Supplemental Material

Data S1.

Detailed Description of Cohorts

GeneBank Study: The Cleveland Clinic GeneBank study is a single site sample repository generated from consecutive patients undergoing elective diagnostic coronary angiography or elective cardiac computed tomographic angiography with extensive clinical and laboratory characterization and longitudinal observation. Subject recruitment occurred between 2001 and 2006. 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. Coronary artery disease (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 $\geq$ 50% stenosis of one or more major epicardial vessel, and/or a history of known CAD (documented MI, CAD, or history of revascularization). The GeneBank Study has been used previously for discovery and replication of novel genes and risk factors for atherosclerotic disease[1-4]. Plasma glycine levels were measured in blood samples obtained upon entry into GeneBank. Genome-wide genotyping was carried out on 3031 GeneBank subjects of European ancestry using the Affymetrix Genome-Wide Human Array 6.0 SNP chip. After conversion of genomic coordinates to GRCh37/hg19, exclusion of SNPs with duplicates, call rates <97%, minor allele frequencies (MAFs) <1%, and without chromosome and base pair position, and exclusion of 44 subjects with genotype call rates <90%, 642,766 were available for imputation in 2972 participants. Imputation was carried out on the forward (+)

strand using the University of Michigan Imputation Server (https://imputationserver.sph.umich.edu) and data from the 1000 Genomes Project (Phase 3, Version 5). Application of the same quality control filters described above to the 46,180,700 imputed SNPs, with the addition of excluding SNPs with Hardy-Weinberg equilibrium p-values <0.0001 and imputation Rsq scores <0.3, resulted in 8,986,545 autosomal SNPs that were available for analysis in 1276 GeneBank subjects for whom plasma glycine levels were also available. All patients provided written informed consent prior to being enrolled in GeneBank and the study was approved by the Institutional Review Board of the Cleveland Clinic.

FINRISK: FINRISK (FR) surveys are cross-sectional, population-based studies conducted every five years since 1972 to monitor risk of chronic diseases. For each survey, a representative random sample was selected from 25- to 74-year-old inhabitants of different regions in Finland. The survey included a questionnaire and a clinical examination, at which a blood sample was drawn, with linkage to national registries of cardiovascular disease and other health outcomes. The study protocol has been described elsewhere[5]. Study participants were followed up through December 31, 2012. Eligible individuals from FINRISK surveys conducted in 1992, 1997, 2002, and 2007 (total n=27 838) were genotyped in three separate batches and analyzed separately to avoid batch effects, followed by a meta-analysis for glycine levels as described previously[6]. Genome-wide genotyping was carried out on an Illumina core-exome chip. After quality controls, including SNP call rates ≥ 95%, minor allele frequencies (MAFs) ≥ 1%, and sample call rates ≥ 95%, identity-by-descent (IBD) ≤ 0.1, without sex mismatches, duplicates, and heterozygosity outliers by eye from distribution, 273,113 SNPs was available for imputation. IMPUTE2 was used for imputation based on 1000 Genomes Project March 2012 version.

Further exclusions included p for Hardy–Weinberg equilibrium ≤ $1.0x10^{-6}$ and imputation info ≤ 0.4[6].

Cardiovascular Risk in Young Finns Study (YFS): The Cardiovascular Risk in Young Finns Study (YFS) is a population based prospective cohort study. It was conducted at five medical schools in Finland (Turku, Helsinki, Kuopio, Tampere and Oulu) with the aim of studying the levels of cardiovascular risk factors in children and adolescents in different parts of the country. The latest follow-up was conducted in 2007 at which serum samples were used for metabolomics analyses. The study and data collection protocols have been described in detail previously[7]. Genome-wide SNP data were generated from a custom Illumina BeadChip containing 670,000 SNPs and CNV probes. The custom content on the custom 670K array replaced some poor performing SNPs on the Human610 BeadChip and added more CNV content, and includes 546,677 SNPs passing QC from 594,210 SNPs on the chip. The custom 670K chip shares 562,643 SNPs in common with the Illumina Human610 BeadChip. Genotypes were called using Illumina's clustering algorithm. A total of 2,556 samples were genotyped. After initial clustering, we removed 2 subjects for poor call rates (CR<0.90), and 54 samples failed subsequent QC filters (i.e., duplicated samples, heterozygosity, low call rate, or custom SNP fingerprint genotype discrepancy). The following filters were then applied to the remaining data: MAF 0.01, GENO 0.05, MIND 0.05, and HWE $1\times10-6$. Three individuals were removed for low genotyping (MIND>0.05), 11,766 markers were excluded based on HWE test ($P\leq1\times10^{-6}$), 7,746 SNPs failed missingness test (GENO>0.05), 34,596 SNPs failed frequency test (MAF<0.01), and one individual failed gender check. A final list of 546,677 SNPs passed QC and allele frequency filters[8]. IMPUTE2 was used for imputation based on 1000 Genomes

