, Moncada S. The role of induction of nitric oxide synthesis in the altered responses of jugular veins from endotoxaemic rabbits. *Br. J. Pharmacol.* 1992; 106:459–63. [PubMed: 1393271]
- Wang TJ, Larson MG, Levy D, Benjamin EJ, Corey D, Leip EP, Vasani RS. Heritability and genetic linkage of plasma natriuretic peptide levels. *Circulation.* 2003; 108:13–16. [PubMed: 12821537]
- Wang TJ, Larson MG, Vasani RS, Cheng S, Rhee EP, McCabe E, Lewis GD, Fox CS, Jacques PF, Fernandez C, et al. Metabolite profiles and the risk of developing diabetes. *Nat. Med.* 2011; 17:448–453. [PubMed: 21423183]
- Waterworth DM, Ricketts SL, Song K, Chen L, Zhao JH, Ripatti S, Aulchenko YS, Zhang W, Yuan X, Lim N, et al. Genetic variants influencing circulating lipid levels and risk of coronary artery disease. *Arterioscler. Thromb. Vasc. Biol.* 2010; 30:2264–2276. [PubMed: 20864672]
- Wilk JB, Chen TH, Gottlieb DJ, Walter RE, Nagle MW, Brandler BJ, Myers RH, Borecki IB, Silverman EK, Weiss ST, et al. A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet.* 2009; 5:e1000429. [PubMed: 19300500]
- Wu C, Hu Z, He Z, Jia W, Wang F, Zhou Y, Liu Z, Zhan Q, Liu Y, Yu D, et al. Genome-wide association study identifies three new susceptibility loci for esophageal squamous-cell carcinoma in Chinese populations. *Nat. Genet.* 2011; 43:679–684. [PubMed: 21642993]
- Yang Q, Kottgen A, Dehghan A, Smith AV, Glazer NL, Chen MH, Chasman DI, Aspelund T, Eiriksdottir G, Harris TB, et al. Multiple genetic loci influence serum urate levels and their relationship with gout and cardiovascular disease risk factors. *Circ. Cardiovasc. Genet.* 2010; 3:523–530. [PubMed: 20884846]
- Yasuda K, Miyake K, Horikawa Y, Hara K, Osawa H, Furuta H, Hirota Y, Mori H, Jonsson A, Sato Y, et al. Variants in *KCNQ1* are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet.* 2008; 40:1092–7. [PubMed: 18711367]

HIGHLIGHTS

- 31 genetic loci have associations with 64 distinct plasma metabolites
- Heritability has a greater impact than clinical factors on metabolite variance
- GWAS data for all metabolites (loci with $P < 1.0 \times 10^{-3}$) are provided as a resource
- Integrating gene-metabolite data highlights a role for *AGXT2* in lipid metabolism

