Dimethylguanidino valeric acid is a marker of liver fat and predicts diabetes

John F. O'Sullivan, … , Kathleen E. Corey, Robert E. Gerszten

Research Article   Cardiology   Metabolism

Unbiased, "nontargeted" metabolite profiling techniques hold considerable promise for biomarker and pathway discovery, in spite of the lack of successful applications to human disease. By integrating nontargeted metabolomics, genetics, and detailed human phenotyping, we identified dimethylguanidino valeric acid (DMGV) as an independent biomarker of CT-defined nonalcoholic fatty liver disease (NAFLD) in the offspring cohort of the Framingham Heart Study (FHS) participants. We verified the relationship between DMGV and early hepatic pathology. Specifically, plasma DMGV levels were correlated with biopsy-proven nonalcoholic steatohepatitis (NASH) in a hospital cohort of individuals undergoing gastric bypass surgery, and DMGV levels fell in parallel with improvements in post-procedure cardiometabolic parameters. Further, baseline DMGV levels independently predicted future diabetes up to 12 years before disease onset in 3 distinct human cohorts. Finally, we provide all metabolite peak data consisting of known and unidentified peaks, genetics, and key metabolic parameters as a publicly available resource for investigations in cardiometabolic diseases.

Find the latest version:

http://jci.me/95995/pdf

Dimethylguanidino valeric acid is a marker of liver fat and predicts diabetes

John F. O'Sullivan,[1,2] Jordan E. Morningstar,[3] Qiong Yang,[4,5] Baohui Zheng,[1] Yan Gao,[6] Sarah Jeanfavre,[7] Justin Scott,[7] Celine Fernandez,[8] Hui Zheng,[9] Sean O'Connor,[10] Paul Cohen,[10] Ramachandran S. Vasan,[4,11] Michelle T. Long,[12] James G. Wilson,[6] Olle Melander,[8,13] Thomas J. Wang,[14] Caroline Fox,[4] Randall T. Peterson,[1] Clary B. Clish,[7] Kathleen E. Corey,[15] and Robert E. Gerszten[3,7]

[1]Cardiovascular Research Center, Division of Cardiology, Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA. [2]Charles Perkins Centre and Heart Research Institute, The University of Sydney, Sydney, Australia. [3]Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA. [4]Framingham Heart Study of the National Heart, Lung, and Blood Institute and Boston University School of Medicine, Framingham, Massachusetts, USA. [5]Biostatistics Department, Boston University School of Public Health, Boston, Massachusetts, USA. [6]University of Mississippi Medical Center, Jackson, Mississippi, USA. [7]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [8]Department of Clinical Sciences, Lund University, Malmö, Sweden. [9]Biostatistics Department, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA. [10]The Rockefeller University, Laboratory of Molecular Metabolism, New York, New York, USA. [11]Cardiology Division, Boston Medical Center, and [12]Gastroenterology Division, Boston Medical Center, Boston University School of Medicine, Boston, Massachusetts, USA. [13]Center of Emergency Medicine, Skåne University Hospital, Malmö, Sweden. [14]Cardiology Division, Vanderbilt University, Nashville, Tennessee, USA. [15]Gastroenterology Division, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA.

Unbiased, "nontargeted" metabolite profiling techniques hold considerable promise for biomarker and pathway discovery, in spite of the lack of successful applications to human disease. By integrating nontargeted metabolomics, genetics, and detailed human phenotyping, we identified dimethylguanidino valeric acid (DMGV) as an independent biomarker of CT-defined nonalcoholic fatty liver disease (NAFLD) in the offspring cohort of the Framingham Heart Study (FHS) participants. We verified the relationship between DMGV and early hepatic pathology. Specifically, plasma DMGV levels were correlated with biopsy-proven nonalcoholic steatohepatitis (NASH) in a hospital cohort of individuals undergoing gastric bypass surgery, and DMGV levels fell in parallel with improvements in post-procedure cardiometabolic parameters. Further, baseline DMGV levels independently predicted future diabetes up to 12 years before disease onset in 3 distinct human cohorts. Finally, we provide all metabolite peak data consisting of known and unidentified peaks, genetics, and key metabolic parameters as a publicly available resource for investigations in cardiometabolic diseases.

Introduction

Nonalcoholic fatty liver disease (NAFLD) is defined by the presence of hepatic steatosis (via histology or imaging) when there are no causes for secondary hepatic fat accumulation, and it is histologically further classified into nonalcoholic fatty liver (NAFL) and nonalcoholic steatohepatitis (NASH) (1). NAFLD is frequently correlated with metabolic risk factors such as obesity, diabetes mellitus (DM), and dyslipidemia (1) and has recently been reported as an independent risk factor for atherosclerosis (2–4). However, the interplay between NAFLD, diabetes, and cardiovascular disease (CVD) remains poorly understood (5), and circulating markers of early disease might highlight at-risk populations.

We aimed to investigate the early systemic changes correlated with NAFLD and hypothesized that nontargeted metabolite profiling might provide an orthogonal view into hepatic pathology. As nontargeted metabolomic profiling covers a far greater portion of the metabolome (thousands of potential metabolites) than does targeted profiling (generally several hundred a priori–defined metabolites), it consequently offers greater potential for discovery. However, the metabolite identities of the underlying peaks of interest remain unknown until several challenging steps are completed.

High accurate-mass metabolite databases such as the Human Metabolome Database (HMDB) (6) and METLIN (7) and chemical databases such as ChemSpider (8) and PubChem (9) have been expanding in size. Searching these databases often provides a long list of potential matches for each unknown mass (reported as a mass/charge [m/z] ratio). However, unless the investigator is armed with additional information about the putative metabolite, refining this list can be difficult, with few success stories thus far (10). If commercial standards are not available, unambiguous identification can require labor-intensive, iterative synthesis of candidate molecules and subsequent comparative analysis using liquid chromatography coupled to accurate-mass tandem mass spectrometry (LC-MS/MS) (11).

One approach that has the potential to inform and refine this process is the incorporation of genetic data. Because metabolites play key roles as markers and effectors of cardiometabolic diseases, recent studies have sought to annotate the genetic determinants of circulating metabolite levels (12–19). For example, we

