

## GENETICS

# Loss-of-function variants influence the human serum metabolome

Bing Yu,<sup>1</sup> Alexander H. Li,<sup>1</sup> Ginger A. Metcalf,<sup>2</sup> Donna M. Muzny,<sup>2</sup> Alanna C. Morrison,<sup>1</sup> Simon White,<sup>2</sup> Thomas H. Mosley,<sup>3</sup> Richard A. Gibbs,<sup>2</sup> Eric Boerwinkle<sup>1,2\*</sup>

2016 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC). 10.1126/sciadv.1600800

The metabolome is a collection of small molecules resulting from multiple cellular and biological processes that can act as biomarkers of disease, and African-Americans exhibit high levels of genetic diversity. Exome sequencing of a sample of deeply phenotyped African-Americans allowed us to analyze the effects of annotated loss-of-function (LoF) mutations on 308 serum metabolites measured by untargeted liquid and gas chromatography coupled with mass spectrometry. In an independent sample, we identified and replicated four genes harboring six LoF mutations that significantly affected five metabolites. These sites were related to a 19 to 45% difference in geometric mean metabolite levels, with an average effect size of 25%. We show that some of the affected metabolites are risk predictors or diagnostic biomarkers of disease and, using the principle of Mendelian randomization, are in the causal pathway of disease. For example, LoF mutations in *SLCO1B1* elevate the levels of hexadecanedioate, a fatty acid significantly associated with increased blood pressure levels and risk of incident heart failure in both African-Americans and an independent sample of European-Americans. We show that *SLCO1B1* LoF mutations significantly increase the risk of incident heart failure, thus implicating the metabolite in the causal pathway of disease. These results reveal new avenues into gene function and the understanding of disease etiology by integrating -omic technologies into a deeply phenotyped population study.

## INTRODUCTION

Defining the genetic architecture of health and diseases is a key goal of contemporary human genetics. The human metabolome, a complete set of small molecules reflecting multiple metabolic and physiological processes, holds promise to bridge gene action to disease end points via intermediate phenotypes (1). Annotated loss-of-function (LoF) variants, which are expected to signal nonsense-mediated decay (2), are likely to have large effects on phenotypes and occur more often in genomes of African descent than in genomes of European descent (3). LoF variants that lower risk factor levels and disease risk have been proven to be good predictors of drug efficacy (4, 5). Given the biological role of LoF variants and the metabolome, we leveraged these two -omic technologies in a deeply phenotyped population of African-Americans to systematically investigate the effects of LoF sites on the human metabolome and further influences on disease risk.

## RESULTS

We sequenced the exomes and measured the metabolomes of 1361 African-American (described in table S1) participants from the Atherosclerosis Risk in Communities (ARIC) Study. We defined LoF variation as sequence changes caused by single-nucleotide variants (SNVs) or small insertions and deletions (indels), which are predicted to result in a nonviable transcript or a greatly truncated protein product. A total of 7038 genes harboring 12,522 LoF variant sites (5060 stopgains, 2599 splice sites, and 4863 frameshift indels) were identified with an average of 111.7 heterozygous and 14.5 homozygous LoF sites per person (fig.

S1). We quantified metabolomic profiles in serum using liquid and gas chromatography followed by mass spectrometry (LC/GC-MS). After applying stringent quality control standards for metabolite levels, we analyzed 308 known serum metabolites consisting of 83 amino acids, 16 carbohydrates, 9 cofactors and vitamins, 7 energy metabolites, 136 lipids, 12 nucleotides, 25 peptides, and 20 xenobiotics (table S2).

To increase statistical power and reduce the false-positive rate, we limited the analyses to 324 LoF sites with a minor allele frequency (MAF)  $\geq 5\%$  and 1285 genes with a cumulative minor allele count (cMAC)  $\geq 7$ , using single-variant and gene-based burden tests, respectively. We identified nine genes harboring 18 LoF sites and affecting 10 metabolite levels (Table 1;  $P < 5.0 \times 10^{-7}$ , single-variant test;  $P < 1.3 \times 10^{-7}$ , burden test). Detailed results are shown in table S3, and quantile-quantile (QQ) plots are shown in fig. S2. These sites were related to a 19 to 45% difference in geometric mean metabolite levels, with an average effect size of 31%. Genotypes of all significantly associated sites were validated using an orthogonal laboratory technology (that is, Array-based genotyping, Sequenom genotyping, or Sanger sequencing). As a positive control, we observed that LoF sites in *PCSK9* had lower cholesterol levels compared to the noncarriers ( $P = 5.4 \times 10^{-9}$ ). Except for *PCSK9* and *HAL*, the remaining LoF sites in seven genes were novel findings (*C6orf25*, *CD36*, *FAM198B*, *LRRRC46*, *LRRRC69*, *SLCO1B1*, and *TEX15*; Table 1). To validate our findings, we sequenced the exomes of an additional 559 African-Americans and quantified their metabolomic profiles using LC/GC-MS. In this replication sample, 5 of 10 significant gene-metabolite relationships were reproduced ( $P < 0.05$ ; Table 1). Four genes did not have enough LoF carriers to be analyzed, and one gene-metabolite association did not reach nominal statistical significance, although the direction of the effect was consistent.

