Effect of Genetic and Dietary Perturbation of Glycine Metabolism on
Atherosclerosis in Humans and Mice

Subarna Biswas, MD1,2,*, James R. Hilser, MPH2,3,*, Nicholas C. Woodward, PhD2,3,*, Zeneng Wang, PhD4,5,6, Janet Gukasyan, BS2,3, Ina Nemet, PhD4,5, William S. Schwartzman, MD2,3, Pin Huang, MD2,3, Yi Han, PhD2,3, Zachary Fouladian, BS7, Sarada Charugundla7, Neal J. Spencer, BS2,3, Calvin Pan, MS8, W.H. Wilson Tang, MD4,5,6, Aldons J. Lusis, PhD7,8,9, Stanley L. Hazen, MD, PhD4,5,6, Jaana A. Hartiala, PhD2, and Hooman Allayee, PhD2,3

Departments of 1Surgery, 2Population & Public Health Sciences, and 3Biochemistry & Molecular Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033; Departments of 4Cardiovascular & Metabolic Sciences, Lerner Research Institute, 5Center for Microbiome and Human Health, and 6Cardiovascular Medicine, Heart, Vascular and Thoracic Institute, Cleveland Clinic, Cleveland, OH 44195; Departments of 7Medicine, 8Human Genetics, and 9Microbiology, Immunology, & Molecular Genetics, David Geffen School of Medicine of UCLA, Los Angeles, CA 90095.

*These authors contributed equally to this work.

Address correspondence and reprint requests to:

Hooman Allayee, PhD
Keck School of Medicine of USC
2250 Alcazar Street, CSC202
Los Angeles, CA 90033
Phone: (323) 442-1736
Fax: (323) 442-2764
Email: hallayee@usc.edu

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Objective: Epidemiological and genetic studies have reported inverse associations between circulating glycine levels and risk of coronary artery disease (CAD). However, these findings have not been consistently observed in all studies. We sought to evaluate the causal relationship between circulating glycine levels and atherosclerosis using large-scale genetic analyses in humans and dietary supplementation experiments in mice.

Methods: Serum glycine levels were evaluated for association with prevalent and incident CAD in the UK Biobank. A multi-ancestry genome-wide association study (GWAS) meta-analysis was carried out to identify genetic determinants for circulating glycine levels, which were then used to evaluate the causal relationship between glycine and risk of CAD by Mendelian randomization (MR). A glycine feeding study was carried out with atherosclerosis-prone apolipoprotein E deficient (ApoE−/−) mice to determine the effects of increased circulating glycine levels on amino acid metabolism, metabolic traits, and aortic lesion formation.

Results: Among 105,718 subjects from the UK Biobank, elevated serum glycine levels were associated with significantly reduced risk of prevalent CAD (Quintile 5 vs. Quintile 1 OR=0.76, 95% CI 0.67-0.87; P<0.0001) and incident CAD (Quintile 5 vs. Quintile 1 HR=0.70, 95% CI 0.65-0.77; P<0.0001) in models adjusted for age, sex, ethnicity, anti-hypertensive and lipid-lowering medications, blood pressure, kidney function, and diabetes. A meta-analysis of 13 GWAS datasets (total n=230,947) identified 61 loci for circulating glycine levels, of which 26 were novel. MR analyses provided modest evidence that genetically elevated glycine levels were causally associated with reduced systolic blood pressure and risk of type 2 diabetes, but did not provide evidence for an association with risk of CAD. Furthermore, glycine-supplementation in...
ApoE-/- mice did not alter cardiometabolic traits, inflammatory biomarkers, or development of atherosclerotic lesions.

**Conclusions:** Circulating glycine levels were inversely associated with risk of prevalent and incident CAD in a large population-based cohort. While substantially expanding the genetic architecture of circulating glycine levels, MR analyses and *in vivo* feeding studies in humans and mice, respectively, did not provide evidence that the clinical association of this amino acid with CAD represents a causal relationship, despite being associated with two correlated risk factors.

**Keywords:** glycine, coronary artery disease; genome-wide association study, Mendelian randomization; dietary supplementation; mice.
Introduction

Atherosclerotic coronary artery disease (CAD) is a complex, multi-factorial process in which many of the underlying causal biological mechanisms remain unknown\textsuperscript{1}. For example, of the ~300 genetic risk loci identified for CAD, only one-third harbor genes known to be involved in traditional risk factors, such as elevated lipid levels or blood pressure\textsuperscript{2,3}. Furthermore, lipid lowering therapies and anti-hypertensive medications are only partially effective in reducing CAD risk, and more than 50\% of patients with an acute event do not exhibit these traditional risk factors\textsuperscript{4}. Thus, there is a critical need to identify other mechanisms underlying CAD pathogenesis.

Integrating metabolomics with genetic and clinical data can offer a window into the intricate pathways governing the complex pathophysiology underlying atherosclerosis\textsuperscript{5}. This approach has previously implicated glycine – a simple amino acid – as a potential factor in modulating risk of CAD\textsuperscript{6,7}. For example, dietary glycine supplementation was shown to reduce blood pressure in rats\textsuperscript{8} and platelet aggregation in both humans and mice\textsuperscript{9,10}, thus providing plausible interconnected mechanisms for the protective association observed between glycine levels and cardiovascular risk. However, recent genetic studies have provided conflicting evidence regarding the causal role of glycine in atherosclerosis\textsuperscript{11-13}. One potential explanation for these discrepancies may be due to the pleiotropic associations that several of the genetic variants associated with glycine levels exhibit with other CAD-related traits, thus potentially violating one of the central assumptions of Mendelian randomization (MR). Consequently, it remains unclear whether the epidemiological associations observed between higher glycine levels and reduced risk of CAD represents a true inverse causal relationship. In the present
study, we used complementary clinical, genetic, and dietary supplementation strategies to comprehensively evaluate the causal association of glycine with cardiovascular risk.
Results

Association of Serum Glycine Levels with Prevalent and Incident Risk of CAD in UK Biobank:

We first leveraged data from the UK Biobank to evaluate the association of circulating glycine levels with risk of CAD. The clinical characteristics of the subjects used for these analyses for whom complete clinical, demographic, and covariate were available are shown in Supplemental Table 1. Among 105,718 subjects from the UK Biobank (4,099 CAD cases/101,619 controls) with metabolomics data, higher circulating glycine levels were associated with reduced risk of prevalent CAD at the time of enrollment into UK Biobank (Table 1; Figure 1A). This atheroprotective association was particularly evident among subjects in the highest quintiles of glycine levels compared to subjects in the first quintile (OR=0.87, 95% CI 0.78-0.96; P<0.0001 for Q4 vs. Q1 and OR=0.76, 95% CI 0.67-0.87; P<0.0001 for Q5 vs. Q1) (Table 1; Figure 1A). Among the 101,608 control subjects without CAD at the time of enrollment into UK Biobank and for whom complete data were available, longitudinal analyses over 5,000 days (~13 years) of follow-up revealed that higher baseline circulating glycine levels were also associated with reduced risk of incident CAD (Table 1; Figure 1B). For example, subjects in the third, fourth, and fifth quintiles of glycine levels had 13-30% reduced risk of incident CAD (P<0.0001) compared to individuals with lowest levels of glycine (Table 1; Figure 1B).

Genome-Wide Meta-Analysis for Circulating Glycine Levels: We next carried out a large-scale GWAS meta-analysis to further define the genetic architecture of circulating glycine levels. In total, we combined GWAS summary statistics from 13 datasets comprising 230,947 multi-ancestry subjects (Supplemental Table 2). This analysis revealed 15,230 SNPs distributed among 61 loci that were associated with circulating glycine levels at the genome-wide
significance threshold ($P=5.0\times10^{-8}$) (Figure 2A; Supplemental Table 3). Twenty-six of these loci were newly identified herein as being significantly associated with serum glycine levels (Table 2; Supplemental Figure 1), whereas the remaining 35 had been identified in previous studies\textsuperscript{11,12,14} (Supplemental Table 3). As expected, the larger sample size and power in our meta-analysis increased significance levels at many previously known loci, including $CPS1$ (rs1047891; $P=7.8\times10^{-1101}$) and $GLDC$ (rs1801133; $P=3.9\times10^{-165}$), which remained the two strongest genetic determinants of circulating glycine levels (Supplemental Table 3). In addition, several of the 26 novel loci harbored genes involved in glucose metabolism and insulin regulation, such as $IRS1$ and $PPARG$, whereas genes known to be involved in the metabolism of glycine and other amino acids localized to other loci ($PCCB$, $CYP3A7$)\textsuperscript{15,16} (Supplemental Table 3; Supplemental Figure 1). Using the absolute levels of glycine available in the UK Biobank, we also constructed a weighted genetic risk score (GRS) with all 61 SNPs, which revealed a dose-dependent increase in glycine levels as a function of carrying alleles that were associated with higher glycine ($P$-trend=$1.57\times10^{-4}$). Compared to those in the bottom quintile for carrying glycine-raising alleles (range 104.2-149.0µM), serum glycine levels were significantly increased by $\sim67.0\pm3.0$µM ($P=8.1\times10^{-67}$) among individuals in the top quintile (range 173.0-214.6µM) (Figure 2B).