Project March 2012 version. Further exclusions included p for Hardy–Weinberg equilibrium ≤ $1.0x10^{-6}$ and imputation info ≤ 0.4[6].

Northern Finland Birth Cohort (NFBC): The Northern Finland Birth Cohorts were initiated 20 years apart in 1966 (NFBC66) and 1986 (NFBC86) to examine risk factors involved in pre-term birth and intrauterine growth retardation, and the consequences of these early adverse outcomes on subsequent morbidity and mortality, as described in detail previously[9]. Mothers living in the two northern-most provinces of Finland (Oulu and Lapland) were invited to participate if they had expected delivery dates during 1966 or 186. Individuals still living in the Helsinki area or Northern Finland were asked at age 31 to participate in a detailed biological and medical examination as well as a questionnaire. GWAS analyses for circulating glycine levels, as measured by NMR, were carried out in 4,483 and 3,112 from the NFBC66 and 1986 NFBC86 studies, respectively. Genomic DNA was extracted from whole blood using standard methods and samples were genotyped on the Illumina Infinium 370cnvDuo array at the Broad Institute Biological Sample Repository. All individuals in the study were genotyped with call rates >95%. Individuals with discrepancy between their reported sex and the sex determined from the X chromosome were excluded from analysis. The identity-by-descent (IBD) analysis option of PLINK45 was used to determine possible relatedness among sample subjects and identify sample duplications and sample contamination (the latter identified as individuals who seemed to be related to nearly everyone in the sample). If the sample duplication issue could not be resolved by external means, both samples were excluded. All apparently contaminated samples were also excluded. For pairs of individuals identified to be related at the level of half-sibs or closer in the IBD analysis, the subject with less complete genotyping was excluded. Variants were excluded from the analysis if the call rate in the final sample was <95%, if the P value from a test of

Hardy-Weinberg Equilibrium (HWE) was <0.0001, or if the MAF was <1%[10]. This resulted in 335,118 SNPs that were available for imputation. IMPUTE2 was used for imputation based on 1000 Genomes Project March 2012 version, with further exclusions for p for Hardy–Weinberg equilibrium $\leq 1.0x10^{-6}$ and imputation info ≤ 0.4[6].

The Metabolic Syndrome in Men (METSIM) Study. METSIM is a population-based study that recruited 10,197 Finnish men from the city of Kuopio in Eastern Finland between 2005-2010. The aims of METSIM are to investigate nongenetic and genetic factors associated with the risk of type 2 diabetes and cardiovascular disease, and with cardiovascular risk factors[11]. The protocol included a detailed phenotyping of the participants, an oral glucose tolerance test, fasting laboratory measurements, including proton NMR measurements, mass spectrometry metabolomics, as well as adipose tissue biopsies and stool samples in a subset of participants. Participants were genotyped on the Human OmniExpresss-12v1_C BeadChip (OmniExpress) and Infinium HumanExome-12 v1.0 BeadChip (Exome Chip) platforms. Quality controls included sample-level controls for sex and relatedness confirmation, sample duplication, and detection of sample genetic ancestry outliers using principal component analysis. Based on these quality control measures, 14 samples with sex chromosome anomalies, 18 with evidence of participant duplication, 12 population outliers, and 9 samples with non-Mendelian inheritance inconsistencies were removed. In addition, one individual from each of seven monozygotic twin pairs was removed. Variants with low mapping quality of probes to genome build GRCh37, low genotype completeness (<95% and<98% for the OmniExpress and ExomeChip, respectively), or Hardy-Weinberg equilibrium $P<10^6$ were also filtered out. OmniExpress variants passing quality control with SHAPEIT v2 were phased and imputed using minimac v2. For imputation, a