Homozygous LoF genotypes are rare in the population at each site, uncommon within each individual across the genome, and more likely

<sup>1</sup>Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX 77030, USA. <sup>2</sup>Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA. <sup>3</sup>Department of Medicine (Geriatrics), University of Mississippi Medical Center, Jackson, MS 39216, USA.

\*Corresponding author. Email: eric.boerwinkle@uth.tmc.edu

**Table 1. Ten significant gene-metabolite associations identified among African-Americans in ARIC.** Decanoylcarnitine and octanoylcarnitine were identified using the single-variant test, and the remains were identified using the gene-based burden test. Gene-related human pleiotropy was obtained through OMIM (Online Mendelian Inheritance in Man) ([www.ncbi.nlm.nih.gov/omim](http://www.ncbi.nlm.nih.gov/omim)) and the NHGRI GWAS catalog ([www.genome.gov/gwastudies](http://www.genome.gov/gwastudies)). NA, not applicable; LDL, low-density lipoprotein; MI, myocardial infarction; 5-HETE, 5-hydroxyicosatetraenoic acid; CHD, coronary heart disease; CVD, cardiovascular disease; MCAD, medium-chain acyl-CoA dehydrogenase.

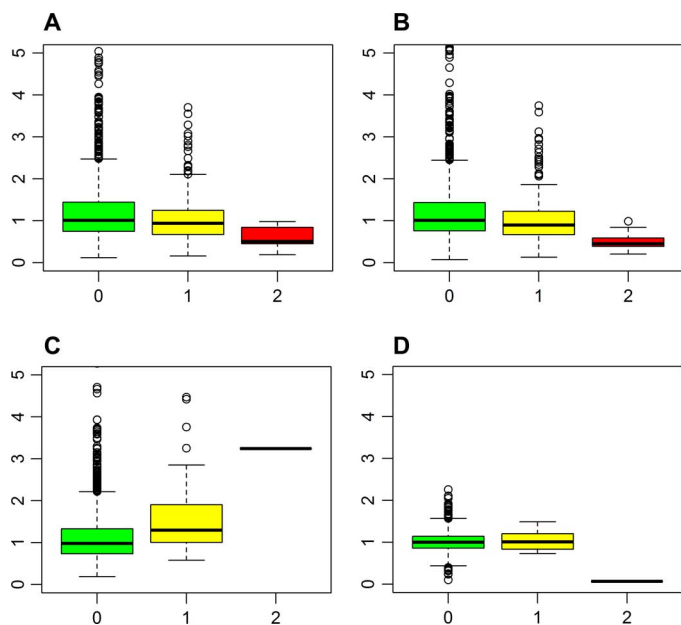
Metabolite	Pathway	Gene	P	$\beta$ (SE)	cMAC	Gene-related human pleiotropy	Metabolite-related human phenotypes	Replication P
Decanoylcarnitine	Lipid	<i>CD36</i>	$2.7 \times 10^{-7}$	-0.19 (0.04)	221	Platelet glycoprotein IV deficiency, CHD	MCAD deficiency (7)	0.03
Octanoylcarnitine	Lipid	<i>CD36</i>	$3.9 \times 10^{-8}$	-0.20 (0.04)	221	Platelet glycoprotein IV deficiency, CHD	MCAD deficiency (7)	0.03
Histidine	Amino acid	<i>HAL</i>	$2.3 \times 10^{-13}$	0.23 (0.03)	26	Histidinemia	Rheumatoid arthritis (23), CHD (24)	0.05
Methionine sulfoxide	Amino acid	<i>C6orf25</i>	$1.3 \times 10^{-8}$	-0.45 (0.08)	9	—	Aging (25)	NA
Mannose	Carbohydrate	<i>TEX15</i>	$7.9 \times 10^{-9}$	-0.70 (0.12)	10	—	—	NA
Cholesterol	Lipid	<i>PCSK9</i>	$5.4 \times 10^{-9}$	-0.21 (0.04)	30	LDL, CHD, MI, statin response, familial hypercholesterolemia	Atherosclerosis (26), CVD (27)	0.0002
Deoxycarnitine	Lipid	<i>LRRC69</i>	$8.5 \times 10^{-16}$	-0.42 (0.05)	17	—	—	0.6
Hexadecanedioate	Lipid	<i>SLCO1B1</i>	$2.2 \times 10^{-9}$	0.38 (0.06)	67	Hexadecanedioate, tetradecanedioate, octadecanedioate, bilirubin, hyperbilirubinemia, statin-induced myopathy, statin response	—	0.0001
5-HETE	Lipid	<i>FAM198B</i>	$4.5 \times 10^{-9}$	-0.38 (0.07)	14	—	—	NA
Urate	Nucleotide	<i>LRRC46</i>	$1.1 \times 10^{-7}$	-0.44 (0.08)	10	—	Gout (28), type 2 diabetes (29), CVD (30) mortality (31)	NA