We next carried out PheWAS analyses, which revealed that 54 of the 61 loci identified for glycine levels also exhibited pleiotropic associations with levels of other circulating metabolites or known CAD risk factors, such as lipid levels and blood pressure (Supplemental Table 3). Of the remaining seven loci, three harbored genes encoding components of the glycine cleavage system ($AMT$, $GLDC$, and $GCSH$)\textsuperscript{17} (Supplemental Table 3; Supplemental Figure 1). Genes localizing to the remaining four loci included those involved in mitochondrial regulation
of oxidative stress (VWA8)\textsuperscript{18}, transcription and chromosome modulators (ZNF763, DHX38)\textsuperscript{19, 20}, or solute transport (AQP9)\textsuperscript{21}. Altogether, the lead variants at the 61 loci explained 15.6\% of the variance in glycine levels, while the seven non-pleiotropic SNPs alone explained 2.9\%. Of the seven non-pleiotropic variants, one SNP (rs61757601) was also significantly associated with CAD (P=1.3E-14) (Table 2).

\textit{Mendelian Randomization (MR) Analysis with Glycine-Associated SNPs and Risk of CAD}: We next evaluated whether the clinical association of glycine levels with risk of CAD represented a causal relationship using various MR analyses. Variants identified from our meta-analysis for glycine levels were used as genetic instruments for the exposure and the results of a recently published large-scale GWAS for CAD\textsuperscript{2, 3} were used for the outcome. Weighted median and inverse variance weighted MR analyses with all 61 glycine-associated SNPs yielded modest inverse associations with CAD (OR=0.96, 95\% CI 0.94-0.97; P=8.6E-07 for weighted median and OR=0.93, 95\% CI 0.88-0.98; P=0.01 for inverse variance weighted) (Figure 3A). However, analyses with MR Egger, which takes into account pleiotropic effects of variants, did not reveal significant evidence for a causal association between genetically increased glycine levels and risk of CAD (OR=0.97, 95\% CI 0.91-1.03; P=0.28) (Figure 3A). We also carried out weighted median and inverse variance weighted MR with the seven non-pleiotropic SNPs, which also did not yield causal evidence for the association between glycine and risk of CAD (OR=0.97, 95\% CI 0.89-1.05; P=0.39 for weighted median and OR=0.93, 95\% CI 0.73-1.17; P=0.52 for inverse variance weighted) (Figure 3B). Similarly, the most restrictive MR model that only included non-pleiotropic variants at three loci harboring genes involved in glycine cleavage system also did not reveal significant evidence for a causal association between glycine and risk of CAD.
(OR=0.96, 95% CI 0.89-1.05; P=0.39 for weighted median and OR=0.97, 95% CI 0.90-1.05; 
P=0.45 for inverse variance weighted) (Figure 3C). By comparison, MR analyses with the 
seven non-pleiotropic loci and known CAD risk factors did provide modest evidence for an 
inverse causal relationship between genetically increased glycine levels and systolic blood 
pressure and risk of T2D, but not with diastolic blood pressure, lipid levels, or T2D-related 
metabolic traits (Supplemental Table 5).

Cardiometabolic Effects of Glycine Supplementation in Mice: To complement the human 
studies, we next carried out a physiologically relevant feeding study in mice to evaluate the 
effect of dietary glycine supplementation on cardiometabolic traits and development of 
atherosclerosis (Table 3). Amino acid-defined chow diets were developed with either 2% or 
0.3% glycine content that maintained nitrogen balance (Supplemental Table 6), would perturb 
glycine levels similar to the natural variation observed in humans, and would avoid the potential 
confounding effects that high fat/high cholesterol atherogenic diets could have on glycine 
metabolism. Given these considerations, we also selected ApoE/− mice as the mouse model with 
which to carry out dietary supplementation since aortic lesions develop in this strain on a chow 
diet. After 16 weeks of feeding, fasting glycine levels were increased in ApoE/− mice fed the 2% 
glycine diet, with this elevation being highly significant in male mice and nearly significant in 
female mice (Figure 4A; Table 4). Non-fasting glycine levels were also robustly elevated in 
both male and female fed the glycine-enriched diet (Figure 4B; Table 4). Compared to the 
control diet, there were no significant differences in body weight, glucose and insulin-related 
metabolic traits, or lipid levels as a result of glycine supplementation, with the exception of a 
modest decrease in triglyceride levels in male mice (Figure 4C-J; Table 3). By comparison,
male, but not female, *ApoE-/-* mice in the glycine supplementation group exhibited small differences in plasma levels of glycine-related metabolites, various amino acids, inflammatory biomarkers, and blood cell traits (*Table 4; Supplemental Tables 7-8*). However, atherosclerotic lesion area at the aortic root or along the entire aorta by *en face* analysis was not significantly different between *ApoE-/-* mice in the glycine or control diet groups (*Figure 4K-L; Table 3*).
Discussion

Multiple indirect associations have suggested that glycine levels could be a protective biomarker of CAD risk. However, genetic studies have produced equivocal evidence for a causal link between glycine and CAD, which implies that this amino acid may be correlated with other as yet unrecognized causal biomarkers, metabolites, or pathogenic mechanisms. In the present study, we observed consistent epidemiological associations between circulating glycine levels and both prevalent and incident CAD in the UK Biobank. We also conducted the largest GWAS meta-analysis of glycine levels to date by combining summary statistics from >230,000 subjects in the UK Biobank and 12 other datasets. While substantially expanding our understanding of the genetic architecture of circulating glycine levels, various types of MR analyses in humans coupled with a large and well-controlled feeding experiment did not provide evidence that glycine plays a direct causal atheroprotective role in humans or mice.

Our large-scale genetic analysis identified 61 loci significantly associated with circulating glycine levels and strengthened the association signals at 35 previously known loci. Notably, the identification of 26 novel loci further revealed the diversity of metabolic pathways associated with glycine levels. For example, two novel loci harbored genes with well-known roles in glucose metabolism (IRS1 and PPARG). In addition, recent studies have implicated HNF4A and PROX1-AS1 in insulin resistance and T2D22,23, whereas loci harboring FAM13A, RSPO3, and EBPL have been associated with body fat distribution and hepatic steatosis24-26. Genes at other loci are likely involved in the synthesis or degradation of glycine itself. For instance, HOGA1 catalyzes the breakdown of hydroxyproline into glyoxylate, which can serve as a substrate for glycine production27. By comparison, DLD mediates the oxidation and activation of the enzyme encoded by GCSH, which in turn provides the substrate for the AMT enzyme in the oxidative...
cleavage of glycine. GCSH, DLD, and AMT, together with GLDC, are the four components of the glycine cleavage system that metabolizes glycine into ammonia and carbon dioxide for subsequent detoxification through the urea cycle. In this regard, GLDC was one of the most strongly associated loci for glycine levels in our meta-analysis. Moreover, arginase, encoded by ARG1, catalyzes conversion of arginine to ornithine in the final step of the urea cycle and is one of the major routes for removal of ammonia produced through the degradation of nitrogen-containing compounds, including glycine. However, the biological mechanisms underlying association of most of the loci for circulating glycine levels still remain to be determined.

A nonsynonymous Thr1405Asn substitution (rs1047891) in the gene encoding the rate-limiting enzyme of the urea cycle, CPS1, remained as the most significantly associated variant for glycine levels (P=7.8E-1101), consistent with prior studies. The results of a recent large-scale GWAS also revealed the glycine-raising allele of rs1047891 (1405Asn) to be associated with decreased risk of CAD, although not at the genome-wide significance threshold. The association of CPS1 with risk of CAD was previously shown to be female-specific - similar to the sexually dimorphic association patterns observed with levels of various metabolites including glycine, and more recently, reduced risk of T2D, in East Asians. Taken together, these observations suggest that glycine may have protective cardiometabolic properties. However, CPS1 is one of the most pleiotropically associated loci in the genome and rs1047891 has been linked to multiple other CAD risk factors, such as trimethylamine N-oxide (TMAO), blood cell traits, and uric acid levels. These observations would thus argue that the protective association of CPS1 with risk of CAD could also be due to one or more of these other traits, either individually or collectively, rather than glycine per se. This notion is supported by the observation that the second most significantly associated locus for glycine levels (GLDC)
was not associated with risk of CAD² or T2D³, whereas the loci most strongly associated with CAD (PIGV, TRIB1, and SLC22A3) had smaller effect sizes on glycine levels than CPS1 and GLDC.