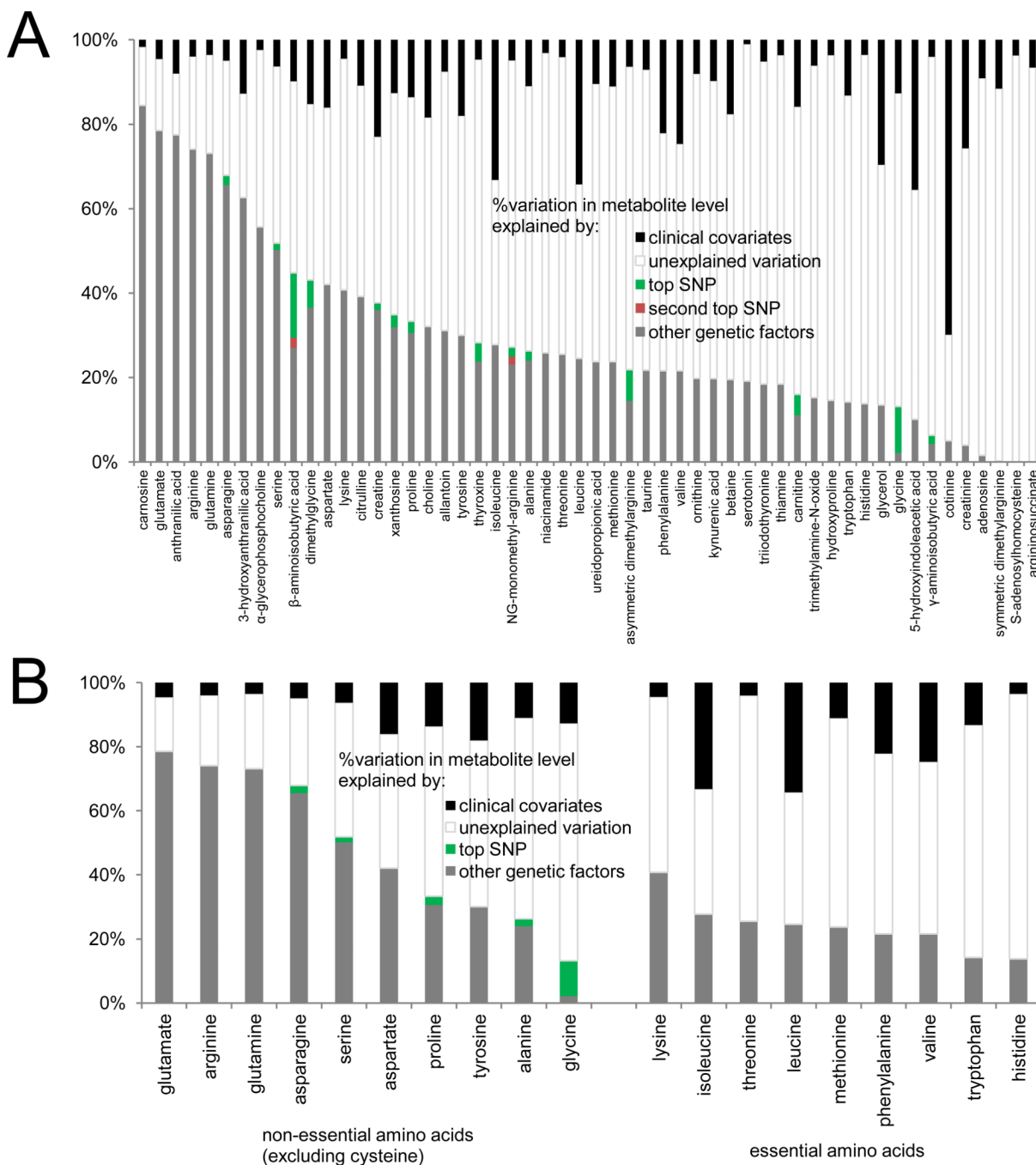


Figure 1. see also Figure S1 and Table S1. Genetic architecture of the human plasma metabolome

(A) The percent inter-individual variation for positively charged, polar metabolites explained by clinical (black) and genetic factors: top SNP (green), second top SNP (red), other genetic factors (grey). Clinical factors include age, sex, systolic blood pressure, anti-hypertensive medication use, diabetes mellitus, smoking status, body-mass index, and prevalent cardiovascular disease. (B) Data summarized for non-essential and essential amino acids.