to lead to extreme phenotypes. Here, we observed three genes (*CD36*, *SLCO1B1*, and *LRRC69*), each of which had homozygous stopgain mutations influencing four metabolite levels (Fig. 1). Each additional LoF allele in *CD36* (c.975T>G) had a lowering effect on octanoylcarnitine and decanoylcarnitine levels, and each additional LoF allele in *SLCO1B1* (c.481+1G>T) demonstrated an increasing effect on hexadecanedioate levels. *LRRC69* (c.933+2T>A) homozygous LoF status appeared to have a marked effect on deoxycarnitine levels ( $P = 6.8 \times 10^{-35}$ ). We next genotyped these three variants in an independent sample of 508 ARIC African-Americans and replicated our findings with *CD36* and *SLCO1B1* ( $P < 0.017$ , Bonferroni correction); no homozygote was observed in the replication genotype samples for *LRRC69*.

*CD36* is known as a fatty acid translocase that enhances cellular fatty acid uptake, which is a key step in energy metabolism (6). We discovered that LoF mutation in *CD36* influences serum octanoylcarnitine and decanoylcarnitine levels—two acylcarnitine biomarkers of medium-chain acyl-coenzyme A dehydrogenase (MCAD) deficiency (7)—and replicated this mutation. MCAD deficiency [MIM (Mendelian Inheritance in Man): 201450] is a rare metabolic disorder caused by homozygous or compound heterozygous mutations in the *ACADM* gene, which prevents the body from converting certain fatty acids to energy, especially under fasting conditions. Our results suggest an additional role of *CD36* in regulating acylcarnitine levels in addition to *ACADM*.

In genome-wide association studies (GWAS), common non-functional variants tagging *SLCO1B1* have been reported to be associated with fatty acid levels, including hexadecanedioate (8). Hexadecanedioic acid is a long-chain dicarboxylic acid, generated from fatty acid  $\omega$ -oxidation and thereafter metabolized by  $\beta$ -oxidation in peroxisomes. *SLCO1B1* encodes a protein that mediates the cellular uptake of numerous endogenous compounds and is involved in the metabolism and clearance of multiple drug compounds, including (the popular cholesterol-lowering drug) statins (9). Here, we show that LoF variants in *SLCO1B1* have a profound effect on serum hexadecanedioate levels ( $P = 2.2 \times 10^{-9}$ ). We next genotyped the LoF site in *SLCO1B1* (c.481+1G>T) among the entire ARIC participants (10,263 European-Americans and 3543 African-Americans) and demonstrated that one copy of mutated T allele was associated with 29% increased risk for heart failure (HF) [hazard ratio (HR), 1.29;  $P = 0.048$ ], indicating a direct moderate effect of *SLCO1B1* on HF. One other disease relationship cascade, *TEX15* LoF variants with mannose levels and mannose levels with diabetes and HF (fig. S3), was unsurprising because of the correlation between mannose and glucose.