Prior MR analyses in humans with sample sizes smaller than in our present study had suggested that genetically higher glycine levels were associated with reduce risk of CAD¹²,¹³. We also used various forms of MR with larger numbers of subjects and genetic instruments to evaluate whether glycine levels were causally associated with risk of CAD. When using lead variants at all 61 identified loci, MR analyses did yield evidence that genetically higher glycine levels were modestly associated with reduced risk of CAD. However, most of the glycine-associated loci (i.e., CPS1) had pleiotropic effects on other CAD-related traits, thus violating one of the fundamental assumptions of MR analysis. Therefore, we could not conclude, based on these data alone, that the atheroprotective association of glycine with CAD was attributable entirely to this amino acid. The lack of significant evidence for a causal association between genetically increased glycine levels and risk of CAD from the MR Egger analyses, which takes into account pleiotropic effects of variants, further suggested that glycine may not be causally associated with risk of CAD. To address this issue, we specifically carried out MR analyses with the seven glycine-associated loci for which there were no other reported associations, including the variants at three of the four loci harboring genes of the glycine cleavage complex (AMT, GLDC, and GCSH). However, MR analyses with both sets of non-pleiotropic variants also did not provide evidence for a causal relationship between higher glycine levels and lower risk of CAD. Collectively, these results do not provide compelling evidence for glycine having direct cardioprotective properties. MR analyses with the seven non-pleiotropic variants did suggest that genetically higher glycine levels may be associated with a modest causal effect on decreased
systolic blood pressure and risk of T2D. These observations suggest that the inverse clinical associations we and others observe between glycine levels and risk of CAD could be indirect and possibly mediated through effects on peripheral metabolism and/or the vasculature.

To complement the human studies, we also carried out a comprehensive feeding study in ApoE−/− mice with isocaloric, amino acid-defined diets. Despite elevating both fasting and non-fasting glycine levels, particularly in males, the glycine-enriched diet did not reduce aortic lesion development. For the most part, glycine supplementation also did not change plasma levels of lipid and metabolic traits, other amino acids, inflammatory biomarkers, or blood cell parameters, particularly among female mice who are more prone to atherosclerosis than male mice.

Importantly, our study was well-powered to detect differences in these phenotypes since, for example, there were at least ~20 mice of each sex (and in some cases even >30 mice) in most of the experimental groups, including those used to evaluate aortic lesion formation. Taken together with our human studies, these data further argue against glycine having atheroprotective properties. By comparison, a previous study had suggested that glycine feeding in combination with a Western diet could mitigate atherosclerosis in ApoE−/− mice39. However, the effects of the diets on circulating glycine or lipid levels after supplementation were not shown and only male mice were used, with fewer animals in each experimental group39 than in our study.

Interestingly, other rodent feeding studies have shown improvements in insulin sensitivity with glycine supplementation, although these effects were primarily observed in the context of high fat diets as well40. By contrast, chow diets supplemented with glycine did not lead to metabolic changes even with longer feeding durations41. Our results would be consistent with these latter observations since we also used chow diets with modified glycine content.
While our results point to interesting clinical and genetic associations with glycine levels, we also note certain limitations of our study. First, glycine measurements in the UK Biobank were only available at the time of enrollment and may fluctuate over time. There may also have been residual confounding in our multivariate regression models for prevalent and incident CAD, since some risk factors remain unknown or unmeasured. Furthermore, because the meta-analysis involved several different datasets, metabolomics measurements were carried out on different platforms, in either serum or plasma, and at different levels of dietary intake. However, the error stemming from these variations would likely bias the results towards the null and decrease the likelihood of identifying significant genetic associations. In addition, unlike the wide variation observed in overall glycine levels in the clinical analyses, the genetic instruments used in our MR analyses, even in aggregate and when only carried out with non-pleiotropic variants, may not have provided sufficient variation in glycine levels to detect significant causal associations with risk of CAD. By comparison, the stringent criteria we used to select genetic instruments in MR analyses resulted in evidence for causal associations between glycine and decreased systolic blood pressure and risk of T2D, consistent with prior studies.12, 14. Last, it is also possible that our feeding study did not elevate glycine, at least with respect to fasting levels, sufficiently to reduce atherosclerosis. However, we elected to use diets that would lead to elevations of glycine levels that would be physiologically relevant to humans. In the UK Biobank, individuals in the top quintile for carrying glycine-raising alleles at all 61 identified loci had glycine levels that were on average ~67μM higher than those in the bottom quintile. Furthermore, the 1405Asn variant of CPS1, which is the strongest genetic determinant of glycine levels in humans and is associated with reduced risk of CAD in women, increases glycine by ~50μM per allele.32. Thus,
the increased glycine levels observed in our feeding study with ApoE-/- mice were comparable to the effects of naturally occurring genetic variants on glycine levels in humans.

In summary, our results provide a more complete picture of the genetic architecture of glycine metabolism in humans and offer numerous opportunities to further elucidate glycine metabolism in mammals. While not obtaining evidence for a direct causal relationship between glycine and atherosclerosis in humans or mice, our results still suggest glycine could indirectly mitigate CAD risk, possibly through metabolic processes and/or blood pressure regulation. Additional studies will be required to explore these possibilities.
Methods

Study Populations: The UK Biobank is a large, multi-site cohort that recruited participants between 40-69 years of age who were registered with a general practitioner of the UK National Health Service (NHS). Between 2006-2010, a total of 503,325 individuals were enrolled through 22 assessment centers in the UK. At enrollment, extensive data on demographics, ethnicity, education, lifestyle indicators, imaging of the body and brain, and disease-related outcomes were obtained through questionnaires, health records, and/or clinical evaluations. Blood samples were also collected at baseline for measurement of serum biomarkers that are either established disease risk factors or routinely measured as part of clinical evaluations. All study participants provided informed consent and the study was approved by the North West Multi-centre Research Ethics Committee. The present analyses with the UK Biobank were approved by the Institutional Review Board of USC Keck School of Medicine.

The Cleveland Clinic GeneBank study is a single site sample repository generated from ~10,000 consecutive patients undergoing elective diagnostic coronary angiography or elective cardiac computed tomographic angiography with extensive clinical and laboratory characterization and longitudinal observation (https://clinicaltrials.gov/ct2/show/NCT00590200). Subject recruitment occurred between 2001 and 2007. Ethnicity was self-reported and information regarding demographics, medical history, and medication use was obtained by patient interviews and confirmed by chart reviews. All clinical outcome data were verified by source documentation. CAD was defined as adjudicated diagnoses of stable or unstable angina, myocardial infarction (MI) (adjudicated definition based on defined electrocardiographic changes or elevated cardiac enzymes), angiographic evidence of ≥ 50% stenosis of one or more major epicardial vessel, and/or a history of known CAD (documented MI, CAD, or history of
revascularization). Plasma glycine levels at enrollment were quantified in 3,037 subjects of northern European ancestry in three batches (n=384, 882, and 1,771) by stable isotope dilution high performance liquid chromatography with online electrospray ionization tandem mass spectrometry (LC/MS)\textsuperscript{43}. The GeneBank cohort has been used previously for discovery and replication of novel genes and risk factors for atherosclerotic disease\textsuperscript{11,32,44-49}. Written informed consent was obtained from all participants prior to enrollment. The GeneBank study has been approved by the Institutional Review Board of the Cleveland Clinic and the present analyses were approved by the Institutional Review Board of USC Keck School of Medicine.