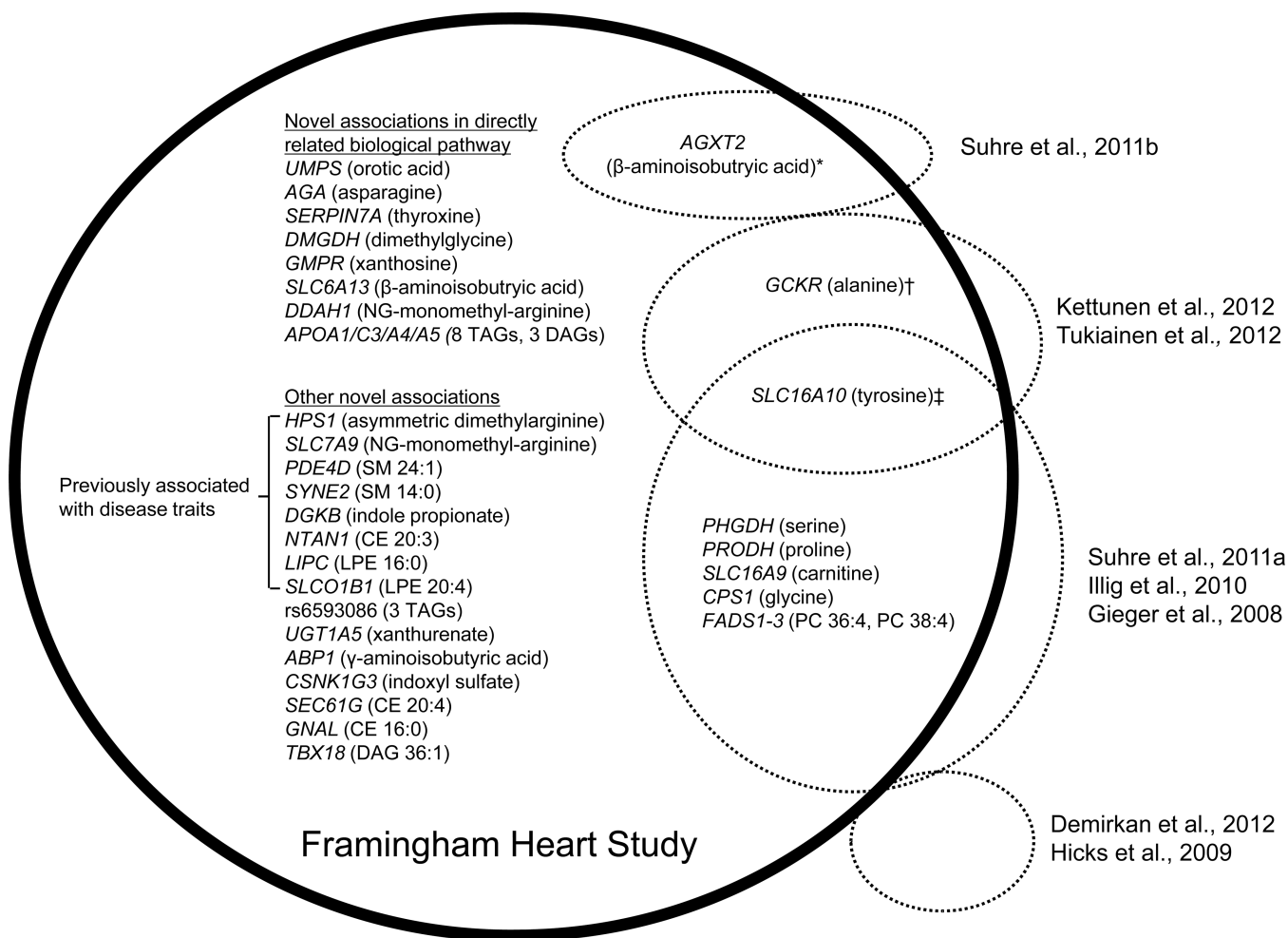


Figure 2. see also Figures S2 and S3. Thirty-one genome-wide significant loci associated with metabolites in FHS

Loci with significant metabolite associations in FHS are depicted within the black circle. Overlap with prior studies (dotted circles) is indicated for 8 loci. Prior studies found an association between common variants at **AGXT2* and urinary levels of β-aminoisobutyric acid, †*GCKR* and the alanine/glutamine ratio, and ‡*SLC16A10* and the tyrosine/isoleucine and alanine/tyrosine ratios.