reference panel of 20.9M variants from the GoT2D study (including SNVs, indels and large deletions) based on the whole genome sequence of 2874 Europeans, including 1004 Finnish individual, was used. Following imputation, variants directly genotyped on the ExomeChip were added. In cases of common markers between imputed and genotyped variants, the directly genotyped calls from the ExomeChip were used. Subsequently, 16,607,533 variants with high imputation quality (i.e. minimac RSQ0.3) were carried forward for single-variant association testing. GWAS analyses for circulating glycine levels, as measured by NMR, were carried out in a subset of 8545 non-diabetic men as described previously[12]. The institutional review boards of the University of Kuopio and Kuopio University approved the METSIM study. Written informed consent was obtained from each participant.

Supplemental Table Legends (see Excel file):

Table S1. Results of 12 Loci Significantly Associated with Circulating Glycine Levels Stratified by Metabolomics Platform.

Table S2. Results of 12 Loci Significantly Associated with Circulating Glycine Levels Stratified by Sex.

Table S3. PheWAS Results for 12 Loci Significantly Associated with Circulating Glycine Levels.

Table S4. Association of 12 Glycine-associated Loci with CAD in CARDIoGRAM+C4D and UK Biobank.

Table S5. Individual and Joint SNP Effect Associations and Mendelian randomization analysis of Glycine-associated Loci with Traditionally CAD Risk Factors.