Table 1. Baseline characteristics of FHS, MDC, and JHS cohorts

Clinical characteristics	FHS	MDC		JHS	
		Cases	Controls	Cases	Controls
	(n = 1,066)	(n = 196)	(n = 126)	(n = 133)	(n = 465)
Age, yr, mean ± SD	40 ± 9	57 ± 6	58 ± 6	61 ± 10	62 ± 11
Female, n (%)	526 (53%)	106 (54%)	70 (56%)	86 (65%)	279 (60%)
BMI, kg/m^2, mean ± SD	26.6 ± 5.3	28.6 ± 5.2	27.9 ± 4.3	33.4 ± 6.8	30.2 ± 6.0
Waist, in., mean ± SD	36.3 ± 5.7	36.4 ± 6.1	35.4 ± 5.8	41.5 ± 5.4	39.0 ± 5.6
SBP, mmHg, mean ± SD	117 ± 14	147 ± 19	142 ± 16	132 ± 20	129 ± 16
DBP, mmHg, mean ± SD	75 ± 9	90 ± 10	88 ± 9	77 ± 9	75 ± 9
HOMA-IR, mean ± SD	1.05 ± 0.84	2.8 ± 1.7	2.2 ± 1.6	4.7 ± 2.4	3.3 ± 1.9
Glucose, mg/dl, mean ± SD	96 ± 19	99 ± 8	95 ± 7.5	98 ± 10	91 ± 8
HDL, mg/dl, mean ± SD	60 ± 18	48 ± 13	50 ± 16	51 ± 14	54 ± 16
Triglycerides, mg/dl, mean ± SD	109 ± 66	148 ± 67	135 ±63	118 ± 57	100 ± 53
Creatinine, mg/dl, mean ± SD	0.80 ± 0.15	0.99 ± 0.28	0.95 ± 0.15	1.0 ± 0.6	1.0 ± 0.2

reported a GWAS of 217 plasma metabolite traits measured in 2,076 participants of the Framingham Heart Study (FHS). Further, and as corroborated by other published metabolite GWAS, many of the strongest associations we identified were at loci that encode transporters or enzymes with a direct biochemical relationship with the metabolite (20, 21). Therefore, we hypothesized that in the context of nontargeted profiling, a genetic association with an unknown metabolite could aid in its identification by highlighting a specific enzymatic reaction. Here, we first identified metabolite peaks correlated with hepatic fat as assessed by CT in a community-based population free of cardiometabolic disease. In GWAS, our top metabolite peak mapped to a specific aminotransferase, providing the key insight required for its unambiguous identification. We then validated the putative biomarker in a biopsy-proven cohort of NAFLD and determined its predictive capability for future cardiometabolic disease in other cohorts. In addition, all data (for known metabolites and unidentified peaks) are being made publicly available through the database of Genotypes and Phenotypes (dbGaP; accession number: pht005145.v1.p10) (http://www.ncbi.nlm.nih.gov/gap; representative summary data for metabolite peaks correlated with hepatic fat as well as GWAS data for the top finding are also included in this article). This resource enables researchers to cross-reference their own metabolites of interest in a richly phenotyped cohort and use the available genetic data to inform the identification of unknown peaks — the most challenging step in nontargeted metabolomics.

Results

m/z 202.1185 is correlated with CT-defined liver fat. We applied nontargeted LC-MS discovery (see Methods) to a cohort of 1,066 FHS Generation III participants (Table 1). Of these, 470 participants underwent a CT scan as part of a comprehensive assessment of adipose depots, including liver fat, which was quantified using the liver-to-phantom ratio (LPR) (22). The top-20 metabolite peaks that correlated with hepatic fat (age- and sex-adjusted) are shown in Table 2. The list includes glutamate, which has been previously correlated with cross-sectional and incident cardiometabolic disease (23). A comprehensive list of all peaks

correlated with liver fat and many other phenotypes, as well as GWAS data, is being made publicly available via dbGaP (http://www.ncbi.nlm.nih.gov/gap). In age- and sex-adjusted analyses, the most highly significant correlation was between CT-defined liver fat and an unknown metabolite with a mass-to-charge (m/z) ratio of 202.1185 (Table 2) (P value = 2.28×10^{-24}). This correlation was highly significant after adjustment for multiple covariates including age, sex, smoking history, alcohol consumption, HDL concentration, triglyceride concentration, homeostasis model assessment of insulin resistance (HOMA-IR), hypertension, glucose, and BMI (β coefficient = 0.174, P value = 1.16×10^{-10}) and remained so after further adjustment for liver function tests (alanine transaminase [ALT] and aspartate transaminase [AST]; P value = 4.12×10^{-8}). Receiver-operator characteristic (ROC) curves showed that this peak also improved discrimination for liver fat as compared with AST and ALT (P = 0.014; Supplemental Figure 4; supplemental material available online with this article; https://doi.org/10.1172/JCI95995DS1). Of note, there were no significant associations between SNPs in the *PNPLA3* locus and levels of this unknown metabolite (representative SNP rs738409; P = NS).

GWAS informs the identification of m/z 202.1185. We searched the Human Metabolome Database (6) for a list of possible candidate matches for m/z 202.1185 (Supplemental Figure 1A). However, comparison of MS/MS fragmentation of multiple authentic

Table 2. Top-20 metabolite peaks correlated with liver fat (age- and sex-adjusted)

	Metabolite	m/z	RT	β	SE	P value
1	5836	202.1185	7.79	0.0197	0.00193	2.28×10^{-24}
2	474	551.5034	1.61	0.0204	0.00211	4.08×10^{-22}
3	1648	386.2536	1.99	0.0212	0.00243	3.41×10^{-18}
4	606	606.6179	1.66	0.0195	0.00226	7.11×10^{-18}
5	523	612.5556	1.63	0.0192	0.00229	6.69×10^{-17}
6	578	634.6491	1.65	0.0191	0.00229	7.59×10^{-17}
7	613	578.5864	1.66	0.019	0.00228	8.28×10^{-17}
8	5893	116.1073	7.87	0.0187	0.00231	5.80×10^{-16}
9	5830	223.972	7.77	0.0193	0.00242	1.26×10^{-15}
10	5826	261.9278	7.77	0.0195	0.00246	2.61×10^{-15}
11	Glutamate	148.0603	7.8	0.0171	0.00217	3.46×10^{-15}
12	4388	795.5713	5.54	0.0179	0.00228	4.00×10^{-15}
13	468	577.5186	1.61	0.0172	0.0022	6.35×10^{-15}
14	537	313.2733	1.63	0.017	0.00218	7.38×10^{-15}
15	510	341.3046	1.62	0.0164	0.00213	1.72×10^{-14}
16	4507	575.5026	5.65	0.0191	0.00251	2.92×10^{-14}
17	418	549.4876	1.57	0.0167	0.00222	5.82×10^{-14}
18	4477	747.5711	5.62	0.017	0.00227	5.86×10^{-14}
19	612	632.6336	1.66	0.0171	0.00228	5.88×10^{-14}
20	4478	603.534	5.62	0.0167	0.00222	6.25×10^{-14}

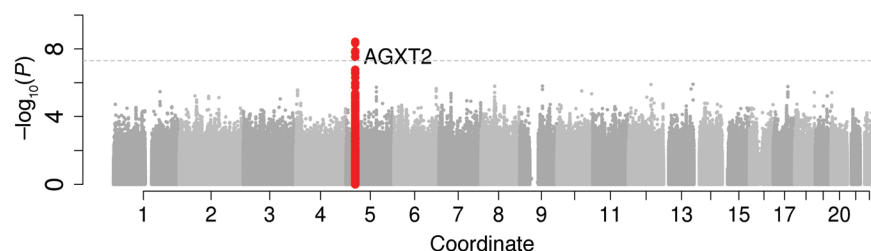β, effect size for 1-unit increase in metabolite levels on liver fat (negative LPR); SE, standard error.

Figure 1. *m/z* 202.1185 is associated with the gene *AGXT2*. Manhattan plot of the GWAS for *m/z* 202.1185 revealed an association with several SNPs at the *AGXT2* locus. See also Supplemental Figure 2.

chemical standards for potential matches with the MS/MS fragmentation of *m/z* 202.1185 failed to make a definitive identification (representative illustration in Supplemental Figure 1B).