To determine whether this fatty acid is involved in the development of HF in a general population, we examined its association with incident HF among ARIC African-Americans ( $n = 1792$ ), who had a median follow-up time of 22 years. We show that high levels of hexadecanedioate



**Fig. 1. Distribution of metabolite levels among LoF mutation carriers in ARIC.** The x axis indicates the number of LoF allele carried, and the y axis indicates the metabolite levels. **(A)** Octanoylcarnitine levels among *CD36* (c.975T>G) LoF mutation carriers. **(B)** Decanoylcarnitine levels among *CD36* (c.975T>G) LoF mutation carriers. **(C)** Hexadecanedioate levels among *SLC01B1* (c.481+1G>T) LoF mutation carriers. **(D)** Deoxycarnitine levels among *LRRC69* (c.933+2T>A) LoF mutation carriers.

are associated with increased risk of incident HF beyond the effect of traditional risk factors (HR, 1.22;  $P = 3.0 \times 10^{-7}$ ). To generalize its predictability of HF, we further measured this metabolite in 1421 ARIC European-Americans with 23.5 years of median follow-up time. Again, hexadecanedioate was positively and significantly related to incident HF risk in European-Americans (HR, 1.09;  $P = 0.005$ ). The relationship between hexadecanedioate levels and HF in both ancestries remained significant even in those without a history of previous myocardial infarction. In addition, in both European-Americans and African-Americans, high hexadecanedioate levels were related to high systolic blood pressure and diastolic blood pressure levels, which are important risk factors for HF (all  $P < 0.004$ ; table S4) (10). These observational data in humans are consistent with recent feeding studies in rats, showing that adding hexadecanedioate to the diet increased blood pressure levels (11). Vascular reactivity to noradrenaline was significantly increased in mesenteric resistance arteries among hexadecanedioate-treated rats (11), highlighting a possible vascular mechanism that underlies hexadecanedioate and blood pressure association.

## DISCUSSION

The transition from GWAS to exome sequencing-based rare variant studies has proven difficult, requiring large sample sizes and often leading to novel variants in known genes rather than the discovery of novel genes (12). Our study focused on the human metabolome, an intermediate phenotype that bridges gene effects to clinical end points, to promote novel gene discoveries. Because the samples originated from a large longitudinal cohort study, we were able to explore

the predictive relationships of the observed findings on the onset of disease and the maintenance of health. Because of their increased genetic diversity (13) and undue burden of disease, this study focused on African-Americans. Future studies are encouraged to include other ethnic groups. Our findings illuminate the value of using -omic studies in a deeply phenotyped population to generate new hypotheses and enhance our understanding of gene function and disease etiology.

## MATERIALS AND METHODS

### Study population and metabolome measurements

The ARIC Study is a prospective epidemiological study designed to investigate the etiology and predictors of cardiovascular disease (CVD). The ARIC Study enrolled 15,792 individuals aged 45 to 64 years from four U.S. communities (Forsyth County, NC; Jackson, MS; suburbs of Minneapolis, MN; and Washington County, MD) in 1987–1989 (baseline) who were followed up for four completed visits in 1990–1992, 1993–1995, 1996–1998, and 2011–2013. A detailed description of the ARIC study design and methods is published elsewhere (14). Basic CVD risk factors were measured at each visit, and CVD end points, including HF, were ascertained annually using telephone interviews and hospital medical record review. In ARIC, incident HF was defined as the first hospitalization or death from HF for those without a prior HF hospitalization. The diagnosis of HF was based on *International Classification of Diseases, Ninth Revision* (ICD-9) code 428.3. Individuals were followed up for events from baseline to 31 December 2011, and those who were lost to follow-up were censored at the date of last contact. Incident type 2 diabetes was defined in diabetes-free participants at baseline as having any of the following during a follow-up visit until visit 4: (i) a fasting glucose level  $\geq 7.0$  mM, (ii) a nonfasting glucose level  $\geq 11.1$  mM, (iii) use of a diabetes medication, or (iv) a self-reported physician diagnosis.

Metabolite profiling was measured using fasting serum samples collected at the baseline examination in 1987–1989 among ARIC European-Americans and African-Americans. For the discovery African-American samples, a total of 602 metabolites were detected and semiquantified by Metabolon Inc. using an untargeted LC/GC-MS-based metabolomic quantification protocol (15, 16). Metabolites were excluded if (i) more than 50% of the samples had values below the detection limit or (ii) they had unknown chemical structures. After this assessment, a total of 308 named metabolites were included in the present study. For the replication samples, we focused on metabolites that were significantly identified in the discovery samples.