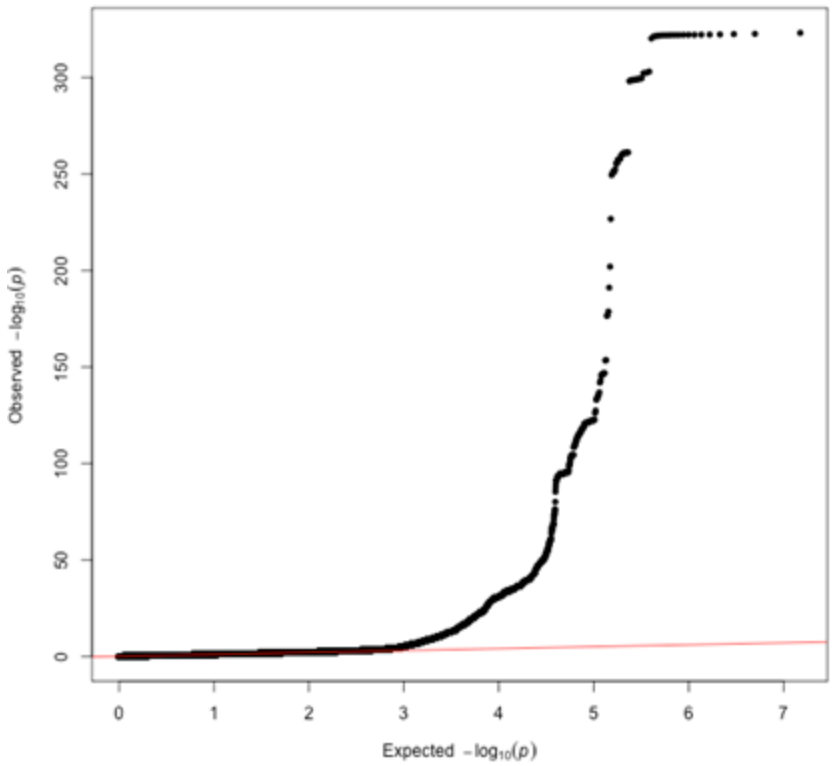
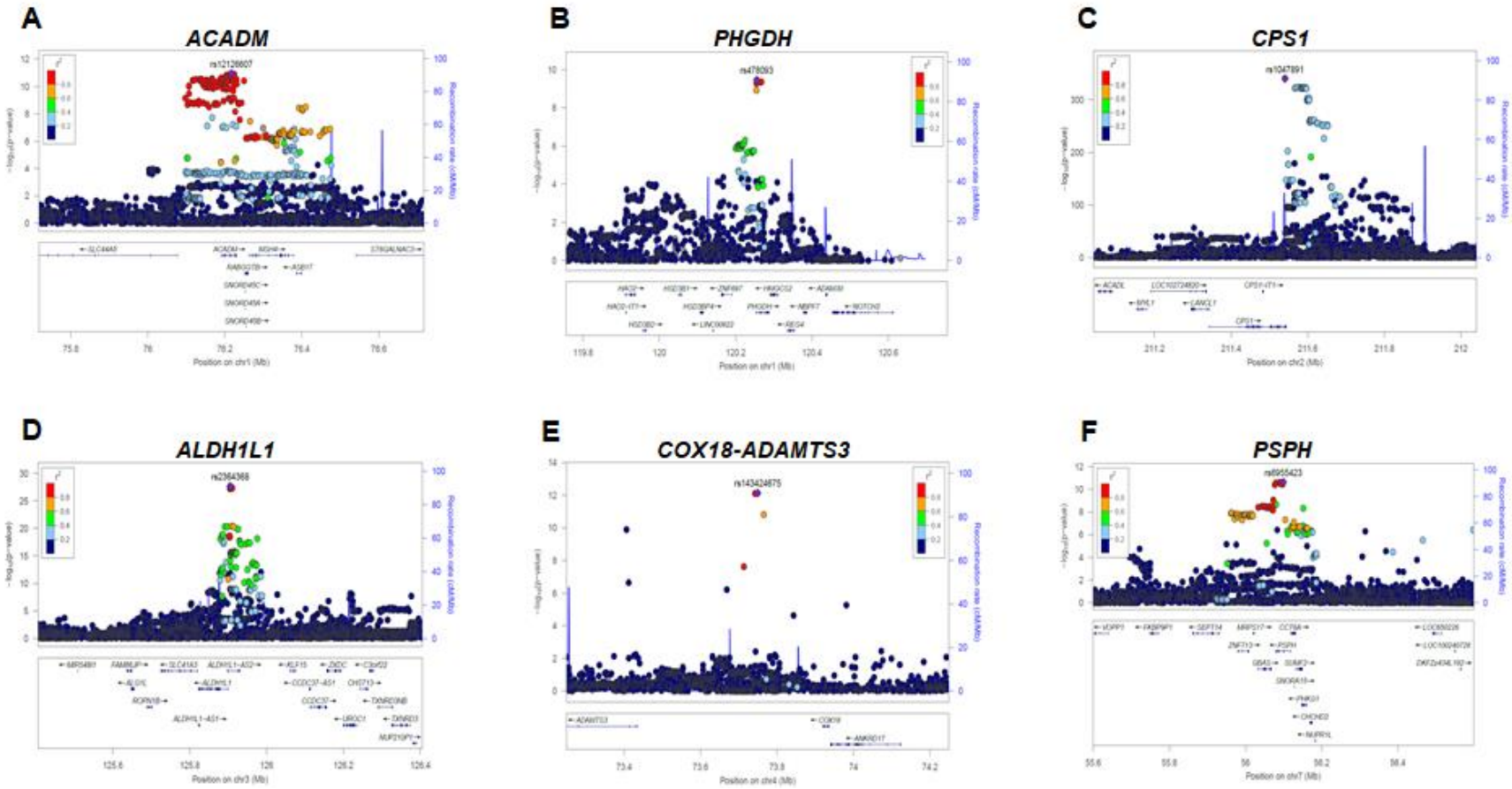


Figure S1. Quantile-quantile (Q-Q) plot of GWAS meta-analysis results for circulating glycine levels in 30,118 subjects. The observed versus the expected p-values from the meta-analyses for glycine levels are shown in the Q-Q plot. These analyses yielded a genomic inflation factor (λ) of 1.035, indicating that the GWAS meta-analyses were not confounded by underlying population stratification.