**Clinical Definitions in UK Biobank:** All subjects in the UK Biobank with serum glycine levels were included in the clinical analyses with CAD regardless of ancestry. CAD cases were defined as having been assigned ICD10 codes I21, I22, I23, I25.2, I24.0, I24.8, I24.9, I25.0, I25.1, I25.4, I25.8, or I25.9, which included doctor-diagnosed and self-reported ischemic heart disease (Data-Fields 6150 and 20002, respectively) on or before the date of enrollment (Data-Field 53). Time to incident CAD was derived by calculating the number of days from the date when the participant without a prior CAD diagnosis attended the baseline assessment (Date-Field 53) to the date on which the first ICD10 code for CAD was assigned to the subject (Date-Field 41262 and 41280). Type 2 diabetes (T2D) status was defined based on ICD10 code E11 and kidney function was based on estimated glomerular filtration rate (eGFR), as calculated using serum creatinine, age, sex, and self-reported ethnic background (Data-Field 21000)\textsuperscript{50}. Antihypertensive and lipid-lowering medication use were also based on self-reported data (Data-Fields 6153, 6177 and 20003).
Clinical Analyses in UK Biobank: Glycine levels were derived from nuclear magnetic resonance (NMR)-based metabolomics analyses that were carried out on serum samples at the baseline blood draw from a random subset of ~121,000 UK Biobank subjects. Serum glycine levels were categorized into quintiles and evaluated for association with risk of prevalent CAD by logistic regression. This analysis included 101,619 controls and 4,099 cases with CAD at the time of enrollment in the UK Biobank and the model was adjusted for age, sex, self-reported ethnicity, anti-hypertensive medications, lipid-lowering medications, eGFR, and systolic blood pressure. Cox proportional hazards regression models were used to evaluate quintiles of glycine levels with incident risk of CAD with adjustment for age, sex, self-reported ethnicity, anti-hypertensive medications, lipid-lowering medications, and T2D at baseline. Association of serum glycine levels with risk incident CAD was evaluated only among participants who were defined as controls in the prevalent CAD logistic regression analysis and for whom complete data were available (n=101,608). Incident CAD was defined based on participants who were assigned an ICD10 code for CAD after enrollment into UK Biobank and up to October 31, 2022 (maximum follow-up period was capped at 5,000 days). Participants assigned an ICD10 code for CAD within 14 days of enrollment into UK Biobank were excluded and those who died from causes not classified under ICD10 codes for CAD were censored at the time of death. Trend and quintile comparison P-values were calculated for both logistic and time-to-event analyses. All statistical analyses were carried out with SAS 9.4 (SAS Institute, Inc, Cary, NC).

GWAS for Circulating Glycine Levels in the Genebank and UK Biobank Cohorts:
Genome-wide genotyping in GeneBank was carried out with either the Affymetrix Genome-Wide Human Array 6.0 Chip (n=3,031) or the Illumina Infinium Global Screening Array-24 v2.0.
Prior to imputation, genomic coordinates of SNPs on each genotyping platform were first converted to GRCh37/hg19. Quality control steps included removal of duplicate SNPs as well as those with call rates <97%, minor allele frequencies (MAFs) <1%, and without chromosome and base pair position. Individuals with genotype call rates <90%, of African American ancestry, and outliers from PCA analysis were also excluded, resulting in 671,968 SNPs in 2,972 participants genotyped with the Affymetrix 6.0 Chip, and 539,533 SNPs in 1,624 participants genotyped with the GSA Chip. Imputation was carried out for unmeasured SNPs on the forward (+) strand using 1000 Genomes Project (Phase 3 release, v5) and Haplotype Reference Consortium (vr1.1 2016) as reference panels through the University of Michigan Imputation Server [https://imputationserver.sph.umich.edu](https://imputationserver.sph.umich.edu). After imputation, subjects with discordant sex were excluded and SNPs that had Hardy-Weinberg equilibrium P-values <0.0001, were duplicates or multiallelic, had imputation quality scores <0.3, and with MAFs <1% were removed. This resulted in 9,185,470 SNPs available for analysis in 3,037 GeneBank subjects with plasma glycine measurements. Circulating glycine levels were analyzed by linear regression with age, sex, and genotyping array as covariates using PLINK (v1.9) [http://www.cog-genomics.org/plink/1.9/](http://www.cog-genomics.org/plink/1.9/)

In UK Biobank, quality control of samples, DNA variants, and imputation were performed by the Wellcome Trust Centre for Human Genetics. Briefly, ~90 million SNPs imputed from the Haplotype Reference Consortium, UK10K, and 1000 Genomes imputation were available in UK Biobank. After filtering on autosomal SNPs with INFO scores >0.8 (directly from the UK Biobank) and with MAFs >1%, 9,560,226 variants were available in 117,152 UK Biobank subjects with serum glycine measurements. A GWAS analysis was performed with BOLT-LMM (v2.3.4) using a standard (infinitesimal) mixed model to correct for
population structure due to relatedness and ancestral heterogeneity, and with adjustment for age, sex, the first 20 principal components, and genotyping array. The genome-wide significance thresholds for GWAS analyses in the GeneBank and UK Biobank cohorts were set at P=5.0E-08.

**Meta-analysis for Circulating Glycine Levels:** A meta-analysis for circulating glycine levels was carried by combining the GWAS results generated in GeneBank and UK Biobank with publicly available summary statistics from 13 previously published datasets. In total, 230,947 multi-ancestry subjects from 13 datasets were included in the meta-analysis (Supplemental Table 2). GWAS summary statistics from Lotta et al. were first imputed with summary statistics imputation (SSimp) software using the European population from the 1000 Genomes Project (Phase 3 release, v5) as a reference panel for LD computation. After filtering on autosomal SNPs with MAF >1%, this imputation resulted in Z-scores and P-values for association of 7,667,668 SNPs with circulating glycine levels in the cohorts used by Lotta et al. We next harmonized the summary GWAS data from the remaining studies to also match the data in the GeneBank and UK Biobank cohorts and converted genomic coordinates for all datasets to GRCh37/hg19. A weighted Z-score meta-analysis was performed by combining the Z-score and P-value summary level data for SNPs that were common to at least two datasets (Supplemental Table 1), as implemented in METAL. The genome-wide threshold for significant associations in the GWAS meta-analyses was set at P=5.0E-08. Manhattan and quantile-quantile (QQ) plots were generated using the ‘qqman’ package. A locus was defined as novel if the lead SNP was in weak or no linkage disequilibrium (LD; r²<0.2) with genome-wide significant variants reported previously for circulating glycine levels. Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) program was used to define independent SNPs at glycine-associated loci as those that also yielded P<5.0E-08.
were at least 500kb away from the lead variant, and were in low LD ($r^2<0.2$) with other variants at the locus based on LD information from the 1000 Genomes Project. To determine whether loci identified for glycine levels were also associated with other clinical traits, CAD risk factors, and metabolites, a phenome-wide association study (PheWAS) was carried out using FUMA and publicly available resources, such as the GWAS Catalog (https://www.ebi.ac.uk/gwas/home), PhenoScanner⁶², and the UCSC Genome Browser (https://genome.ucsc.edu/). The significance threshold for PheWAS analyses was set at $P=5.0E-08$ with a LD cut off of $r^2\geq0.8$ for proxy SNPs.

**Genetic Risk Score Analyses.** Primary level data from the UK Biobank were used to generate weighted genetic risk scores (GRS) with all 61 loci for glycine levels. For each variant, the number of alleles associated with increased glycine was multiplied by its respective effect size obtained from the GWAS analysis in the UK Biobank and summed together across all variants to generate the GRS. Association between quintiles of weighted GRS and glycine levels was tested using linear regression, with adjustment for age, sex, the first 20 principal components, and genotyping array, in R (v4.2.0)⁶³.

**Mendelian Randomization (MR) Analyses.** To evaluate whether glycine levels were causally associated with risk of CAD, we used the results of our GWAS analysis in the UK Biobank (since we had primary metabolomics data with absolute concentrations of glycine levels available in this large dataset) and publicly available summary results from a recently published large-scale multi-ancestry GWAS meta-analysis for CAD². Effect sizes for three groups of genetic instruments for glycine levels as the exposure were taken from the GWAS results in the
UK Biobank. Group 1 included variants at all 61 glycine-associated loci, Group 2 only included variants at the seven loci that did not exhibit pleiotropic associations with other traits or metabolites, and Group 3 only included variants at the three non-pleiotropic loci that harbored genes involved in the glycine cleavage system. MR analyses were carried out using weighted median, inverse variance weighted, and MR Egger methods, as implemented in the “TwoSampleMR” package\textsuperscript{64,65} for R. MR analyses with the seven non-pleiotropic loci were also carried out for blood pressure, body mass index (BMI), glucose/insulin traits, circulating lipid levels, and risk of T2D using effect sizes from previously published GWAS meta-analyses\textsuperscript{38, 66-69}, as the outcomes.

**Animal Husbandry and Glycine Supplementation.** Animal studies were performed with approval and in accordance with the guidelines of the Institutional Animal Research Committees of the University of Southern California. All mice were housed in a temperature-controlled facility with a 12-h light/dark cycle and fed ad libitum with free access to water. Male and female \textit{ApoE}\textsuperscript{-/-} mice (stock number: 002052) were purchased from the Jackson Laboratories (Bar Harbor, Maine), bred in-house for this study, and maintained on a chow diet (PicoLab: #5053) until initiation of the glycine feeding studies. At 8-10 weeks of age, age-matched male and female \textit{ApoE}\textsuperscript{-/-} mice were placed on customized chow and amino acid defined diets containing either 0.3% glycine (Research Diets: A18011701, New Brunswick, NJ) or 2% glycine (Research Diets: A18011702, New Brunswick, NJ).