To further inform the *m/z* data, we performed a GWAS on this unknown metabolite peak by imputation of 15.6 million SNPs (see Methods). We found that *m/z* 202.1185 was significantly ($P = 3.79 \times 10^{-9}$) associated with multiple SNPs near the gene encoding for alanine glyoxylate aminotransferase 2 (AGXT2) (Figure 1 and Supplemental Figure 2). AGXT2 was originally described in the context of transamination of alanine and glyoxylate, resulting in the formation of pyruvate and glycine (Supplemental Figure 3A) (24). However, subsequent studies have shown that this enzyme can catalyze the transamination of other previously unanticipated metabolite pairs, including β-aminoisobutyric acid (BAIBA) and pyruvate, asymmetric dimethylarginine (ADMA) and glyoxylate, as well as alanine and Y,δ-dioxovaleric acid (DOVA) (25). Therefore, we sought to determine whether *m/z* 202.1185 was either a product or substrate of AGXT2-mediated transamination.

Metabolite databases can also be searched using the fragment ions instead of the parent mass of a compound. We searched the METLIN database (7) using the 2 dominant fragments of 202.1185: 70.0659 and 71.0612, with a mass error tolerance of 5 or fewer parts per million (ppm) (Supplemental Figure 1C). ADMA was suggested as the potential candidate metabolite. While the chromatographic retention time (RT) as well as mass ruled out ADMA as the actual parent mass, we determined that the product of a transamination reaction (causing the substitution of an amine with a carbonyl group) (Figure 2A) of ADMA by AGXT2 would have an *m/z* within 1 ppm of our unknown metabolite, which is well within the acceptable mass error using this technology (26). Although this compound is not present in metabolite databases (HMDB [ref. 6], METLIN [ref. 7]) or chemical (ChemSpider [ref. 8], PubChem [ref. 9]), it has been described in the literature before and labeled α-keto-δ-(N^{G}N^G-dimethylguanidino)-valeric acid (DMGV) (27, 28).

Confirmation of DMGV as the correct identity of m/z 202.1185. Given the lack of an authentic standard, DMGV was synthesized by reflux of ADMA with trifluoroacetic acid anhydride and then hydrolysis using sodium hydroxide, followed by purification with preparative chromatography. We spiked DMGV into human plasma, and reacquired data using the LC-MS platform alongside a human plasma sample. We performed MS/MS fragmentation of both the standard and the unknown. As demonstrated in Figure 2, B and C, the chromatographic RT and the MS/MS fragmentation spectra match exactly.
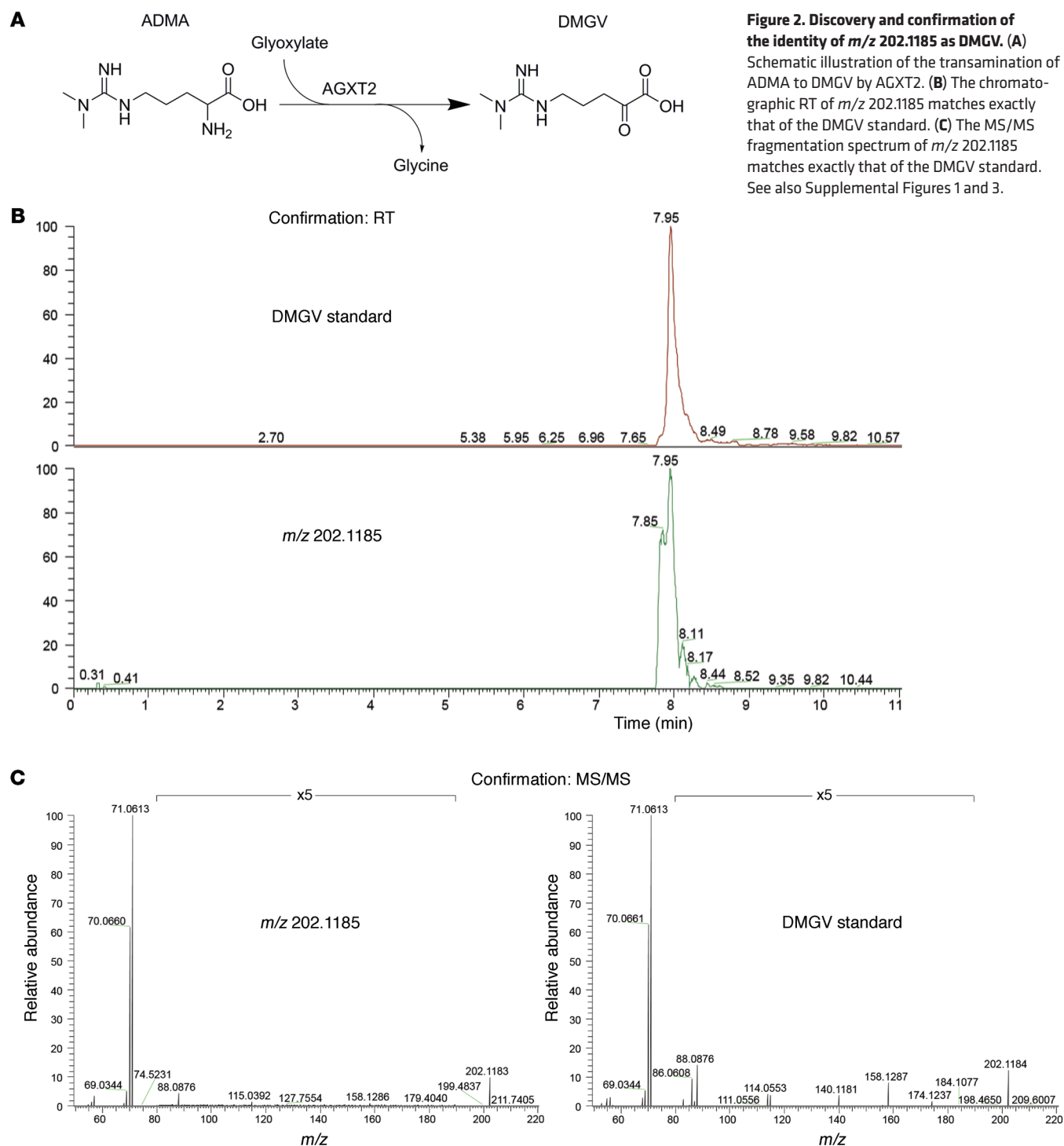
DMGV is correlated with biopsy-proven NASH. To better evaluate the relationship between DMGV levels and hepatic pathology, we profiled plasma samples from a cohort of patients with biopsy-proven NASH and biopsy-proven normal histology matched for age, sex, and BMI (controls; Table 3). DMGV levels were significantly elevated in patients with NASH as compared with levels in matched controls ($n = 36$ per group, 2.48 ± 1.31 vs. 1.71 ± 0.99 AU, $P = 0.007$) (Figure 3A). We performed multiple logistic regression of the correlation of DMGV with biopsy-proven NASH, adjusting for clinical and biochemical factors. DMGV remained significantly correlated with biopsy-proven NASH in the presence of the following covariates: age and sex (odds ratio [OR] = 1.97, $P = 0.009$); age, sex, and BMI (OR = 1.96, $P = 0.011$); and age, sex, and ALT, a marker of liver injury (OR = 1.98; $P = 0.016$).