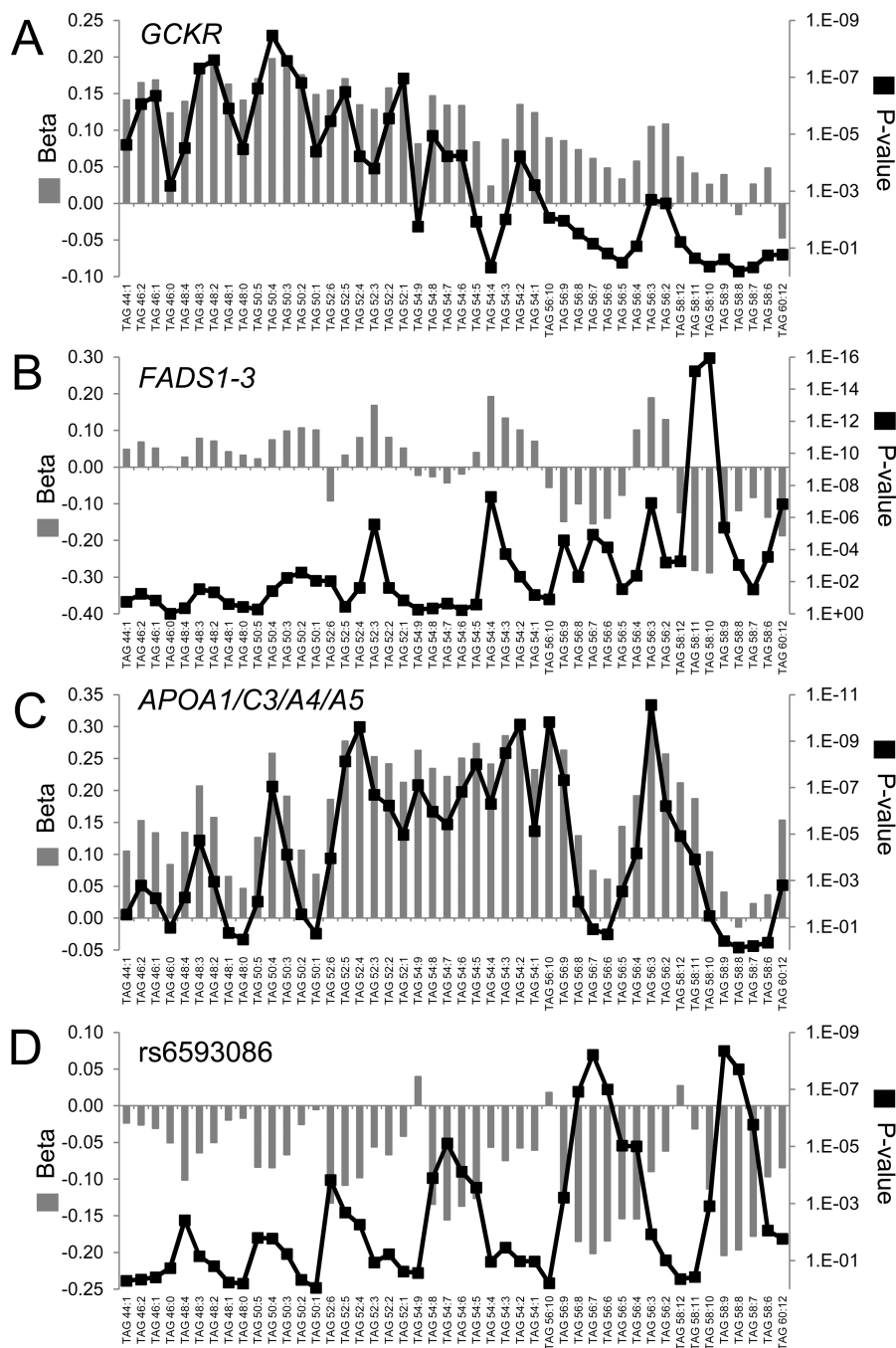


Figure 3. Metabolite profiling demonstrates distinct patterns of TAG associations for select loci For the 46 triacylglycerols (TAGs) monitored by our platform, the beta coefficient (left y-axis) and P-value (right y-axis) of association are shown for the top variant at (A) *GCKR*, (B) *FADS1-3*, (C) *APOA1/C3/A4/A5*, and (D) rs6593086.

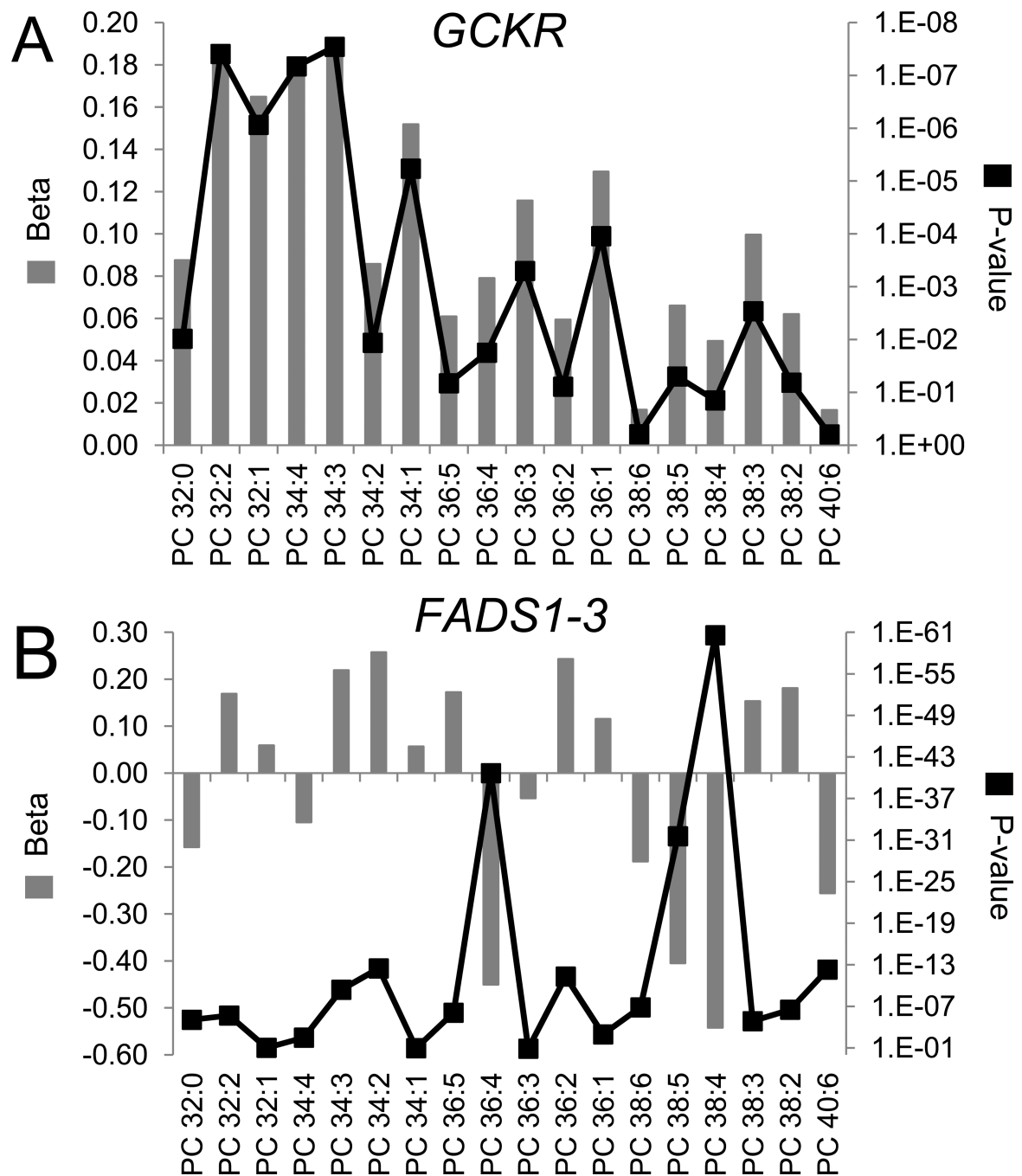


Figure 4. Phosphatidylcholine associations for the *GCKR* and *FADS1-3* loci

For the 18 phosphatidylcholines (PCs) monitored by our platform, the beta coefficient (left y-axis) and P-value (right y-axis) of association are shown for the top variant at (A) *GCKR* and (B) *FADS1-3*.