### Whole exome sequencing and variant validation

The annotated exome was captured by NimbleGen's VCRome2.1 (Roche NimbleGen), and the captured exons were sequenced using Illumina HiSeq 2000. The Burrows-Wheeler Aligner was used to align sequences to the hg19 reference genome (17). Allele calling and variant call file construction were performed using the Atlas2 suite (18) (Atlas-SNP and Atlas-Indel). SNVs were excluded if they had a posterior probability  $< 0.95$ , a total depth of coverage  $< 6\times$ , an allelic fraction  $< 0.1$ , 99% of reads in a single direction, and homozygous reference alleles with  $< 6\times$  coverage. Low-quality indels were excluded if they had a minimum total depth  $< 30$ , an allelic fraction  $< 0.2$  for heterozygous variants and  $< 0.8$  for homozygous variants, and variant reads  $< 10\times$ .

Variants were annotated using ANNOVAR (19) and dbNSFP v2.0 (20) according to the reference genome GRCh37 and National Center for Biotechnology Information Reference Sequence. LoF variants included in this study were defined as premature stop codons occurring in the exon, essential splice site disrupting, and indels predicted to disrupt the downstream reading frame.

Following statistical analysis, all significantly associated variants were validated using an orthogonal laboratory technology (that is, Array-based genotyping, Sequenom genotyping, or Sanger sequencing). All reported results were validated with 100% accuracy, indicating identical genotypes in applied genotyping or sequencing methods.

For the purpose of replication, three LoF sites—*CD36* (c.975T>G), *SLCO1B1* (c.481+1G>T), and *LRRC69* (c.933+2T>A), having both homozygous and heterozygous genotypes in the study sample and which were significantly related to a clinical phenotype—were genotyped in an independent sample of 508 ARIC African-Americans. The calling and quality control methods were described elsewhere (18).

### Statistical analyses

Metabolite levels below the detectable limit of the assay were imputed with the lowest detected value for that metabolite in all samples, and all metabolite values were natural log-transformed before the analyses. Single-variant analyses on sites with MAF  $\geq 5\%$  and burden T5 tests (21) that evaluate the joint effects of rare alleles (MAF  $< 5\%$ ) in a gene were conducted on each metabolite after adjusting for age, gender, estimate glomerular filtration rate calibrated (eGFR) (22), and population structure. Statistical significance was defined as  $P < 5.0 \times 10^{-7}$  for single-variant tests (Bonferroni correction of 99,792 tests: 324 variants  $\times$  308 metabolites) and  $P < 1.3 \times 10^{-7}$  for the T5 tests (Bonferroni correction of 395,780 tests: 1285 genes  $\times$  308 metabolites).

All available African-American samples with metabolomic data were used to estimate the association between metabolite levels and the longitudinal onset of four clinical end points, including coronary heart disease (CHD), HF, chronic kidney disease, and type 2 diabetes. Metabolite levels were standardized before the analysis. Cox proportional hazard models adjusting for age, gender, body mass index, hypertension, anti-hypertensive medication use, diabetes, high-density lipoprotein, total cholesterol, current smoking, eGFR, and prevalent CHD were generally applied for incident disease analyses. Prevalent cases were excluded for the corresponding incident disease analyses. The proportional hazard assumption was examined and not rejected. Statistical significance was defined as  $P < 4.5 \times 10^{-5}$  (Bonferroni correction of 1232 tests: 4 diseases  $\times$  308 metabolites).

Cox proportional hazard models were separately applied in ARIC European-Americans and African-Americans to evaluate whether the LoF variant in *SLCO1B1* was associated with incident HF adjusting for age, gender, and population structure. Inverse variance fixed-effect meta-analyses were used to combine the results across the races to obtain an overall disease risk estimate, SE, and  $P$  value.

Three LoF sites—*CD36* (c.975T>G), *SLCO1B1* (c.481+1G>T), and *LRRC69* (c.933+2T>A), having both homozygous and heterozygous genotypes in the study sample and which were significantly related to a clinical phenotype—were replicated in an independent sample of 508 ARIC African-Americans. Metabolite levels were natural log-transformed before the replication analyses along with adjustment for age, gender, eGFR, and population structure. Statistical significance for the replication study was defined as  $P < 0.017$  (Bonferroni correction of three tests). All the analyses were performed using R (www.r-project.org).

## SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/2/8/e1600800/DC1>

table S1. Baseline characteristics of African-Americans in ARIC having metabolomic and exome sequence data.

table S2. List of 308 named metabolites measured in ARIC.

table S3. Eighteen significant LoF sites with metabolite associations identified among African-Americans in ARIC.

table S4. Relationship between hexadecanedioate and blood pressure levels among European-Americans and African-Americans in ARIC.

fig. S1. Distribution of the number of LoF sites per individual among ARIC African-Americans.

fig. S2. QQ plots of the expected and observed  $-\log P$  values for 10 identified metabolites.

fig. S3. Pathways among genes, metabolites, and diseases identified among ARIC African-Americans.