We also performed correlation analysis of DMGV levels with steatosis, hepatocyte ballooning, lobular inflammation, and NAFLD activity score (NAS) (which is a composite of all 3 parameters) in the age-, sex-, and BMI-matched patients. Of these parameters, DMGV was most significantly correlated with hepatocyte ballooning (Table 4). Taken together, these data extend the imaging findings by further tying levels of this small molecule to hepatic pathology.

DMGV predicts future development of type 2 diabetes. We next examined whether baseline DMGV levels in disease-free individuals could predict future type 2 diabetes (T2D). In the FHS Generation 3 cohort, in which the correlation between DMGV and hepatic fat was identified, baseline DMGV levels were correlated with incident T2D in the 19 subjects who developed overt disease in the ensuing 6 years of follow-up (OR = 1.86, $P = 0.0005$). To extend this observation made with a small number of incident cases, we measured DMGV levels in 196 individuals with incident T2D and 126 controls in the Malmö Diet and Cancer (MDC) study, with a mean follow-up of 12.8 years (Table 1). DMGV was a strong, independent predictor of T2D in the replication cohort (OR = 1.6 per SD increment, $P = 8.6 \times 10^{-4}$, $n = 322$, adjusting for age, sex, glucose, and BMI; Table 5). We then investigated whether DMGV levels predicted incident DM in a third cohort, in this case, African American participants in the Jackson Heart Study (JHS). We measured DMGV in 133 cases of incident DM and 465 controls. In this group as well, we found that DMGV was an independent predictor of T2D, even after adjusting for age, sex, BMI, and baseline fasting glucose (mean follow-up 7.5 years, OR = 1.3 per SD increment, $P = 0.03$, $n = 598$) (Table 5). Of note, in both cohorts, DMGV was predictive of incident disease, even after further adjustment for levels of branched-chain amino acids (BCAAs), emerging biomarkers of prediabetes (29, 30) (MDS: OR = 1.6 per SD increment; $P = 9.2 \times 10^{-4}$; $n = 322$; JHS: OR = 1.3 per SD increment; $P = 0.03$; $n = 598$).

DMGV levels are significantly decreased following weight loss surgery. We assessed whether DMGV levels track with weight loss following bypass surgery, which has been shown to improve NAFLD (31). We performed metabolomic profiling on our tar-

Figure 2. Discovery and confirmation of the identity of *m/z* 202.1185 as DMGV. (A) Schematic illustration of the transamination of ADMA to DMGV by AGXT2. (B) The chromatographic RT of *m/z* 202.1185 matches exactly that of the DMGV standard. (C) The MS/MS fragmentation spectrum of *m/z* 202.1185 matches exactly that of the DMGV standard. See also Supplemental Figures 1 and 3.

geted platform using plasma samples from a clinical cohort of 39 roux-en-Y gastric bypass (RYGB) patients. All patients had NAFLD defined by biopsy at the time of operation and had blood samples taken at this time. Patients returned to the clinic 2 and 6 months after their operation and had clinical parameters measured and plasma taken for clinical biochemistry (Table 6). At both 2 and 6 months after RYGB, patients had dramatic weight loss and also significant improvements in metabolic parameters, such as glucose levels, systolic blood pressure

(SBP), diastolic blood pressure (DBP), total cholesterol, LDL cholesterol, and triglycerides. Compared with baseline, DMGV levels were significantly decreased at both the 2-month (2.23 ± 1.77 vs. 1.67 ± 1.18 AU, $P = 0.01$, $n = 39$) and 6-month (2.23 ± 1.77 vs. 1.33 ± 1.17 AU, $P = 0.0003$, $n = 39$) time points (Figure 3B). Of note, there was no significant correlation between baseline levels of DMGV and levels of ALT ($\rho = 0.08$, $P = 0.58$), a marker of liver injury. These data suggest that DMGV may report on a different pathway related to liver disease.

Table 3. Clinical and biochemical characteristics of biopsy-proven NASH cohort

Clinical characteristics	Normal (*n* = 36)	NASH (*n* = 36)	*P* value
Age at biopsy, yr, mean ± SD	40 ± 9.7	43.1 ± 11.8	0.23
BMI, kg/m^2, mean ± SD	46.5 ± 7.0	47.9 ± 7.4	0.44
AST, IU/l, median (range)	16.5 (0–61.0)	24.5 (13–179)	<0.001
ALT, IU/l, median (range)	27.0 (14.0–147.00)	48 (19–288)	<0.001
LDL, mg/dl, median (range)	89.8 (29.4–178.6)	91.6 (48.2–179.4)	0.92
Total cholesterol, mg/dl, median (range)	163 (80–160)	160 (101–276)	0.99
Trig, mg/dl, median (range)	91 (44–266)	125 (39–418)	0.03
HDL, mg/dl, median (range)	45 (25–105)	38.5 (25–73)	0.007
Insulin, mU/l, median (range)	20 (4–118)	32.5 (11–99)	0.002
Glucose, median (range)	91 (65–224)	119.5 (74–263)	0.01
HOMA-IR, median (range)	4.8 (0.9–65.3)	8.6 (2.4–40.1)	0.001

Discussion

There is an important need for early biomarkers of cardiometabolic disease. While investigators have begun to apply unbiased metabolomics techniques to human samples, replicated findings of disease-predictive markers are few. To address this gap in knowledge, we used nontargeted metabolomic profiling, an unbiased, hypothesis-generating approach aimed to optimize the opportunity for novel insight. Once we determined the unknown metabolite peak of the strongest correlation with NAFLD, we incorporated genomic data to help identify the metabolite and to help place the metabolite in a pathway. Using this integrative approach, we identified what we believe to be a completely novel correlation of a poorly understood metabolite — DMGV — with liver fat. We first confirmed the correlation of DMGV with bona fide early liver disease as assessed by biopsy and then demonstrated that baseline levels of DMGV predict incident diabetes as early as 12 years before disease onset. Further, DMGV levels fall significantly (*P* = 0.0003) with weight loss (following gastric bypass surgery), which has been shown to improve NAFLD.

DMGV has been measured in the blood of rats and humans in 2 prior studies (27, 28), but it has never been correlated with metabolic disease and had not been added to biochemical databases. In our analyses, DMGV was an independent predictor of future T2D in the FHS, MDC, and JHS cohorts. All individuals in these cohorts were free of diabetes at the time of blood sampling, minimizing interference from concomitant lifestyle or medical interventions. Furthermore, DMGV remained predictive of future diabetes in the MDC and JHS cohorts after correcting for age, sex, BMI, and fasting glucose levels.

As weight loss is a cornerstone of NAFLD therapy and weight loss surgery is correlated with high rates of NASH regression (31), it was important to determine whether DMGV was modulated by this intervention. In the RYGB cohort, patients lost significant (*P* < 0.0001) weight at both 2 months and 6 months after their operation. DMGV was significantly reduced at both 2 months (*P* = 0.01) and 6 months (*P* = 0.0003) after the operation, and it is notable that DMGV levels did not correlate with ALT, a known marker of liver pathology.