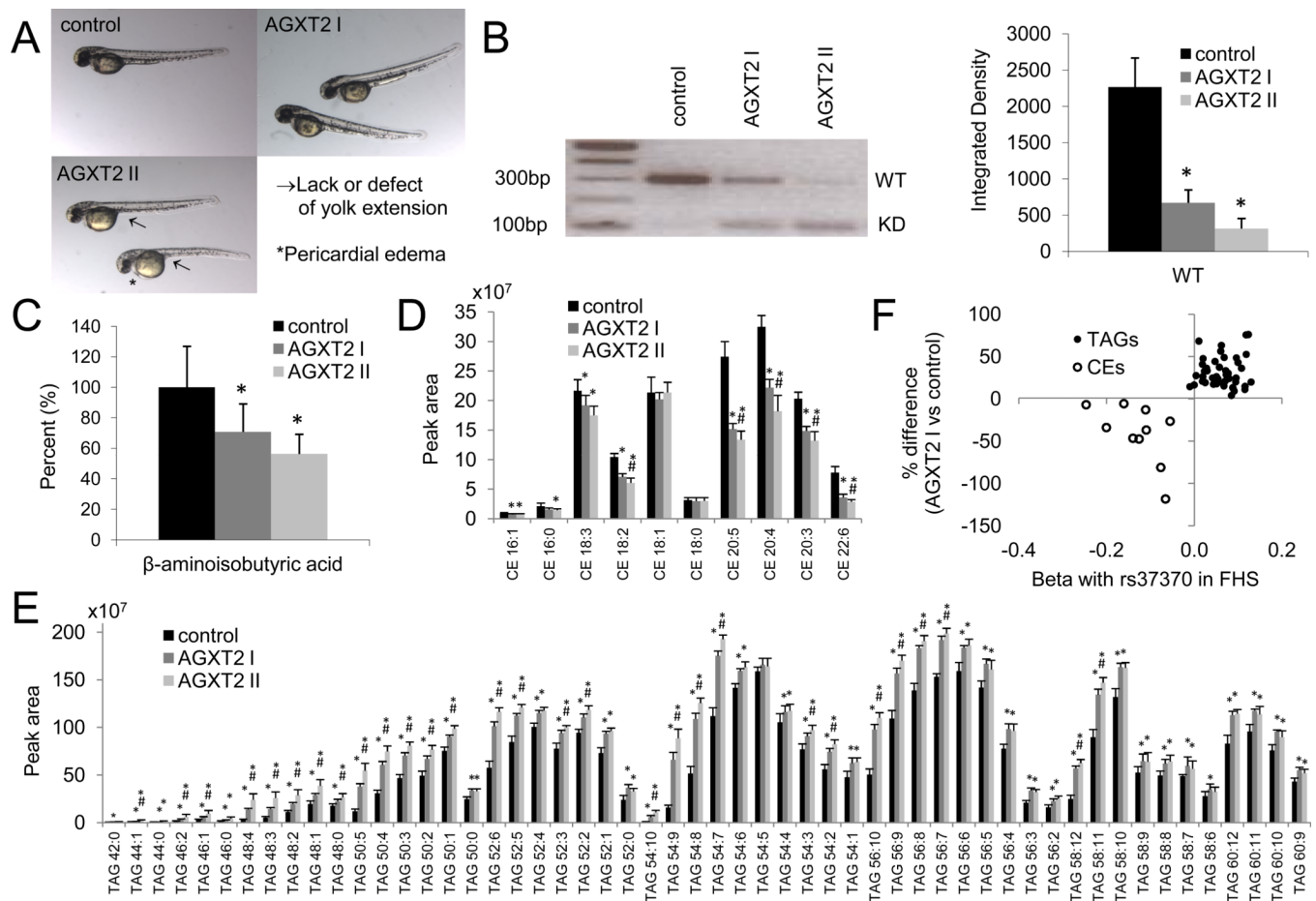


Figure 5. see also Figures S4 and S5 and Table S4. AGXT2 modulates lipid homeostasis in zebrafish

(A) Zebrafish embryos were injected with 2nL of a 400uM solution of a splice blocking *agxt2* morpholino oligonucleotide (MO) or control MO. At 48 hpf, *agxt2* MO injection resulted in morphants with both normal phenotype (upper right panel, “AGXT2 I”) and morphants with defective yolk sac extension and pericardial edema (lower left panel, “AGXT2 II”). (B) RT-PCR of *agxt2* mRNA for control, AGXT2 I and AGXT2 II morphants at 48 hpf; the 300 nucleotide fragment is the WT *agxt2* transcript and the 109 nucleotide fragment is the knocked-down (KD) *agxt2* transcript. The relative extent of WT *agxt2* expression, normalized to *rpl13-α*, was estimated from band intensities; data are presented as mean ± SD, *P<0.05 for comparison to control. (C) β-aminoisobutyric acid levels in control, AGXT2 I and AGXT2 II morphants at 48 hpf; data are presented as mean ± SD, *P<0.05 for comparison to control. (D) CE and (E) TAG levels for control, AGXT2 I and AGXT II larvae at 5 dpf; data are presented as mean ± SD, *P<0.05 for comparison to control, #P<0.05 for comparison to AGXT2 I. (F) For TAGs (solid circles) and CEs (hollow circles), plot of each metabolite’s association with rs37370 (*AGXT2*) in FHS on the x-axis versus the percent difference for each metabolite in AGXT I versus control zebrafish on the y-axis.