## REFERENCES AND NOTES

1. K. Suhre, C. Gieger, Genetic variation in metabolic phenotypes: Study designs and applications. *Nat. Rev. Genet.* **13**, 759–769 (2012).
2. Y.-F. Chang, J. S. Imam, M. F. Wilkinson, The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* **76**, 51–74 (2007).
3. D. G. MacArthur, S. Balasubramanian, A. Frankish, N. Huang, J. Morris, K. Walter, L. Jostins, L. Habegger, J. K. Pickrell, S. B. Montgomery, C. A. Albers, Z. D. Zhang, D. F. Conrad, G. Lunter, H. Zheng, Q. Ayub, M. A. DePristo, E. Banks, M. Hu, R. E. Handsaker, J. A. Rosenfeld, M. Fromer, M. Jin, X. J. Mu, E. Khurana, K. Ye, M. Kay, G. I. Saunders, M.-M. Suner, T. Hunt, I. H. A. Barnes, C. Amid, D. R. Carvalho-Silva, A. H. Bignell, C. Snow, B. Yngvadottir, S. Campstead, D. N. Cooper, Y. Xue, I. G. Romero; 1000 Genomes Project Consortium, J. Wang, Y. Li, R. A. Gibbs, S. A. McCarrroll, E. T. Dermizakis, J. K. Pritchard, J. C. Barrett, J. Harrow, M. E. Hurles, M. B. Gerstein, C. Tyler-Smith, A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).
4. Myocardial Infarction Genetics Consortium Investigators, Inactivating mutations in *NPC1L1* and protection from coronary heart disease. *N. Engl. J. Med.* **371**, 2072–2082 (2014).
5. J. C. Cohen, E. Boerwinkle, T. H. Mosley Jr., H. H. Hobbs, Sequence variations in *PCSK9*, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.* **354**, 1264–1272 (2006).
6. A. Bonen, S. E. Campbell, C. R. Benton, A. Chabowski, S. L. M. Coort, X.-X. Han, D. P. Y. Koonen, J. F. C. Glatz, J. J. F. P. Luiken, Regulation of fatty acid transport by fatty acid translocase/CD36. *Proc. Nutr. Soc.* **63**, 245–249 (2004).
7. D. H. Chace, S. L. Hillman, J. L. K. Van Hove, E. W. Naylor, Rapid diagnosis of MCAD deficiency: Quantitative analysis of octanoylcarnitine and other acylcarnitines in newborn blood spots by tandem mass spectrometry. *Clin. Chem.* **43**, 2106–2113 (1997).
8. S.-Y. Shin, E. B. Fauman, A.-K. Petersen, J. Krumsiek, R. Santos, J. Huang, M. Arnold, I. Erte, V. Forgetta, T.-P. Yang, K. Walter, C. Menni, L. Chen, L. Vasquez, A. M. Valdes, C. L. Hyde, V. Wang, D. Ziemek, P. Roberts, L. Xi, E. Grundberg; Multiple Tissue Human Expression Resource (MuTHER) Consortium, M. Waldenberger, J. B. Richards, R. P. Mohney, M. V. Milburn, S. L. John, J. Trimmer, F. J. Theis, J. P. Overington, K. Suhre, M. J. Brodman, C. Gieger, G. Kastenmüller, T. D. Spector, N. Soranzo, An atlas of genetic influences on human blood metabolites. *Nat. Genet.* **46**, 543–550 (2014).
9. SEARCH Collaborative Group, E. Link, S. Parish, J. Armitage, L. Bowman, S. Heath, F. Matsuda, I. Gut, M. Lathrop, R. Collins, *SLCO1B1* variants and statin-induced myopathy—A genome-wide study. *N. Engl. J. Med.* **359**, 789–799 (2008).
10. J. E. Udelson, Heart failure with preserved ejection fraction. *Circulation* **124**, e540–e543 (2011).
11. C. Menni, D. Graham, G. Kastenmüller, N. H. J. Alharbi, S. M. Alsanosi, M. McBride, M. Mangino, P. Titcombe, S.-Y. Shin, M. Psatha, T. Geisendorfer, A. Huber, A. Peters, R. Wang-Sattler, T. Xu, M. J. Brodman, J. Trimmer, C. Reichel, R. P. Mohney, N. Soranzo, M. H. Edwards, C. Cooper, A. C. Church, K. Suhre, C. Gieger, A. F. Dominiczak, T. D. Spector, S. Padmanabhan, A. M. Valdes, Metabolomic identification of a novel pathway of blood pressure regulation involving hexadecanedioate. *Hypertension* **66**, 422–499 (2015).
12. TG and HL Working Group of the Exome Sequencing Project, J. Crosby, G. M. Peloso, P. L. Auer, D. R. Crosslin, N. O. Stitzel, L. A. Lange, Y. Lu, Z. Z. Tang, H. Zhang, G. Hindy, N. Masca, K. Stirrups, S. Kanoni, R. Do, G. Jun, Y. Hu, H. M. Kang, C. Xue, A. Goel, M. Farrall, S. Duga, P. A. Merlini, R. Asselta, D. Girelli, O. Olivieri, N. Martinelli, W. Yin, D. Reilly, E. Speliotes, C. S. Fox, K. Hveem, O. L. Holmen, M. Nikpay, D. N. Farlow, T. L. Assimes, N. Franceschini, J. Robinson, K. E. North, L. W. Martin, M. DePristo, N. Gupta, S. A. Escher, J. H. Jansson, N. Van Zuydam, C. N. Palmer, N. Wareham, W. Koch, T. Meitinger, A. Peters, W. Lieb, R. Erbel, I. R. Konig, J. Kruppa, F. Degenhardt, O. Gottesman, E. P. Bottinger, C. J. O'Donnell, B. M. Psaty, C. M. Ballantyne, G. Abecasis, J. M. Ordovas, O. Melander, H. Watkins, M. Orho-Melander, D. Ardisson, R. J. Loos, R. McPherson, C. J. Willer,