**Blood Measurements.** Following 16 weeks of glycine supplementation, blood was collected from the submandibular vein of mice after a 12hr overnight fast and mice were euthanized. One
week prior to euthanization, non-fasting blood was also obtained from a subset of mice two hours after the beginning of the feeding cycle. Plasma levels of glycine and other metabolites/amino acids were quantified by LC/MS, as described above for the GeneBank cohort. Plasma total cholesterol, high-density lipoprotein (HDL) cholesterol, triglyceride, glucose, and insulin levels were determined by enzymatic colorimetric assays, as described previously\textsuperscript{70,71}. Combined very low-density lipoprotein (VLDL) cholesterol and low-density lipoprotein (LDL) cholesterol levels were calculated by subtracting HDL cholesterol from total plasma cholesterol levels. Plasma insulin levels were measured in duplicate using Mouse Ultrasensitive Insulin ELISA kits (Alpco Inc: 80-INSMSU-E10, Salem, NH). Homeostasis modeling assessment insulin resistance (HOMA-IR) was calculated according to the formula: 
$\left[\text{glucose (mmol/L)} \times \text{insulin (mIU/ml)}\right]/22.5$\textsuperscript{72}. Inflammatory cytokines were measured using a multiplexed immunoassay kit (Meso Scale Discovery, K15048G-1, Rockville, MD) and complete blood count profiles were determined using the HEMAVET\textsuperscript{®} 950FS Multi-species Hematology System (Drew Scientific, Miami Lakes, FL).

**Aortic Lesion and En Face Analyses.** After euthanization, hearts were perfused through the left ventricle with approximately 15ml of phosphate-buffered saline (PBS) and characterized for atherosclerotic lesion formation, as described previously\textsuperscript{73}. The hearts were then removed, placed in 10% formalin solution, and transferred to 30% sterile sucrose for 48hrs before being embedded in Optimal Cutting Temperature (OCT) compound (Fisher Scientific, 23-730-571). Serial interrupted 10µm thick aortic cryosections were cut starting at the origins of the aortic valve leaflets. Every 8th section was stained with Oil Red O and hematoxylin for quantification of atherosclerotic lesion area. For each mouse, total aortic lesion size was determined in blinded
fashion by summing the lesion areas of 10 sections using Image-J software (NIH). *En face* analysis was carried out on a different set of euthanized mice after first perfusing the heart and aorta with 15ml of PBS, followed by a formal sucrose solution (4% paraformaldehyde/7.5% sucrose/10mM sodium phosphate buffer/2mM EDTA/20mM butylated hydroxytoluene) for 15mins, and rinsing again with 10ml of PBS. Aortas from the aortic arch to iliac bifurcation were removed carefully under a microscope (Richter Optica S6-RLT) and the surrounding adventitial fat tissue was dissected away. The aorta was then opened longitudinally from the aortic root to iliac bifurcation and pinned on a black rubber plate filled with PBS. The aortas were then incubated in 70% ethanol for 5mins, stained with Sudan-IV solution (5mg/ml Sudan-IV in 70% ethanol and 100% acetone) for 15mins, and de-stained with 80% ethanol for 3mins. Aortas were then briefly rinsed under running tap water to remove any residual ethanol, then submerged in PBS for image capture. Images were taken with a digital camera and Sudan-IV stained atherosclerotic lesion area was calculated using Image-J software. Lesion area along the aortic arch, descending aorta, and abdominal aorta was calculated as the percentage of total area. All image capture and quantitation for the *en face* analyses were done in a blinded fashion.

**Statistical Analyses.** Differences in measured variables between control and glycine-supplemented mice were determined by unpaired Student’s *t*-tests (PRISM v6.01, GraphPad Software, Boston, MA). Values are expressed as mean ± SE and differences and were considered statistically significant at P<0.05 or at the significance threshold after correcting for multiple comparisons.
References

7. Ding Y, Pedersen ER, Svingen GF, Helgeland O, Gregory JF, Loland KH, Meyer K, Tell GS, Ueland PM and Nygard OK. Methylenetetrahydrofolate dehydrogenase 1 polymorphisms modify the associations of plasma glycine and serine with risk of acute


Data availability

Individual level data used in the present study are available upon application to the UK Biobank (https://www.ukbiobank.ac.uk/). Summary statistics from the meta-analysis for glycine levels will be posted to a public repository. Summary statistics for glycine levels from all other datasets used in the present study are available through their respective publications. All other relevant data are available upon request from the authors.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported, in part, by NIH Grants R01HL133169 and R01HL148110. The GeneBank study was supported in part by NIH grants P01HL098055, P01HL076491, and R01HL103931. Mass Spectrometry instrumentation used for the GeneBank study was housed in a facility supported in part through a Shimadzu Center of Excellence award. We gratefully acknowledge the UK Biobank Resource for providing access to their data under Application Number 33307. The sponsors had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Disclosures

Z.W. and S.L.H. are named as coinventors on pending and issued patents held by the Cleveland Clinic relating to cardiovascular diagnostics and therapeutics and have the right to receive

The copyright holder for this preprint (which was not certified by peer review) is the author/lunder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license.
royalty payments for inventions or discoveries related to cardiovascular diagnostics or therapeutics from Cleveland Heart Lab, Quest Diagnostics, and Procter & Gamble Company. S.L.H. also reports having been paid as a consultant from Procter & Gamble Company and having received research funds from Procter & Gamble Company and Roche. All other authors report no conflicts.

Author contributions

Table 1. Association of Serum Glycine Levels with Prevalent CAD at Time of Enrollment in UK Biobank and Incident Risk of CAD Over 5,000 Days of Follow-up.

<table>
<thead>
<tr>
<th>Quintile for Serum Glycine Levels</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prevalent CAD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range (µM)</td>
<td>0.0006-116.0</td>
<td>116.1-139.4</td>
<td>139.5-164.3</td>
<td>164.4-204.9</td>
<td>205.0-749.1</td>
</tr>
<tr>
<td>Cases/ Controls</td>
<td>1,189/19,931</td>
<td>980/20,189</td>
<td>866/20,282</td>
<td>679/20,461</td>
<td>385/20,756</td>
</tr>
<tr>
<td>OR (95% CI)</td>
<td>1</td>
<td>0.91 (0.83-1.00)</td>
<td>0.91 (0.83-1.01)</td>
<td><strong>0.87 (0.78-0.96)</strong></td>
<td>*<strong>0.76 (0.67-0.87)</strong></td>
</tr>
<tr>
<td><strong>Incident CAD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range (µM)</td>
<td>0.0006-116.5</td>
<td>116.6-140.0</td>
<td>140.1-165.1</td>
<td>165.2-206.2</td>
<td>206.3-749.1</td>
</tr>
<tr>
<td>Cases/ Controls</td>
<td>1,990/18,297</td>
<td>1,783/18,606</td>
<td>1,511/18,791</td>
<td>1,322/18,995</td>
<td>895/19,418</td>
</tr>
<tr>
<td>HR (95% CI)</td>
<td>1</td>
<td>0.94 (0.88-1.01)</td>
<td>***0.87 (0.81-0.92)</td>
<td>***0.84 (0.78-0.90)</td>
<td>***0.70 (0.65-0.77)</td>
</tr>
</tbody>
</table>

Data are shown as OR or HR with 95% CIs for association of glycine quintiles with prevalent and incident CAD, respectively, where quintile 1 is the reference group. Model for prevalent CAD is adjusted for age, sex, ethnicity, anti-hypertensive medications, lipid-lowering medications, eGFR, and systolic blood pressure. Model for incident CAD is adjusted for age, sex, ethnicity, anti-hypertensive medications, lipid-lowering medications, and T2D at baseline. *P<0.05; **P <0.01; ***P<0.0001 for comparisons to quintile 1.
Table 2. 26 Novel Loci Identified for Circulating Glycine Levels.