Table 1**Sample clinical characteristics**

Characteristics of the 2,076 individuals in the Framingham Offspring Cohort who underwent metabolite profiling and GWAS are shown. Data represent means (standard deviation) unless otherwise noted.

| Characteristic | N=2,076 |
|---|----------------|
| Age, years | 55 (10) |
| Women, n (%) | 1062 (51) |
| Systolic blood pressure, mmHg | 127 (19) |
| Diastolic blood pressure, mmHg | 75 (10) |
| Body-mass index, kg/m ² | 27.6 (5.0) |
| Antihypertensive treatment, n (%) | 424 (20) |
| Diabetes mellitus, n (%) | 137 (7) |
| Current smoker, n (%) | 396 (19) |
| Prevalent cardiovascular disease, n (%) | 49 (2) |
| Total cholesterol, mg/dl | 207 (37) |
| HDL cholesterol, mg/dl | 49 (15) |

Table 2
see also Figures S2 and S3 and Tables S2 and S3. Genome-wide significant loci associated with metabolites

31 genetic loci with at least one genome-wide significant metabolite association are shown. Citations are provided for previously established associations between these loci and other traits.

| Locus | Trait | SNP | ch | Major/ Allele | MAF | Beta | s.e. | P-value | Previous genome-wide associations |
|-----------------|-------------------------------|-----------|------|------------------|------|---------|------|---------|--|
| <i>AGXT2</i> | β -aminoisobutyric acid | rs37370 | 5 | T/C | 0.10 | 1.02 | 0.05 | 5.8E-83 | β -aminoisobutyric acid (Suhre et al., 2011b) |
| <i>FADS1-3</i> | PC 38:4 | rs174545 | 11 | C/G | 0.33 | -0.54 | 0.03 | 2.7E-61 | Phospholipids (Lemaitre et al., 2011) |
| | PC 36:4 | rs102275 | | T/C | 0.33 | -0.45 | 0.03 | 1.9E-41 | Total, HDL, and LDL cholesterol (Teslovich et al., 2010) |
| | LPC 20:4 | rs174550 | | T/C | 0.33 | -0.45 | 0.03 | 9.7E-41 | Triglycerides (Teslovich et al., 2010) |
| | CE 20:4 | rs174547 | | T/C | 0.33 | -0.44 | 0.03 | 4.7E-37 | Lipid sub-species (Kettunen et al., 2012; Suhre et al., 2011a) |
| | PC 38:5 | rs174535 | | T/C | 0.33 | -0.41 | 0.03 | 1.4E-32 | Fasting glucose (Dupuis et al., 2010) |
| | LPC 20:5 | rs174548 | | C/G | 0.29 | -0.36 | 0.04 | 1.6E-23 | Crohn's disease (Franke et al., 2010) |
| | CE 20:5 | rs174548 | | C/G | 0.29 | -0.34 | 0.04 | 7.8E-21 | Resting heart rate (Eijgelsheim et al., 2010) |
| | TAG 58:10 | rs174548 | | C/G | 0.29 | -0.30 | 0.04 | 8.1E-17 | |
| | LPE 18:2 | rs4246215 | | G/T | 0.36 | 0.29 | 0.04 | 2.4E-16 | |
| | TAG 58:11 | rs174556 | | C/T | 0.29 | -0.30 | 0.04 | 2.5E-16 | |
| | LPE 20:4 | rs174548 | | C/G | 0.29 | -0.30 | 0.04 | 7.0E-16 | |
| | LPC 22:6 | rs174550 | | T/C | 0.33 | -0.26 | 0.04 | 1.3E-13 | |
| | PC 34:2 | rs174574 | | C/A | 0.33 | 0.26 | 0.04 | 1.3E-13 | |
| | PC 40:6 | rs174535 | | T/C | 0.33 | -0.26 | 0.04 | 3.5E-13 | |
| | PC 36:2 | rs174541 | | T/C | 0.36 | 0.25 | 0.04 | 2.6E-12 | |
| | PC 34:3 | rs174576 | | C/A | 0.33 | 0.23 | 0.04 | 1.7E-10 | |
| | CE 16:0 | rs174548 | | C/G | 0.29 | -0.23 | 0.04 | 1.8E-10 | |
| LPC 20:3 | rs968567 | C/T | 0.15 | 0.31 | 0.05 | 4.4E-10 | | | |
| CE 18:3 | rs174535 | T/C | 0.33 | -0.21 | 0.04 | 5.5E-09 | | | |
| TAG 54:4 | rs174550 | T/C | 0.33 | 0.19 | 0.04 | 4.8E-08 | | | |
| <i>CPS1</i> | Glycine | rs7422339 | 2 | C/A | 0.31 | 0.53 | 0.03 | 2.4E-58 | Creatinine production (Kottgen et al., 2010) |
| | Creatine | rs7422339 | | C/A | 0.31 | 0.24 | 0.04 | 2.5E-11 | Homocysteine (Lange et al., 2010) Fibrinogen (Danik et al., 2009) |
| <i>HPS1</i> | Asymmetric dimethylarginine | rs6584192 | 10 | T/G | 0.37 | -0.41 | 0.03 | 3.5E-33 | |
| <i>DMGDH</i> | Dimethylglycine | rs248386 | 5 | C/A | 0.19 | 0.50 | 0.04 | 6.6E-33 | |
| <i>SLC16A9</i> | Carnitine | rs1171617 | 10 | T/G | 0.23 | -0.42 | 0.04 | 5.9E-26 | Carnitine (Suhre et al., 2011a) Uric acid (Kolz et al., 2009) |
| <i>SERPINA7</i> | Thyroxine | rs7883218 | X | A/G | 0.12 | -0.39 | 0.04 | 1.3E-20 | |
| <i>PRODH</i> | Proline | rs2078743 | 22 | G/A | 0.09 | 0.49 | 0.06 | 2.2E-14 | Proline (Kettunen et al., 2012) |
| <i>GMPR</i> | Xanthosine | rs9477074 | 6 | T/C | 0.37 | 0.25 | 0.03 | 3.0E-14 | LDL cholesterol (Waterworth et al., 2010) |