AGXT2 has been described as a promiscuous aminotransferase, catalyzing the transamination of several substrates (Supplemental Figure 3), including BAIBA, pyruvate, glyoxyl-

ate, ADMA, and alanine (25). Our group and others have previously reported associations between common variants in the *AGXT2* gene with circulating metabolite levels in human blood (17, 20, 32) and urine (17). Our data therefore extend the findings of previous studies by demonstrating that DMGV is a bona fide product of AGXT2 metabolism in humans. We have previously shown that knockdown of *agxt2* in zebrafish using morpholinos modulates cholesterol and triacylglycerol metabolism (20). Quantitative trait locus analyses performed in inbred mouse strains have also suggested a potential role for *Agxt2* and related pathways in liver metabolism (33), while SNPs in the *AGXT2* locus have nominal association with diabetes. BAIBA, a substrate of AGXT2, is inversely correlated with cardiometabolic traits in humans and causes browning of white fat and hepatic β oxidation in cells and mice (32). Preliminary experimental studies in our laboratory demonstrated increased expression of AGXT2 in the livers of mice fed a high-fat diet (HFD), with a commensurate increase in circulating DMGV levels (Supplemental Figure 5). Our working model is that, during NAFLD devel-
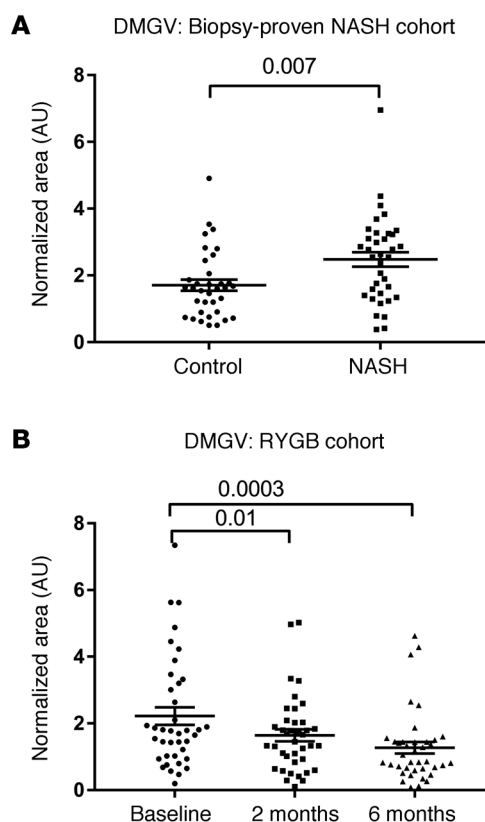
Figure 3. DMGV levels are elevated in biopsy-proven NASH and are modulated following weight loss surgery. (**A**) DMGV was significantly elevated in biopsy-proven NAFLD cases compared with controls (Mann-Whitney *U* test, *n* = 36 per group, 2.48 ± 1.31 vs. 1.71 ± 0.99 AU, *P* = 0.007). (**B**) Levels of DMGV at each time point before and after RYGB, demonstrating a significant reduction at 2 months and 6 months (paired *t* tests, *n* = 39, *P* = 0.01 and *P* = 0.0003, respectively). See also Supplemental Figures 1, 3, and 4.

Table 4. Correlation of DMGV with NAFLD histological parameters (*n* = 72)

	NAS	Steatosis	Ballooning	Lobular inflammation
Spearman's correlation coefficient (*r*)	0.30284	0.31114	0.33012	0.23351
P value	0.0103	0.0083	0.0049	0.05

opment, decreased DDAH1 enzymatic activity (34–36) and decreased BAIBA levels in cardiometabolic disease (32) lead to an increased concentration of ADMA that binds AGXT2 more freely (due to decreased levels of BAIBA, the major substrate of AGXT2) and thereby leads to increased conversion to DMGV by AGXT2 (see Supplemental Figure 6). Furthermore, our murine liver AGXT2 expression studies suggest that in NALFD, AGXT2 expression is increased and that, by implication, there is greater enzymatic activity causing conversion of ADMA to DMGV. Future studies must determine whether DMGV itself confers functional effects in model systems.

Others have also incorporated genetics to streamline the identification of unknown metabolites (37–39). A recent multistep approach (37) used a GWAS of known and unknown metabolites and correlation analysis of known metabolites with unknown ones, and then used published metabolic pathway information to guide the construction of networks that displayed these relationships as a function of metabolite class. The authors used this network information to refine a list of putative metabolite database matches for metabolites of interest, upon which they performed iterative comparison of unknowns with chemical standards, allowing identification of several of these metabolites. One limitation of this approach is that metabolite databases, while constantly increasing in size, remain incomplete, and thus metabolites such as DMGV that are not in these databases cannot be identified using this approach. Furthermore, this group leveraged previously reported disease correlations (37), whereas our disease correlations were derived directly from our own study populations. Finally, we prioritized our unknown metabolites by correlation with disease (DMGV was the top unknown metabolite correlated with liver fat), focusing our workflow on disease biology from the outset.

Our study has potential limitations. Although the Malmö and Jackson cohorts allowed us to examine incident DM, none of the cohorts had phenotyping sufficient to study incident cases of NAFLD, so we were only able to perform cross-sectional analysis of NAFLD cases (defined by CT or biopsy). Second, in our larger cohort that used CT-defined liver fat, we were unable to stratify DMGV levels by severity of NAFLD, as CT cannot differentiate between NAFL and NASH.

Studies are underway to unambiguously identify other metabolites correlated with hepatic fat (Table 2) that might ultimately prove to be better disease markers or functionally relevant small molecules. More broadly, these studies suggest that unbiased metabolite profiling, informed by genetic information, holds promise for the identification of biomarkers and pathways in cardiometabolic disease. Importantly, all the peak data, along with GWAS and phenotypic correlations, are being made publicly available online via dbGaP (accession number: pht005145. v1.p10) as a resource for the scientific community (note: representative summary data for metabolite peaks correlated with hepatic fat as well as GWAS data for the top finding are all included in this article). We anticipate that this resource will serve as a powerful reference tool investigators can use to look up correlations of metabolites (or peaks) identified in their own studies in order to assess their correlations with key clinical parameters in humans and to integrate these correlations with genetic information to aid unambiguous identification and pathway elucidation.

Methods

Subjects

FHS. The FHS Generation 3 cohort enrolled 4,095 individuals from 2002 to 2005 in a community-based longitudinal cohort study. Of 1,006 randomly selected participants, 470 underwent both CT scans of their liver and metabolomic profiling. Cardiometabolic traits and routine biochemical test results were available for all these subjects. Genotyping was performed on the Affymetrix GeneChip Human Mapping 500K Array SetR and 50K Human Gene Focused PanelR, with parameters and data analysis as previously described (20). Briefly, we performed a GWAS on the metabolite peak for *m/z* 202.1185 by imputation of 15.6 million SNPs (1000 Genomes Phase I, version 3, CEU population, March 2012 release, build 37) using a hidden Markov model that was implemented in MACH/minimac (2012-5-29 release) (https://genome.sph.umich.edu/wiki/Minimac).