<table>
<thead>
<tr>
<th>Lead SNP</th>
<th>Chr</th>
<th>Positionb</th>
<th>Nearest Genec</th>
<th>EA/NEA</th>
<th>EAF</th>
<th>N</th>
<th>Z-score</th>
<th>Directiond</th>
<th>P-value</th>
<th>P-het</th>
<th>Beta</th>
<th>SE</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs6695324</td>
<td>1</td>
<td>151,926,762</td>
<td>THEM4</td>
<td>A/T</td>
<td>0.69</td>
<td>222,135</td>
<td>-6.7</td>
<td>---------+---+----</td>
<td>2.2E-11</td>
<td>0.95</td>
<td>0.001</td>
<td>0.005</td>
<td>0.86</td>
</tr>
<tr>
<td>rs1823685</td>
<td>2</td>
<td>61,435,085</td>
<td>USP34</td>
<td>A/G</td>
<td>0.17</td>
<td>230,887</td>
<td>5.5</td>
<td>++++++++?+?+</td>
<td>3.6E-08</td>
<td>0.16</td>
<td>0.002</td>
<td>0.006</td>
<td>0.77</td>
</tr>
<tr>
<td>rs2943635</td>
<td>2</td>
<td>227,077,377</td>
<td>IRS1</td>
<td>T/C</td>
<td>0.69</td>
<td>230,735</td>
<td>-6.5</td>
<td>++++++++?--</td>
<td>1.1E-10</td>
<td>0.41</td>
<td>0.031</td>
<td>0.005</td>
<td>1.6E-10</td>
</tr>
<tr>
<td>rs6802898</td>
<td>3</td>
<td>12,391,207</td>
<td>PPARg</td>
<td>T/C</td>
<td>0.13</td>
<td>230,843</td>
<td>7.1</td>
<td>++++++++?+?+</td>
<td>1.3E-12</td>
<td>0.58</td>
<td>-0.005</td>
<td>0.007</td>
<td>0.51</td>
</tr>
<tr>
<td>rs71636623</td>
<td>3</td>
<td>49,383,260</td>
<td>AMT</td>
<td>CA/C</td>
<td>0.27</td>
<td>124,115</td>
<td>6.9</td>
<td>++?+?+?+++</td>
<td></td>
<td></td>
<td>6.7E-12</td>
<td>0.79</td>
<td>0.016</td>
</tr>
<tr>
<td>rs4431046</td>
<td>3</td>
<td>135,849,123</td>
<td>PCCB</td>
<td>A/C</td>
<td>0.23</td>
<td>230,947</td>
<td>6.7</td>
<td>++++++++??????</td>
<td>2.0E-11</td>
<td>0.72</td>
<td>-0.043</td>
<td>0.005</td>
<td>2.3E-15</td>
</tr>
<tr>
<td>rs12492498</td>
<td>3</td>
<td>152,008,390</td>
<td>MBNL1</td>
<td>A/G</td>
<td>0.69</td>
<td>230,735</td>
<td>6.3</td>
<td>++++++++?++</td>
<td>4.0E-10</td>
<td>0.34</td>
<td>0.007</td>
<td>0.005</td>
<td>0.11</td>
</tr>
<tr>
<td>rs9996922</td>
<td>4</td>
<td>77,437,749</td>
<td>SROOM3</td>
<td>A/G</td>
<td>0.73</td>
<td>222,234</td>
<td>6.1</td>
<td>++++-?+-??</td>
<td>1.3E-09</td>
<td>0.90</td>
<td>-0.021</td>
<td>0.006</td>
<td>7.7E-04</td>
</tr>
<tr>
<td>rs13107325</td>
<td>4</td>
<td>103,188,709</td>
<td>SLC39A8</td>
<td>T/C</td>
<td>0.07</td>
<td>212,417</td>
<td>-8.0</td>
<td>-?++?-?--</td>
<td>1.3E-15</td>
<td>0.43</td>
<td>-0.008</td>
<td>0.011</td>
<td>0.50</td>
</tr>
<tr>
<td>rs7766106</td>
<td>6</td>
<td>127,455,138</td>
<td>RSP03</td>
<td>T/C</td>
<td>0.48</td>
<td>230,947</td>
<td>-6.9</td>
<td>++++++++---</td>
<td>4.7E-12</td>
<td>0.34</td>
<td>0.012</td>
<td>0.004</td>
<td>4.7E-03</td>
</tr>
<tr>
<td>rs961329826</td>
<td>6</td>
<td>131,871,861</td>
<td>ARGI</td>
<td>G/GA</td>
<td>0.76</td>
<td>211,049</td>
<td>6.5</td>
<td>++++++++??</td>
<td>8.4E-11</td>
<td>0.37</td>
<td>-0.011</td>
<td>0.006</td>
<td>0.08</td>
</tr>
<tr>
<td>rs45446698</td>
<td>7</td>
<td>99,332,948</td>
<td>CYP3A7</td>
<td>T/G</td>
<td>0.96</td>
<td>212,663</td>
<td>-6.1</td>
<td>+?++?+-??</td>
<td>1.3E-09</td>
<td>0.01</td>
<td>0.021</td>
<td>0.013</td>
<td>0.10</td>
</tr>
<tr>
<td>rs111865019</td>
<td>7</td>
<td>106,812,246</td>
<td>DLD</td>
<td>A/G</td>
<td>0.73</td>
<td>230,947</td>
<td>-5.8</td>
<td>+--------++</td>
<td>8.3E-09</td>
<td>0.99</td>
<td>-0.026</td>
<td>0.005</td>
<td>1.7E-07</td>
</tr>
<tr>
<td>rs912322963</td>
<td>8</td>
<td>17,445,755</td>
<td>SLC7A2</td>
<td>A/G</td>
<td>0.02</td>
<td>208,491</td>
<td>7.0</td>
<td>+??+?+?+??+</td>
<td>2.1E-12</td>
<td>0.12</td>
<td>0.017</td>
<td>0.019</td>
<td>0.38</td>
</tr>
<tr>
<td>rs7828742</td>
<td>8</td>
<td>116,960,729</td>
<td>TRPS1</td>
<td>A/G</td>
<td>0.40</td>
<td>230,779</td>
<td>7.7</td>
<td>++++++++??</td>
<td>1.5E-14</td>
<td>0.20</td>
<td>-0.004</td>
<td>0.004</td>
<td>0.40</td>
</tr>
<tr>
<td>rs3802650</td>
<td>10</td>
<td>93,577,624</td>
<td>TNKS2</td>
<td>A/G</td>
<td>0.49</td>
<td>170,905</td>
<td>-6.9</td>
<td>++++++++?--</td>
<td>4.8E-12</td>
<td>0.44</td>
<td>-0.007</td>
<td>0.005</td>
<td>0.12</td>
</tr>
<tr>
<td>rs7078003</td>
<td>10</td>
<td>99,359,412</td>
<td>HOGA1</td>
<td>T/C</td>
<td>0.16</td>
<td>230,843</td>
<td>-6.7</td>
<td>-++++?+</td>
<td>2.1E-11</td>
<td>0.16</td>
<td>0.003</td>
<td>0.006</td>
<td>0.63</td>
</tr>
<tr>
<td>rs11191355</td>
<td>10</td>
<td>104,392,497</td>
<td>SUFU</td>
<td>T/C</td>
<td>0.82</td>
<td>230,947</td>
<td>-6.2</td>
<td>++++++++--</td>
<td>6.8E-10</td>
<td>0.74</td>
<td>0.015</td>
<td>0.005</td>
<td>0.01</td>
</tr>
<tr>
<td>rs11177732</td>
<td>12</td>
<td>69,981,111</td>
<td>CCT2</td>
<td>A/G</td>
<td>0.22</td>
<td>230,947</td>
<td>5.7</td>
<td>++++++++?++</td>
<td>1.2E-08</td>
<td>0.76</td>
<td>-0.002</td>
<td>0.005</td>
<td>0.70</td>
</tr>
<tr>
<td>rs41284816</td>
<td>13</td>
<td>50,655,989</td>
<td>EBPL</td>
<td>T/G</td>
<td>0.02</td>
<td>182,987</td>
<td>5.9</td>
<td>+??+++???+??+</td>
<td>3.3E-09</td>
<td>0.74</td>
<td>0.018</td>
<td>0.018</td>
<td>0.32</td>
</tr>
<tr>
<td>rs12889267</td>
<td>14</td>
<td>21,542,766</td>
<td>ARHGEF40</td>
<td>A/G</td>
<td>0.85</td>
<td>230,783</td>
<td>-5.5</td>
<td>++++++++?+?</td>
<td>4.4E-08</td>
<td>0.37</td>
<td>-0.027</td>
<td>0.007</td>
<td>4.6E-05</td>
</tr>
<tr>
<td>rs12882639</td>
<td>14</td>
<td>100,805,977</td>
<td>WARS1</td>
<td>T/C</td>
<td>0.53</td>
<td>230,735</td>
<td>-7.6</td>
<td>+++++??+?+</td>
<td>3.8E-14</td>
<td>0.35</td>
<td>-0.015</td>
<td>0.005</td>
<td>1.6E-03</td>
</tr>
<tr>
<td>rs61757601</td>
<td>16</td>
<td>72,101,525</td>
<td>DHX38</td>
<td>C/G</td>
<td>0.80</td>
<td>216,769</td>
<td>5.5</td>
<td>++++++++?+++</td>
<td>3.0E-08</td>
<td>0.21</td>
<td>-0.040</td>
<td>0.005</td>
<td>1.3E-14</td>
</tr>
<tr>
<td>rs2189338</td>
<td>17</td>
<td>5,326,341</td>
<td>RPA1N</td>
<td>T/C</td>
<td>0.40</td>
<td>218,825</td>
<td>-5.6</td>
<td>++++++++?--</td>
<td>2.3E-08</td>
<td>0.63</td>
<td>0.001</td>
<td>0.005</td>
<td>0.91</td>
</tr>
<tr>
<td>rs23409989</td>
<td>19</td>
<td>48,136,752</td>
<td>ZNF541</td>
<td>T/G</td>
<td>0.13</td>
<td>230,839</td>
<td>-5.7</td>
<td>++++++++?+</td>
<td>9.8E-09</td>
<td>0.10</td>
<td>0.017</td>
<td>0.006</td>
<td>4.6E-03</td>
</tr>
<tr>
<td>rs1800961</td>
<td>20</td>
<td>43,042,364</td>
<td>HNF4A</td>
<td>T/C</td>
<td>0.03</td>
<td>222,832</td>
<td>6.0</td>
<td>+??+++???+?+</td>
<td>2.4E-09</td>
<td>0.09</td>
<td>-0.011</td>
<td>0.014</td>
<td>0.41</td>
</tr>
</tbody>
</table>