- J. Erdmann, A. S. Hall, N. J. Samani, P. Deloukas, H. Schunkert, J. G. Wilson, C. Kooperberg, S. S. Rich, R. P. Tracy, D. Y. Lin, D. Altshuler, S. Gabriel, D. A. Nickerson, G. P. Jarvik, L. A. Cupples, A. P. Reiner, E. Boerwinkle, S. Kathiresan, Loss-of-function mutations in *APOC3*, triglycerides, and coronary disease. *N. Engl. J. Med.* **371**, 22–31 (2014).
13. M. C. Campbell, S. A. Tishkoff, African genetic diversity: Implications for human demographic history, modern human origins, and complex disease mapping. *Annu. Rev. Genomics Hum. Genet.* **9**, 403–433 (2008).
  14. The ARIC Investigators, The Atherosclerosis Risk in Communities (ARIC) Study: Design and objectives. *Am. J. Epidemiol.* **129**, 687–702 (1989).
  15. T. Ohta, N. Masutomi, N. Tsutsui, T. Sakairi, M. Mitchell, M. V. Milburn, J. A. Ryals, K. D. Beebe, L. Guo, Untargeted metabolomic profiling as an evaluative tool of fenofibrate-induced toxicology in Fischer 344 male rats. *Toxicol. Pathol.* **37**, 521–535 (2009).
  16. A. M. Evans, C. D. DeHaven, T. Barrett, M. Mitchell, E. Milgram, Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal. Chem.* **81**, 6656–6667 (2009).
  17. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
  18. D. Challis, J. Yu, U. S. Evani, A. R. Jackson, S. Paithankar, C. Coarfa, A. Milosavljevic, R. A. Gibbs, F. Yu, An integrative variant analysis suite for whole exome next-generation sequencing data. *BMC Bioinformatics* **13**, 8 (2012).
  19. K. Wang, M. Li, H. Hakonarson, ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
  20. X. Liu, X. Jian, E. Boerwinkle, dbNSFP v2.0: A database of human non-synonymous SNVs and their functional predictions and annotations. *Hum. Mutat.* **34**, E2393–E2402 (2013).
  21. B. Li, S. M. Leal, Methods for detecting associations with rare variants for common diseases: Application to analysis of sequence data. *Am. J. Hum. Genet.* **83**, 311–321 (2008).
  22. A. S. Levey, L. A. Stevens, C. H. Schmid, Y. L. Zhang, A. F. Castro III, H. I. Feldman, J. W. Kusek, P. Eggers, F. Van Lente, T. Greene, J. Coresh; CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration), A new equation to estimate glomerular filtration rate. *Ann. Intern. Med.* **150**, 604–612 (2009).
  23. D. A. Gerber, Low free serum histidine concentration in rheumatoid arthritis. A measure of disease activity. *J. Clin. Invest.* **55**, 1164–1173 (1975).
  24. B. Yu, A. H. Li, D. Muzny, N. Veeraghavan, P. S. de Vries, J. C. Bis, S. K. Musani, D. Alexander, A. C. Morrison, O. H. Franco, A. Uitterlinden, A. Hofman, A. Dehghan, J. G. Wilson, B. M. Psaty, R. Gibbs, P. Wei, E. Boerwinkle, Association of rare loss-of-function alleles in *HAL*, serum histidine levels and incident coronary heart disease. *Circ. Cardiovasc. Genet.* **8**, 351–355 (2015).
  25. E. R. Stadtman, H. Van Remmen, A. Richardson, N. B. Wehr, R. L. Levine, Methionine oxidation and aging. *Biochim. Biophys. Acta* **1703**, 135–140 (2005).
  26. J. Plutzky, Peroxisome proliferator-activated receptors in vascular biology and atherosclerosis: Emerging insights for evolving paradigms. *Curr. Atheroscler. Rep.* **2**, 327–335 (2000).