Hospital-based cohort. The hospital-based NASH case-control cohort included patients and controls enrolled at Massachusetts General Hospital. Patients undergoing clinically indicated weight loss surgery had a standard-of-care liver biopsy performed at the time of surgery. Patients with other chronic liver conditions or alcohol use were excluded. Each biopsy was evaluated by a blinded, board-certified hepatopathologist according to the criteria established by Kleiner et al. (40). NASH was defined by the presence of at least grade 1 steatosis, lobular inflammation, and hepatocyte ballooning. Normal liver histology had no evidence of steatosis, portal or lobular inflammation, hepatocyte ballooning, or fibrosis.

Table 5. DMGV prediction of T2D in MDC and JHS cohorts

Model	Model 1		Model 2	
	Adjusted for age, sex		Adjusted for age, sex, BMI, glucose	
	MDC	JHS	MDC	JHS
OR Q1	1.0 (referent)	1.0 (referent)	1.0 (referent)	1.0 (referent)
OR Q2	1.09 (0.58–2.04)	1.15 (0.6–2.1)	1.15 (0.61–2.19)	1.30 (0.8–2.2)
OR Q3	1.49 (0.79–2.81)	1.87 (1.09–3.2)	1.59 (0.81–3.13)	1.51 (0.9–2.5)
OR Q4	2.71 (1.37–5.38)	2.6 (1.54–4.39)	2.81 (1.34–5.87)	1.79 (1.0–3.04)
P for trend	0.003	0.00002	0.004	0.026

Q1, Quartile 1; Q2, Quartile 2; Q3, Quartile 3; Q4, Quartile 4.

Table 6. Clinical and biochemical characteristics of RYGB cohort (*n* = 39)

Clinical characteristics	Baseline	2 mo	6 mo	*P* value
Sex, % male	25.6			
Age, yr, mean ± SD	46.41 ± 12.54			
Weight, lb, median (range)	283.3 (188.0–483.0)	240.1 (168.0–398.0)	202 (134.0–352.0)	<0.0001
BMI, kg/m^2, median (range)	45 (35.0–65.0)	38.7 (31.7–57.0)	32.5 (24.5–50.0)	<0.0001
Creatinine, mg/dl, median (range)	0.9 (0.6–1.5)	0.8 (0.7–1.4)	0.8 (0.6–1.8)	0.4
SBP, mmHg, median (range)	122.0 (106.0–170.0)	117.0 (0.0–146.0)	120.0 (88.0–150.0)	0.04
DBP, mmHg, median (range)	80 (64–98)	76 (60–98)	72 (48–100)	0.01
Glucose, mg/dl, median (range)	108 (68–252)	NA	89 (71–136)	0.001
HOMA-IR, mass units, median (range)	55.7 (29.2–369.6)	NA	19.3 (3.3–149.3)	<0.0001
Total cholesterol, mg/dl, median (range)	188 (103–267)	NA	147 (102–222)	< 0.0001
LDL, mg/dl, median (range)	96 (39–179)	NA	80.5 (35–141)	0.02
HDL, mg/dl, median (range)	49 (33–84)	NA	51 (33–85)	0.8
Triglycerides mg/dl, median (range)	133 (54–375)	NA	87 (48–179)	0.0001
ALT IU/l, mean ± SD	42.3 ± 38.3	36.3 ± 21.6	23.6 12.5	0.006
AST IU/l, mean ± SD	31.1 ± 19.6	29.6 ± 9.8	24.1 ± 6.8	0.03

MDC study. We determined the predictive ability of DMGV in the MDC study, a Swedish population–based cohort of 28,449 individuals enrolled between 1991 and 1996. From this cohort, 6,103 individuals were randomly selected to participate in the MDC Cardiovascular Cohort (41). Fasting plasma samples were obtained from 5,305 subjects in the MDC Cardiovascular Cohort, 564 of whom had prevalent diabetes or CVD prior to baseline measurements. Of note, 456 subjects had missing covariate data, leaving 4,285 subjects eligible for analysis. For these subjects, using a clinical endpoint of new-onset diabetes until December 2012, there were 126 controls and 196 cases of incident diabetes. The mean ± SD follow-up time was 12.8 ± 5.8 years. Clinical and biochemical characteristics of this cohort are described in Table 1, and detailed descriptions of the clinical assessment, diabetes definition, and subject selection have been previously described (30).

JHS. The JHS is a prospective, community-based cohort study designed to investigate determinants of CVD among African Americans living in the tricounty area (Hinds, Madison, and Rankin counties) of metropolitan Jackson, Mississippi. Metabolite profiles were available. Of the 5,306 JHS participants who completed the baseline clinic visit (2000–2004), 3,406 individuals were included in the NIH Trans-Omics for Precision Medicine (TOPMed) project on the basis of consent that allows genetic analysis and data sharing through the NCBI's dbGaP database. From this cohort, DMGV levels were available for 1,052 participants, including 200 coronary heart disease (CHD) cases and 200 controls (matched by sex, age within 2 years, and BMI within 3 units); 183 chronic kidney disease (CKD) cases and 202 controls (matched by sex, age within 2 years, and BMI within 3 units); and 267 randomly selected samples. Of the 1,052 subjects with DMGV data, 327 had prevalent diabetes, 106 had missing information regarding diabetes status over the follow-up period, and 21 subjects had missing covariate data. For the remaining subjects, using a clinical endpoint of new-onset diabetes until 2013, there were 465 controls and 133 cases of incident diabetes. The mean ± SD follow-up time was 7.5 ± 1.4 years. The clinical and biochemical characteristics of this cohort are described in Table 1.

Plasma samples

EDTA blood samples were collected and immediately centrifuged to separate cellular material from plasma. Aliquots of plasma were frozen on dry ice and stored at –80°C until analysis. Samples (10-μl) were deproteinized with 90 μl acetonitrile/methanol/formic acid (75:25:0.2; v/v/v) (all from MilliporeSigma) containing the deuterated internal standards 25 μM phenylalanine-d$_8$ (Cambridge Isotope Laboratories) and 10 μM valine-d$_8$ (MilliporeSigma). After vortexing, the samples were centrifuged at 20,000 *g* at 4°C for 15 minutes, and the supernatants were transferred to HPLC-quality glass vials with inserts (MicroSolv).

Metabolite profiling

Hybrid platform. Hydrophilic interaction liquid chromatographic (HILIC) analyses of water-soluble metabolites in the positive ionization mode were conducted using an LC-MS system composed of a Shimadzu Nexera X2 U-HPLC (Shimadzu Corp.) coupled to a Q Exactive Hybrid Quadrupole Orbitrap Mass Spectrometer (Thermo Fisher Scientific). Plasma samples (10-μl) were prepared via protein precipitation, with the addition of 9 volumes of acetonitrile/methanol/formic acid (74.9:24.9:0.2; v/v/v) containing stable isotope–labeled internal standards (valine-d8; MilliporeSigma) and phenylalanine-d8 (Cambridge Isotope Laboratories). The samples were centrifuged (10 min, 9,000 *g*, 4°C), and the supernatants were injected directly onto a 150 × 2 mm, 3-μm Atlantis HILIC column (Waters). The column was eluted isocratically at a flow rate of 250 μl/min with 5% mobile phase A (10 mM ammonium formate and 0.1% formic acid in water) for 0.5 minutes, followed by a linear gradient to 40% mobile phase B (acetonitrile with 0.1% formic acid) for 10 minutes. MS analyses were carried out by electrospray ionization in the positive ion mode using full-scan analysis over 70 to 800 *m/z* at 70,000 resolution and a 3-Hz data acquisition rate. Other MS settings were: sheath gas 40, sweep gas 2, spray voltage 3.5 kV, capillary temperature 350°C, S-lens RF 40, heater temperature 300°C, microscans 1, automatic gain control target 1 × 10^6, and maximum ion time 250 ms. Metabolite identities were confirmed using authen-

tic reference standards. Raw data were processed using TraceFinder software (Thermo Fisher Scientific) and Progenesis QI (Nonlinear Dynamics). Product ion mass spectra (MS/MS) were collected using a Q Exactive Focus Hybrid Quadrupole Orbitrap Mass Spectrometer (Thermo Fisher Scientific). MS/MS spectra were collected on mass 202.1182 *m/z*. The collision energies were 10, 20, and 40; the isolation window was 1.4 *m/z*; and the resolution was 35,000.