aAssociation results for CAD are taken from a previously published GWAS meta-analysis.

bSNP base pair (bp) positions are given according to NCBI build 37 of the reference human genome sequence (hg19).

cDefined as gene for which lead SNP yielded expression quantitative trait locus in GTEx Project or gene closest to lead variant based on UCSC Genome Browser.
Order of cohorts for direction of effect: UK Biobank, SOL Hispanic Cohort, Tohoku Medical MegaBank Japan, GeneBank I, GeneBank II, GeneBank II, METSIM, Singapore Cohort, Canadian Longitudinal Study on Aging (CLSA) European, CLSA East Asians, CLSA South Asian, CLSA African, and meta-analysis from Lotta et al\textsuperscript{31}.\footnote{Order of cohorts for direction of effect: UK Biobank, SOL Hispanic Cohort, Tohoku Medical MegaBank Japan, GeneBank I, GeneBank II, GeneBank II, METSIM, Singapore Cohort, Canadian Longitudinal Study on Aging (CLSA) European, CLSA East Asians, CLSA South Asian, CLSA African, and meta-analysis from Lotta et al\textsuperscript{31}.}
Table 3. Effect of Dietary Glycine Supplementation on Fasting Cardiometabolic Traits and Aortic Lesion Formation in ApoE<sup>−/−</sup> Mice.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Females</th>
<th>Males</th>
<th>P-value</th>
<th>Females</th>
<th>Males</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control (n=27)</td>
<td>Glycine (n=32)</td>
<td>P-value</td>
<td>Control (n=31)</td>
<td>Glycine (n=36)</td>
<td>P-value</td>
</tr>
<tr>
<td>a,b Body weight (g)</td>
<td>22.4 ± 0.6</td>
<td>22.2 ± 0.3</td>
<td>0.79</td>
<td>31.4 ± 0.9</td>
<td>29.7 ± 0.5</td>
<td>0.08</td>
</tr>
<tr>
<td>Total cholesterol (mg/dL)</td>
<td>299 ± 18</td>
<td>299 ± 17</td>
<td>0.99</td>
<td>373 ± 27</td>
<td>353 ± 27</td>
<td>0.61</td>
</tr>
<tr>
<td>LDL/VLDL cholesterol (mg/dL)</td>
<td>271 ± 16</td>
<td>273 ± 16</td>
<td>0.92</td>
<td>339 ± 26</td>
<td>324 ± 26</td>
<td>0.69</td>
</tr>
<tr>
<td>HDL cholesterol (mg/dL)</td>
<td>28 ± 2</td>
<td>26 ± 1</td>
<td>0.18</td>
<td>34 ± 1</td>
<td>29 ± 2</td>
<td>0.03</td>
</tr>
<tr>
<td>Triglycerides (mg/dL)</td>
<td>43 ± 3</td>
<td>40 ± 2</td>
<td>0.40</td>
<td>58 ± 4</td>
<td>44 ± 3</td>
<td>0.008</td>
</tr>
<tr>
<td>Glucose (mg/dL)</td>
<td>128 ± 6</td>
<td>135 ± 6</td>
<td>0.35</td>
<td>130 ± 6</td>
<td>134 ± 7</td>
<td>0.69</td>
</tr>
<tr>
<td>Insulin (pg/mL)</td>
<td>453 ± 68</td>
<td>447 ± 85</td>
<td>0.96</td>
<td>523 ± 58</td>
<td>481 ± 88</td>
<td>0.70</td>
</tr>
<tr>
<td>d HOMA-IR</td>
<td>3.8 ± 0.5</td>
<td>3.9 ± 0.6</td>
<td>0.88</td>
<td>4.6 ± 0.4</td>
<td>4.1 ± 0.6</td>
<td>0.50</td>
</tr>
<tr>
<td>e,f Aortic lesions (µm&lt;sup&gt;2&lt;/sup&gt;/section)</td>
<td>118,060 ± 18,477</td>
<td>144,598 ± 21,357</td>
<td>0.36</td>
<td>44,414 ± 5,967</td>
<td>48,423 ± 8,097</td>
<td>0.70</td>
</tr>
<tr>
<td>g En Face lesions (%)</td>
<td>2.0 ± 0.4</td>
<td>1.3 ± 0.3</td>
<td>0.14</td>
<td>0.9 ± 0.3</td>
<td>0.5 ± 0.1</td>
<td>0.26</td>
</tr>
</tbody>
</table>

<sup>a</sup>Body weight data in females on the control and glycine diets are from 25 and 30 mice, respectively.

<sup>b</sup>Body weight data in males on the control and glycine diets are from 28 and 33 mice, respectively.

<sup>c</sup>Glucose levels in male mice on the control diet are from 30 animals.

<sup>d</sup>Calculated according to the formula: [glucose mg/dL x insulin (µIU/ml)]/405.

<sup>e</sup>Aortic lesion data in females on the control and glycine diets are from 16 and 18 mice, respectively.

<sup>f</sup>Aortic lesion data in males on the control and glycine diets are from 22 and 27 mice, respectively.