  27. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III), Third report of the national cholesterol education program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III) final report. *Circulation* **106**, 3143–3421 (2002).
  28. H. K. Choi, D. B. Mount, A. M. Reginato, Pathogenesis of gout. *Ann. Intern. Med.* **143**, 499–516 (2005).
  29. A. Dehghan, M. van Hoek, E. J. G. Sijbrands, A. Hofman, J. C. M. Witteman, High serum uric acid as a novel risk factor for type 2 diabetes. *Diabetes Care* **31**, 361–362 (2008).
  30. M. R. Hayden, S. C. Tyagi, Uric acid: A new look at an old risk marker for cardiovascular disease, metabolic syndrome, and type 2 diabetes mellitus: The urate redox shuttle. *Nutr. Metab.* **1**, 10 (2004).
  31. L. K. Niskanen, D. E. Laaksonen, K. Nyssönen, G. Alftan, H. M. Lakka, T. A. Lakka, J. T. Salonen, Uric acid level as a risk factor for cardiovascular and all-cause mortality in middle-aged men: A prospective cohort study. *Arch. Intern. Med.* **164**, 1546–1551 (2004).
- Acknowledgments:** We acknowledge the CHARGE Consortium for its essential role in the development of this article as well as for providing support. We also thank the staff and participants of the ARIC Study for their important contributions. **Funding:** The ARIC Study was carried out as a collaborative study that was supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367, and R01HL086694; National Human Genome Research Institute (NHGRI) contract U01HG004402; and NIH contract HHSN268200625226C. The metabolome measurement work was obtained through support from the National Genome Research Institute (HG004402). The DNA sequence data work was obtained through support from the National Heart, Lung, and Blood Institute (HL102419) and NHGRI (HG003273 and HG006542) of the NIH. **Author contributions:** B.Y. performed statistical analyses. A.H.L. performed variant quality control and annotation. G.A.M., D.M.M., and S.W. ensured that high-quality sequence variants were delivered for analyses. E.B. and T.H.S. were involved with the study design. R.A.G. and E.B. provided materials and project oversight. B.Y., E.B., A.C.M., and A.H.L. prepared the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.
- Submitted 14 April 2016  
 Accepted 3 August 2016  
 Published 31 August 2016  
 10.1126/sciadv.1600800
- Citation:** B. Yu, A. H. Li, G. A. Metcalf, D. M. Muzny, A. C. Morrison, S. White, T. H. Mosley, R. A. Gibbs, E. Boerwinkle, Loss-of-function variants influence the human serum metabolome. *Sci. Adv.* **2**, e1600800 (2016).

## Loss-of-function variants influence the human serum metabolome

Bing Yu, Alexander H. Li, Ginger A. Metcalf, Donna M. Muzny, Alanna C. Morrison, Simon White, Thomas H. Mosley, Richard A. Gibbs and Eric Boerwinkle

*Sci Adv* 2 (8), e1600800.  
DOI: 10.1126/sciadv.1600800

### ARTICLE TOOLS

<http://advances.sciencemag.org/content/2/8/e1600800>

### SUPPLEMENTARY MATERIALS

<http://advances.sciencemag.org/content/suppl/2016/08/29/2.8.e1600800.DC1>

### REFERENCES

This article cites 31 articles, 6 of which you can access for free  
<http://advances.sciencemag.org/content/2/8/e1600800#BIBL>

### PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)