Synthesis and purification of DMGV

Synthesis of DMGV was done using the protocol described by Klein et al. (42). Briefly, ADMA (MilliporeSigma) was refluxed under anhydrous conditions with trifluoroacetic acid anhydride (MilliporeSigma) to make an ADMA-trifluoroacetamide intermediate, and then hydrolysis was performed with sodium hydroxide to yield the crude DMGV mixture. Purification was done by flash column chromatography using a Sephadex G-10 Gel (MilliporeSigma). Fractions that formed a positive reaction with 2,4-dinitrophenylhydrazine were combined and dried. Purity was assessed by LC-MS.

DMGV assay on the targeted MS platform

Deproteinized plasma extracts were subjected to normal-phase HILIC using a 150 × 2.1 mm Atlantis HILIC column (Waters) and the following mobile phases: mobile phase A: 10 mM ammonium formate and 0.1% formic acid (v/v); and mobile phase B: acetonitrile with 0.1% formic acid (v/v) (both from MilliporeSigma). The samples were injected directly onto the HILIC column that was eluted at a flow rate of 250 μl/min with initial conditions of 5% mobile phase A and 95% mobile phase B, followed by a 10.5-minute linear gradient to 60% mobile phase A. The injection volume was 10 μl. The multiplexed LC system was composed of a 1200 Series Pump (Agilent Technologies) and an HTS PAL Autosampler (Leap Technologies) equipped with 2 injection ports and a column selection valve.

The LC system was connected to a 4000 QTrap Triple Quadropole Mass Spectrometer (Applied Biosystems/Sciex) run in positive ion mode. Multiple reaction monitoring (MRM) transitions and chromatographic RTs were calculated for DMGV using infusion of authentic chemical standards. MS analyses were carried out using electrospray ionization (ESI) and MRM scans in positive ion mode. Declustering potentials and collision energies were optimized for DMGV by infusion of the reference standard prior to sample analyses. The dwell time for each transition was 30 ms, the ion spray voltage was 5 kV, and the source temperature was 450°C.

Metabolite peaks were integrated using Sciex MultiQuant software. All metabolite peaks were manually reviewed for peak quality in a blinded manner. In addition, pooled cellular extract samples were interspersed within each analytical run at standardized intervals every 10 injections, enabling the monitoring and correction for temporal drift in MS performance. The nearest neighbor flanking pair of pooled plasma was used to normalize samples in a metabolite-by-metabolite manner. Internal standard peak areas were monitored for quality control, and individual samples with peak areas differing from the group mean by more than 2 SDs were reanalyzed.

Agxt2 mRNA expression in mice

Male C57BL/6J mice were fed a 60% HFD diet to induce NAFLD beginning at 8 weeks of age. At 14 weeks of age, the mice were sacrificed and their livers removed and flash-frozen. RNA was extracted with TRIzol-chloroform followed by column purification (RNeasy; QIAGEN). Reverse transcription was performed using the High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific). Quantitative PCR (qPCR) of samples in duplicate were run on a QuantStudio 6 Flex machine (Applied Biosystems) using Power SYBR Green PCR Master Mix (Applied Biosystems). *Agxt2* mRNA levels were normalized to TATA-binding protein levels using the following primers: *Agxt2* forward, ACCCGAAGGTAAGTGCAGTG; *Agxt2* reverse, CCAGGTCGTTGGCTTCTGAT; *Tbp* forward, GGGTATCTGCTGGCGGTTT; *Tbp* reverse, TGAAATAGTGATGCTGGGCACT. Plasma DMGV was measured by LC/MS as described above for human plasma.

Statistics

All metabolite levels were standardized to the nearest pooled plasma metabolite value within the cohort and then natural logarithmically transformed because of their non-normal distribution. Age- and sex-adjusted (and also BMI- and glucose-adjusted) regression analyses were performed in each study sample to examine the relation of each metabolite (predictor variable) with each clinical metabolic trait (response variables): CT-defined LPR, BMI, fasting glucose, log HOMA-IR, SBP, DBP, log triglycerides, and HDL cholesterol. Conditional regression analyses were performed with adjustment for age and sex. Taking into account the more than 5,000 metabolites that were measured on the Hybrid platform, we used a Bonferroni-corrected threshold of 1×10^{-5}. On the targeted platform, in which we specifically focused on 1 metabolite (DMGV), we used a *P* value threshold of 0.05.

For the FHS GWAS analyses, a linear mixed-effects model accounting for familial relatedness with an additive genetic model with 1 degree of freedom was used (43). Clinical and laboratory continuous variables were compared between 2 independent groups using a Mann-Whitney *U* test. In the RYGB cohort, in which the same patients were followed over 3 time points, which is essentially a paired analysis, we used a Wilcoxon signed-rank test to compare the baseline with each time point. Categorical variables were compared using Fisher's exact test. Statistical analyses were performed using SAS (SAS Institute) and GraphPad Prism 6 (GraphPad Software).

For the MDC analyses, plasma metabolite levels were transformed into their natural logarithm (ln) and then standardized to multiples of 1 SD. Logistic regression analyses to test the correlation between metabolites and the incidence of diabetes during follow-up were performed in SPSS 22 in a model adjusted for age and sex, or for age, sex, BMI, and plasma glucose concentration, at baseline.

Study approval

The IRB of Massachusetts General Hospital approved all metabolomic analyses in the human cohort studies, and all participants provided written informed consent before participating in the studies.

Author contributions

REG conceived, supervised, guided, arranged collaborative projects, and wrote the manuscript. JFOS and JEM discovered and identified the unknown metabolite, tuned it on the targeted platform, performed the metabolomics profiling and analysis in the cohorts using this platform, performed data analysis, and wrote the manuscript. BZ and RTP synthesized and purified DMGV. JS, SJ, and CBC acquired the original metabolomic data

on the Hybrid platform and then performed the acquisition and fragmentation of DMGV to confirm its identity. QY performed statistical analysis. C. Fernandez and OM analyzed the MDC data. YG and JGW analyzed the JHS data. RSV and TJW designed and coordinated the FHS cohort studies. MTL and C. Fox designed the CT liver fat studies and analyzed the data. KEC provided the samples from the biopsy-proven NAFLD and the RYGB studies and performed the analyses for these cohorts. S. O'Connor, and PC performed and analyzed the murine studies. HZ performed analyses on the biopsy-proven NAFLD and the RYGB studies.