| Locus | Trait | SNP | ch | Major/ minor Allele | MAF | Beta | s.e. | P-value | Previous genome-wide associations |
|-----------------------|---|--|----|---|--|--|--|---|--|
| <i>SLC6A13</i> | β -aminoisobutyric acid | rs11613331 | 12 | A/G | 0.49 | -0.23 | 0.03 | 2.0E-12 | Creatinine production (Kottgen et al., 2010) |
| <i>LIPC</i> | LPE 16:0 | rs1532085 | 15 | G/A | 0.38 | 0.24 | 0.03 | 3.7E-12 | Total and HDL cholesterol (Teslovich et al., 2010) Triglycerides (Teslovich et al., 2010) Lipid sub-species (Demirkan et al., 2012; Gieger et al., 2008; Kettunen et al., 2012) Metabolic syndrome (Kristiansson et al., 2012) Macular degeneration (Neale et al., 2010) |
| <i>GCKR</i> | Alanine Lactate α -hydroxybutyrate TAG 50:4 TAG 48:2 TAG 50:3 PC 34:3 PC 32:2 TAG 48:3 | rs1260326 rs1260326 rs1260326 rs1260326 rs1260326 rs1260326 rs1260326 rs1260326 | 2 | C/T C/T C/T C/T C/T C/T C/T C/T | 0.45 0.45 0.45 0.45 0.45 0.45 0.45 0.45 | 0.23 0.24 0.22 0.20 0.19 0.18 0.19 0.18 | 0.03 0.04 0.04 0.03 0.03 0.03 0.03 0.03 | 7.6E-12 3.3E-11 1.3E-09 3.4E-09 2.5E-08 2.6E-08 3.9E-08 4.9E-08 | Triglycerides (Teslovich et al., 2010) Total cholesterol (Teslovich et al., 2010) Phospholipids (Demirkan et al., 2012; Lemaire et al., 2011) CRP (Deleghan et al., 2011; Ridker et al., 2008) Chronic kidney disease (Kottgen et al., 2010) Glycemic traits (Dupuis et al., 2010; Saxena et al., 2010) Uric acid (Kobz et al., 2009; Yang et al., 2010) Alanine/glycine (Kettunen et al., 2012) Platelet count (Gieger et al., 2011) Albumin (Kamatani et al., 2010; Kim et al., 2011) Glucose/mannose (Suhre et al., 2011a) Crohn's disease (Franke et al., 2010) |
| <i>APOA1/C3/A4/A5</i> | DAG 36:2 TAG 56:3 TAG 56:10 TAG 54:2 TAG 52:4 TAG 54:3 DAG 34:2 TAG 52:5 DAG 36:1 TAG 52:3 TAG 54:5 | rs964184 rs964184 rs964184 rs964184 rs964184 rs964184 rs964184 rs964184 rs964184 rs12294259 rs964184 | 11 | C/G C/G C/G C/G C/G C/G C/G C/G C/G C/T C/G | 0.14 0.14 0.14 0.14 0.14 0.14 0.14 0.14 0.14 0.06 0.14 | 0.33 0.32 0.31 0.31 0.30 0.29 0.28 0.28 0.28 0.39 0.26 | 0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.07 0.05 | 1.3E-11 2.8E-11 1.5E-10 1.9E-10 2.4E-10 3.2E-09 6.0E-09 7.4E-09 8.2E-09 1.7E-08 4.8E-08 | Triglycerides (Teslovich et al., 2010) HDL cholesterol (Teslovich et al., 2010) Coronary artery disease (Schunkert et al., 2011) |
| <i>DDAH1</i> | NG-monomethyl-arginine | rs18582 | 1 | A/G | 0.35 | 0.24 | 0.04 | 2.2E-11 | |
| <i>SLCO1B1</i> | LPE 20:4 | rs4149056 | 12 | T/C | 0.15 | 0.30 | 0.05 | 1.1E-10 | Bilirubin (Johnson et al., 2009) Statins-induced myopathy (Link et al., 2008) Lipid sub-species (Suhre et al., 2011a) |
| <i>AGA</i> | Asparagine | rs4690522 | 4 | C/A | 0.32 | 0.25 | 0.04 | 1.4E-10 | |
| <i>SYNE2</i> | SM 14:0 | rs11158519 | 14 | G/A | 0.14 | 0.31 | 0.05 | 3.2E-10 | Lipid sub-species (Illig et al., 2010) Atrial fibrillation (Ellinor et al., 2012) Pulmonary function decline (Imboden et al., 2012) |
| <i>UMPS</i> | Orotic acid | rs9844948 | 3 | C/A | 0.16 | -0.32 | 0.05 | 7.6E-10 | |