<sup>g</sup>En face lesion data are from 8 mice of each sex on the control diet and 9 mice of each sex on the glycine diet.
<table>
<thead>
<tr>
<th>Metabolite (µM)</th>
<th>Control (n=27)</th>
<th>Glycine (n=32)</th>
<th>P-value</th>
<th>Control (n=30)</th>
<th>Glycine (n=36)</th>
<th>P-value</th>
<th>Control (n=9)</th>
<th>Glycine (n=14)</th>
<th>P-value</th>
<th>Control (n=10)</th>
<th>Glycine (n=12)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choline</td>
<td>22 ± 1</td>
<td>21 ± 1</td>
<td>0.47</td>
<td>22 ± 1</td>
<td>22 ± 1</td>
<td>0.73</td>
<td>12.6 ± 1.1</td>
<td>12.2 ± 0.8</td>
<td>0.76</td>
<td>13.1 ± 1.5</td>
<td>11.3 ± 1.1</td>
<td>0.32</td>
</tr>
<tr>
<td>TMAO</td>
<td>2.8 ± 0.4</td>
<td>2.9 ± 0.3</td>
<td>0.87</td>
<td>1.0 ± 0.1</td>
<td>1.1 ± 0.1</td>
<td>0.69</td>
<td>5.5 ± 0.5</td>
<td>9.5 ± 1.6</td>
<td>0.06</td>
<td>2.2 ± 0.3</td>
<td>2.6 ± 0.2</td>
<td>0.35</td>
</tr>
<tr>
<td>Betaine</td>
<td>49 ± 3</td>
<td>50 ± 3</td>
<td>0.88</td>
<td>43 ± 3</td>
<td>45 ± 3</td>
<td>0.73</td>
<td>46 ± 3</td>
<td>47 ± 4</td>
<td>0.77</td>
<td>37 ± 5</td>
<td>23 ± 1</td>
<td>0.01</td>
</tr>
<tr>
<td>Dimethylglycine</td>
<td>8.8 ± 0.3</td>
<td>9.1 ± 0.4</td>
<td>0.52</td>
<td>6.4 ± 0.3</td>
<td>6.2 ± 0.2</td>
<td>0.60</td>
<td>4.1 ± 0.2</td>
<td>4.5 ± 0.4</td>
<td>0.40</td>
<td>2.5 ± 0.2</td>
<td>1.9 ± 0.1</td>
<td>0.02</td>
</tr>
<tr>
<td>Glycine</td>
<td>484 ± 19</td>
<td>536 ± 18</td>
<td>0.053</td>
<td>538 ± 14</td>
<td>639 ± 16</td>
<td>1.8E-05</td>
<td>472 ± 27</td>
<td>926 ± 92</td>
<td>9.6E-04</td>
<td>485 ± 33</td>
<td>1230 ± 103</td>
<td>3.5E-06</td>
</tr>
<tr>
<td>^Citrulline</td>
<td>91 ± 2</td>
<td>94 ± 5</td>
<td>0.57</td>
<td>77 ± 2</td>
<td>68 ± 2</td>
<td>3.4E-04</td>
<td>109 ± 5</td>
<td>102 ± 4</td>
<td>0.30</td>
<td>88 ± 3</td>
<td>75 ± 4</td>
<td>0.02</td>
</tr>
<tr>
<td>Arginine</td>
<td>65 ± 4</td>
<td>73 ± 3</td>
<td>0.10</td>
<td>60 ± 2</td>
<td>61 ± 3</td>
<td>0.75</td>
<td>76 ± 8</td>
<td>72 ± 5</td>
<td>0.63</td>
<td>51 ± 5</td>
<td>54 ± 5</td>
<td>0.69</td>
</tr>
<tr>
<td>Creatinine</td>
<td>12.1 ± 0.3</td>
<td>12.8 ± 0.3</td>
<td>0.17</td>
<td>11.4 ± 0.3</td>
<td>11.6 ± 0.4</td>
<td>0.60</td>
<td>9.5 ± 0.4</td>
<td>9.2 ± 0.3</td>
<td>0.56</td>
<td>10.6 ± 1.5</td>
<td>8.9 ± 1.4</td>
<td>0.40</td>
</tr>
<tr>
<td>Ornithine</td>
<td>139 ± 10</td>
<td>122 ± 6</td>
<td>0.16</td>
<td>115 ± 5</td>
<td>116 ± 7</td>
<td>0.88</td>
<td>119 ± 12</td>
<td>105 ± 10</td>
<td>0.38</td>
<td>158 ± 20</td>
<td>104 ± 10</td>
<td>0.02</td>
</tr>
<tr>
<td>Lysine</td>
<td>128 ± 4</td>
<td>135 ± 4</td>
<td>0.24</td>
<td>106 ± 4</td>
<td>106 ± 44</td>
<td>0.98</td>
<td>189 ± 15</td>
<td>194 ± 18</td>
<td>0.84</td>
<td>170 ± 12</td>
<td>139 ± 8</td>
<td>0.03</td>
</tr>
<tr>
<td>Histidine</td>
<td>131 ± 4</td>
<td>133 ± 3</td>
<td>0.78</td>
<td>126 ± 2</td>
<td>126 ± 4</td>
<td>0.88</td>
<td>139 ± 8</td>
<td>131 ± 11</td>
<td>0.60</td>
<td>156 ± 12</td>
<td>105 ± 6</td>
<td>1.0E-03</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>119 ± 5</td>
<td>120 ± 4</td>
<td>0.87</td>
<td>79 ± 3</td>
<td>84 ± 4</td>
<td>0.32</td>
<td>156 ± 17</td>
<td>144 ± 15</td>
<td>0.63</td>
<td>114 ± 11</td>
<td>85 ± 7</td>
<td>0.03</td>
</tr>
<tr>
<td>Serine</td>
<td>115 ± 6</td>
<td>125 ± 7</td>
<td>0.32</td>
<td>112 ± 4</td>
<td>127 ± 6</td>
<td>0.06</td>
<td>186 ± 13</td>
<td>218 ± 21</td>
<td>0.26</td>
<td>178 ± 23</td>
<td>167 ± 11</td>
<td>0.67</td>
</tr>
<tr>
<td>Proline</td>
<td>150 ± 7</td>
<td>151 ± 4</td>
<td>0.91</td>
<td>129 ± 4</td>
<td>141 ± 8</td>
<td>0.21</td>
<td>241 ± 29</td>
<td>275 ±51</td>
<td>0.63</td>
<td>293 ± 53</td>
<td>202 ± 33</td>
<td>0.14</td>
</tr>
<tr>
<td>Valine</td>
<td>104 ± 4</td>
<td>101 ± 3</td>
<td>0.62</td>
<td>102 ± 4</td>
<td>104 ± 4</td>
<td>0.79</td>
<td>136 ± 8</td>
<td>130 ± 11</td>
<td>0.70</td>
<td>147 ± 11</td>
<td>108 ± 6</td>
<td>4.0E-03</td>
</tr>
</tbody>
</table>
Metabolites were measured by mass spectroscopy in plasmas obtained after either a 4 hour fast (fasting) or after two hours into the feeding cycle (non-fasting). P-values between control and glycine fed groups, separately in males and females and in fasted and non-fasting groups, were derived from 2-sided unpaired t-tests. Differences at P<0.05 are highlighted in bold. TMAO, trimethylamine N-oxide.
Figure 1. Association of Serum Glycine Levels with Risk of CAD in the UK Biobank.
Individuals in the highest quintile of glycine levels had significantly reduced risk of prevalent CAD (OR=0.76, 95% CI 0.67-0.87; P<0.0001) (A) and incident CAD (HR=0.70, 95% CI 0.65-0.77; P<0.0001) (B) compared to individuals in the first quintile. P-value for trend for association with risk of prevalent CAD (A) and log-rank P-value for association with incident risk of CAD (B) across quintiles are also shown, including the number of subjects at risk of incident CAD in each quintile at baseline and the indicated follow-up time (bottom panel in B).
Figure 2. Multi-ancestry GWAS Meta-analysis and Genetic Risk Score (GRS) Analysis for Circulating Glycine Levels. (A) Manhattan plot shows 61 loci significantly associated with circulating glycine levels in 230,947 subjects. Novel (26) and known (35) loci are indicated by red and green dots, respectively. The seven non-pleiotropic loci are indicated by yellow dots, of which two loci on chromosomes 3 and 16 were also novel (purple arrow heads). Genome-wide thresholds for significant (P=5.0E-08) and suggestive (P=5.0E-06) association are indicated by dashed gray lines. P-values are truncated at -log10 (P-value)=40. (B) Serum glycine levels are increased as a function of quintiles of a weighted GRS constructed with the number of glycine-raising alleles carried by individuals in the UK.
Biobank for the 61 loci identified in the meta-analysis (n=23,283/quintile; total n=116,412). Mean glycine levels are shown for each quintile (P-trend=1.57E-4).
Figure 3. Results of MR Analyses to Evaluate Causal Association of Circulating Glycine Levels with Risk of CAD. Effect sizes of lead variants for circulating glycine levels identified in the meta-analysis (x-axis) are
plotted against effect sizes for risk of CAD based on previously published summary statistics (y-axis). Slopes of the regressions are represented by the colored lines and derived from tests of MR by inverse variance weighted (light blue), weighted median (dark blue), or MR Egger (green) methods for all 61 glycine-associated loci (A), the seven non-pleiotropic loci (B), and the three non-pleiotropic loci harboring genes involved in the glycine cleavage system (C).
Figure 4. Effect of Glycine Supplementation on Plasma Cardiometabolic Traits and Atherosclerosis Development in ApoE−/− Mice. Compared to the 0.3% glycine diet (Control), mice fed the 2% glycine content diet (Glycine) had significantly higher fasting (A) and non-fasting (B) glycine levels, particularly among male...
mice. After 16 weeks of glycine supplementation, there were no differences in body weight (C) or fasting plasma levels of total cholesterol (D) and LDL/VLDL (E), whereas fasting levels of HDL cholesterol (F) and triglycerides (G) were decreased. There were also no differences with respect to metabolic traits, including fasting glucose (H) and insulin (I) levels, or HOMA-IR (J). Glycine supplementation did not affect atherosclerotic lesion formation assessed through serial cryosections at the aortic arch (K) or along the entire aorta by *en face* analysis (L). Representative sections of aortic lesions (K) and *en face* aortas stained for lipid content (L) are shown from female mice in the control and glycine-fed groups. Data are represented as mean ± SE. *P*-values are derived from t-tests carried out between control and glycine fed groups separately in males and females. *P*<0.05; **P*<0.01; ***P*<0.001; ****P*<0.0001. N=25-36 for data in panels A and C-K; n=9-14 for data in panel B; and n=8-9 for data in panel L.