1. Chalasani N, et al. The diagnosis and management of non-alcoholic fatty liver disease: practice guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. *Hepatology*. 2012;55(6):2005–2023.

2. Blackett PR, Sanghera DK. Genetic determinants of cardiometabolic risk: a proposed model for phenotype association and interaction. *J Clin Lipidol*. 2013;7(1):65–81.

3. Targher G, et al. Prevalence of nonalcoholic fatty liver disease and its association with cardiovascular disease among type 2 diabetic patients. *Diabetes Care*. 2007;30(5):1212–1218.

4. Targher G, et al. Relations between carotid artery wall thickness and liver histology in subjects with nonalcoholic fatty liver disease. *Diabetes Care*. 2006;29(6):1325–1330.

5. Rinella ME. Nonalcoholic fatty liver disease: a systematic review. *JAMA*. 2015;313(22):2263–2273.

6. Wishart DS, et al. HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res*. 2013;41(Database issue):D801–D807.

7. Smith CA, et al. METLIN: a metabolite mass spectral database. *Ther Drug Monit*. 2005;27(6):747–751.

8. Brumfiel G. Chemists spin a web of data. *Nature*. 2008;453(7192):139.

9. Kim S, et al. PubChem Substance and Compound databases. *Nucleic Acids Res*. 2016;44(D1):D1202–D1213.

10. Zhu ZJ, et al. Liquid chromatography quadrupole time-of-flight mass spectrometry characterization of metabolites guided by the METLIN database. *Nat Protoc*. 2013;8(3):451–460.

11. Kalisiak J, et al. Identification of a new endogenous metabolite and the characterization of its protein interactions through an immobilization approach. *J Am Chem Soc*. 2009;131(1):378–386.

12. Demirkan A, et al. Genome-wide association study identifies novel loci associated with circulating phospho- and sphingolipid concentrations. *PLoS Genet*. 2012;8(2):e1002490.

13. Gieger C, et al. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet*. 2008;4(11):e1000282.

14. Hicks AA, et al. Genetic determinants of circulating sphingolipid concentrations in European populations. *PLoS Genet*. 2009;5(10):e1000672.

15. Illig T, et al. A genome-wide perspective of genetic variation in human metabolism. *Nat Genet*. 2010;42(2):137–141.

16. Kettunen J, et al. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat Genet*. 2012;44(3):269–276.

17. Suhre K, et al. A genome-wide association study of metabolic traits in human urine. *Nat Genet*. 2011;43(6):565–569.

18. Suhre K, et al. Human metabolic individuality in biomedical and pharmaceutical research. *Nature*. 2011;477(7362):54–60.

19. Tukiainen T, et al. Detailed metabolic and genetic characterization reveals new associations for 30 known lipid loci. *Hum Mol Genet*. 2012;21(6):1444–1455.

20. Rhee EP, et al. A genome-wide association study of the human metabolome in a community-based cohort. *Cell Metab*. 2013;18(1):130–143.

21. Rhee EP, et al. An exome array study of the plasma metabolome. *Nat Commun*. 2016;7:12360.

22. Speliotes EK, et al. Liver fat is reproducibly measured using computed tomography in the Framingham Heart Study. *J Gastroenterol Hepatol*. 2008;23(6):894–899.

23. Cheng S, et al. Metabolite profiling identifies pathways associated with metabolic risk in humans. *Circulation*. 2012;125(18):2222–2231.

24. Pavlova NA. [Biosynthesis of glycocoll from glyoxylic and amino acids]. *Dokl Akad Nauk SSSR*. 1955;100(5):947–949.

25. Rodionov RN, Jarzebska N, Weiss N, Lentz SR. AGXT2: a promiscuous aminotransferase. *Trends Pharmacol Sci*. 2014;35(11):575–582.

26. Brenton AG, Godfrey AR. Accurate mass measurement: terminology and treatment of data. *J Am Soc Mass Spectrom*. 2010;21(11):1821–1835.

27. Ogawa T, Kimoto M, Watanabe H, Sasaoka K. Metabolism of NG,NG-and NG,N'G-dimethylarginine in rats. *Arch Biochem Biophys*. 1987;252(2):526–537.

28. Martens-Lobenhoffer J, Rodionov RN, Drust A, Bode-Böger SM. Detection and quantification of α-keto-δ-(N(G),N(G)-dimethylguanidino)valeric acid: a metabolite of asymmetric dimethylarginine. *Anal Biochem*. 2011;419(2):234–240.

29. Newgard CB, et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab*. 2009;9(4):311–326.

30. Wang TJ, et al. Metabolite profiles and the risk of developing diabetes. *Nat Med*. 2011;17(4):448–453.

31. Tolman KG, Dalpiaz AS. Treatment of non-alcoholic fatty liver disease. *Ther Clin Risk Manag*. 2007;3(6):1153–1163.

32. Roberts LD, et al. β-Aminoisobutyric acid induces browning of white fat and hepatic β-oxidation and is inversely correlated with cardiometabolic risk factors. *Cell Metab*. 2014;19(1):96–108.

33. Ferrara CT, et al. Genetic networks of liver metabolism revealed by integration of metabolic and transcriptional profiling. *PLoS Genet*. 2008;4(3):e1000034.

34. Li T, et al. Dimethylarginine dimethylaminohydrolase 1 protects against high-fat diet-induced hepatic steatosis and insulin resistance in mice. *Antioxid Redox Signal*. 2017;26(11):598–609.

35. Davids M, et al. Role of dimethylarginine dimethylaminohydrolase activity in regulation of tissue and plasma concentrations of asymmetric dimethylarginine in an animal model of prolonged critical illness. *Metab Clin Exp*. 2012;61(4):482–490.

36. Ferrigno A, Di Pasqua LG, Berardo C, Richelmi P, Vairetti M. Liver plays a central role in asymmetric dimethylarginine-mediated organ injury. *World J Gastroenterol*. 2015;21(17):5131–5137.

37. Krumsiek J, et al. Mining the unknown: a systems approach to metabolite identification combining genetic and metabolic information. *PLoS Genet*. 2012;8(10):e1003005.

38. Fall T, et al. Non-targeted metabolomics combined with genetic analyses identifies bile acid synthesis and phospholipid metabolism as being associated with incident type 2 diabetes. *Diabetologia*. 2016;59(10):2114–2124.

39. Shen M, et al. Leveraging non-targeted metabolite profiling via statistical genomics. *PLoS ONE*. 2013;8(2):e57667.

40. Kleiner DE, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology*. 2005;41(6):1313–1321.

41. Persson M, Hedblad B, Nelson JJ, Berglund G. Elevated Lp-PLA2 levels add prognostic information to the metabolic syndrome on incidence of cardiovascular events among middle-aged non-diabetic subjects. *Arterioscler Thromb Vasc Biol*. 2007;27(6):1411–1416.

42. Klein C, Schulz G, Steglich W. Umwandlung von Guanidino und Ureido-alpha-aminosauren in alpha-Ketosauren und deren heterocyclische Folgeprodukte. *Liebigs Ann Chem*. 1983;1983(9):1623–1637.

43. Chen MH, Yang Q. GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics*. 2010;26(4):580–581.