| Locus | Trait | SNP | ch | Major/minor Allele | MAF | Beta | s.e. | P-value | Previous genome-wide associations |
|-----------------|--------------------------------|------------|----|--------------------|------|-------|------|---------|--|
| <i>PDE4D</i> | SM 24:1 | rs4700347 | 5 | A/G | 0.16 | 0.28 | 0.05 | 3.4E-09 | Stroke (Song et al., 2006) Esophageal cancer (Wu et al., 2011) Asthma (Himes et al., 2009) Sleepiness (Gottlieb et al., 2007) |
| <i>SEC61G</i> | CE 20:4 | rs11981543 | 7 | C/A | 0.11 | 0.32 | 0.05 | 4.2E-09 | |
| rs6593086 | TAG 58:9 | rs6593086 | 7 | C/G | 0.34 | -0.20 | 0.03 | 4.4E-09 | |
| | TAG 56:7 | rs6593086 | | C/G | 0.34 | -0.20 | 0.03 | 6.1E-09 | |
| | TAG 58:8 | rs6593086 | | C/G | 0.34 | -0.20 | 0.03 | 2.0E-08 | |
| <i>UGT1A5</i> | Xanthurenate | rs4148325 | 2 | C/T | 0.31 | 0.23 | 0.04 | 4.9E-09 | Bilirubin (Johnson et al., 2009; Sanna et al., 2009) |
| <i>SLC7A9</i> | NG-monomethyl-arginine | rs8101881 | 19 | T/C | 0.44 | -0.20 | 0.03 | 5.0E-09 | Chronic kidney disease (Kottgen et al., 2010) Creatinine level (Chambers et al., 2010) Urine lysine/valine (Suhre et al., 2011b) |
| <i>ABP1</i> | γ -aminoisobutyric acid | rs6977081 | 7 | G/T | 0.34 | 0.23 | 0.04 | 5.4E-09 | |
| <i>PHGDH</i> | Serine | rs477992 | 1 | G/A | 0.32 | -0.21 | 0.04 | 6.5E-09 | Serine (Suhre et al., 2011a) |
| <i>SLC16A10</i> | Tyrosine | rs411604 | 6 | G/A | 0.36 | 0.20 | 0.04 | 1.0E-08 | Isoleucine/Tyrosine (Suhre et al., 2011a) Alanine/Tyrosine (Kettunen et al., 2012) |
| <i>CSNK1G3</i> | Indoxylsulfate | rs875480 | 5 | A/G | 0.24 | 0.25 | 0.04 | 1.4E-08 | |
| <i>GNAL</i> | CE16:0 | rs1786573 | 18 | T/C | 0.27 | 0.21 | 0.04 | 2.8E-08 | |
| <i>DGKB</i> | Indole propionate | rs12699655 | 7 | T/G | 0.45 | -0.21 | 0.04 | 3.1E-08 | Glycemic traits (Dupuis et al., 2010) |
| <i>TBX18</i> | DAG 36:1 | rs4510639 | 6 | C/T | 0.25 | 0.21 | 0.04 | 3.5E-08 | |
| <i>NTANI</i> | CE 20:3 | rs3803573 | 16 | C/T | 0.28 | -0.21 | 0.04 | 4.7E-08 | Bone mineral density (Estrada et al., 2012) |

SNP, single nucleotide polymorphism; MAF, minor allele frequency; ch, chromosome; s.e., standard error; PC, phosphatidylcholine; LPC, lysophosphatidylcholine; CE, cholesterol ester; TAG, triacylglycerol; LPE, lysophosphatidylethanolamine; DAG, diacylglycerol; SM, sphingomyelin.