A
ACADM
rs12126607
B
PHGDH
rs478093
C
CPS1
rs1047891
D
ALDH1L1
rs2364368
E
COX18-ADAMTS3
rs143424675
F
PSPH
rs895423

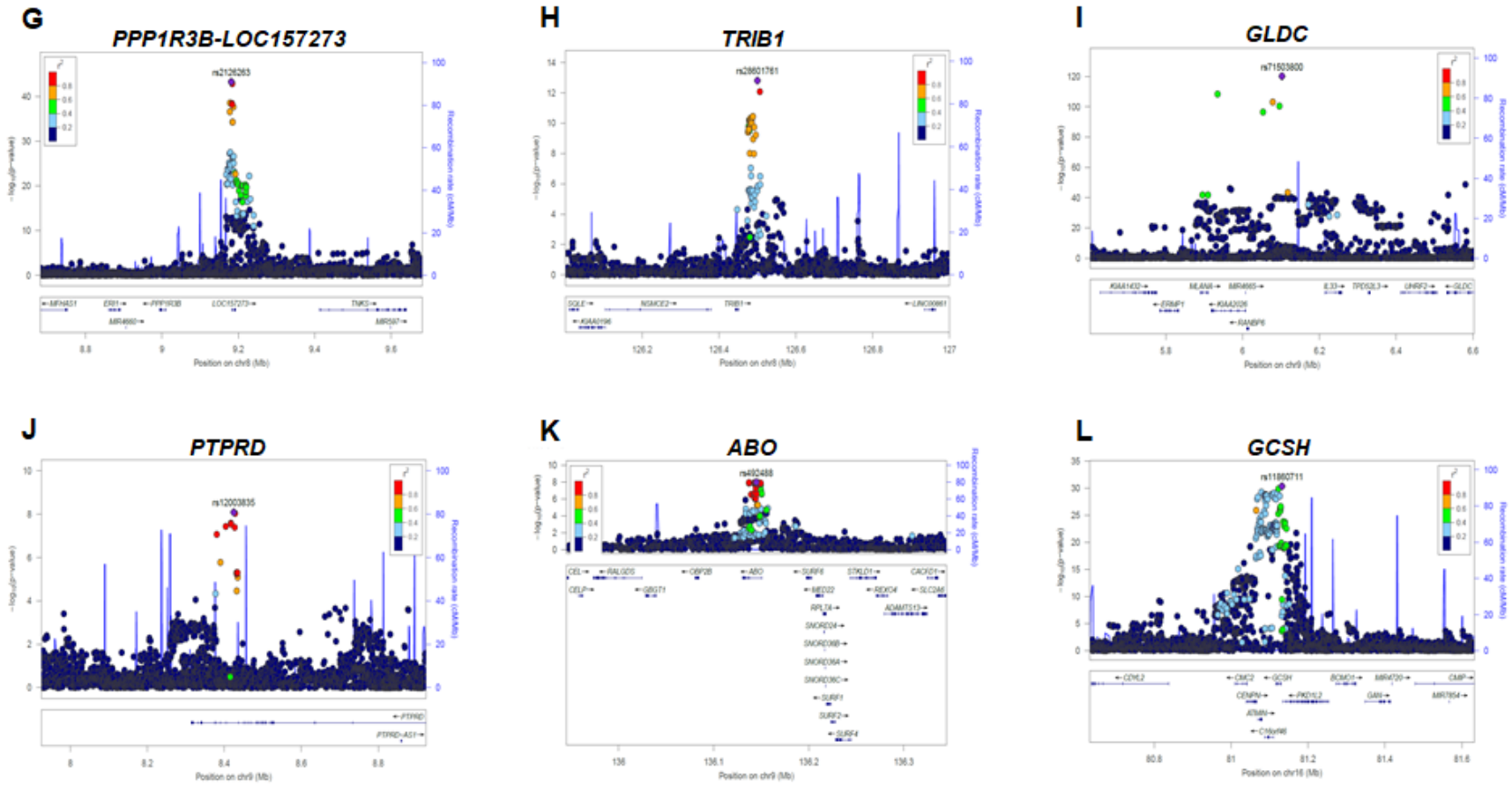


Figure S2. Twelve loci identified for circulating glycine associated levels. Regional plots for the *ACADM*, *PHGDH*, *CPS1*, *ALDH1L1*, *COX18-ADAMTS3*, *PSPH*, *PPP1R3B-LOC157273*, *TRIB1*, *GLDC*, *PTPRD*, *ABO*, and *GCSH* loci are shown in panels A-L. Each region is centered on the lead SNP (purple diamond) and the genes in the interval are indicated in the bottom panel. The degree of linkage disequilibrium (LD) between the lead SNP and other variants is shown as r^2 values according to the color-coded legend in the box.

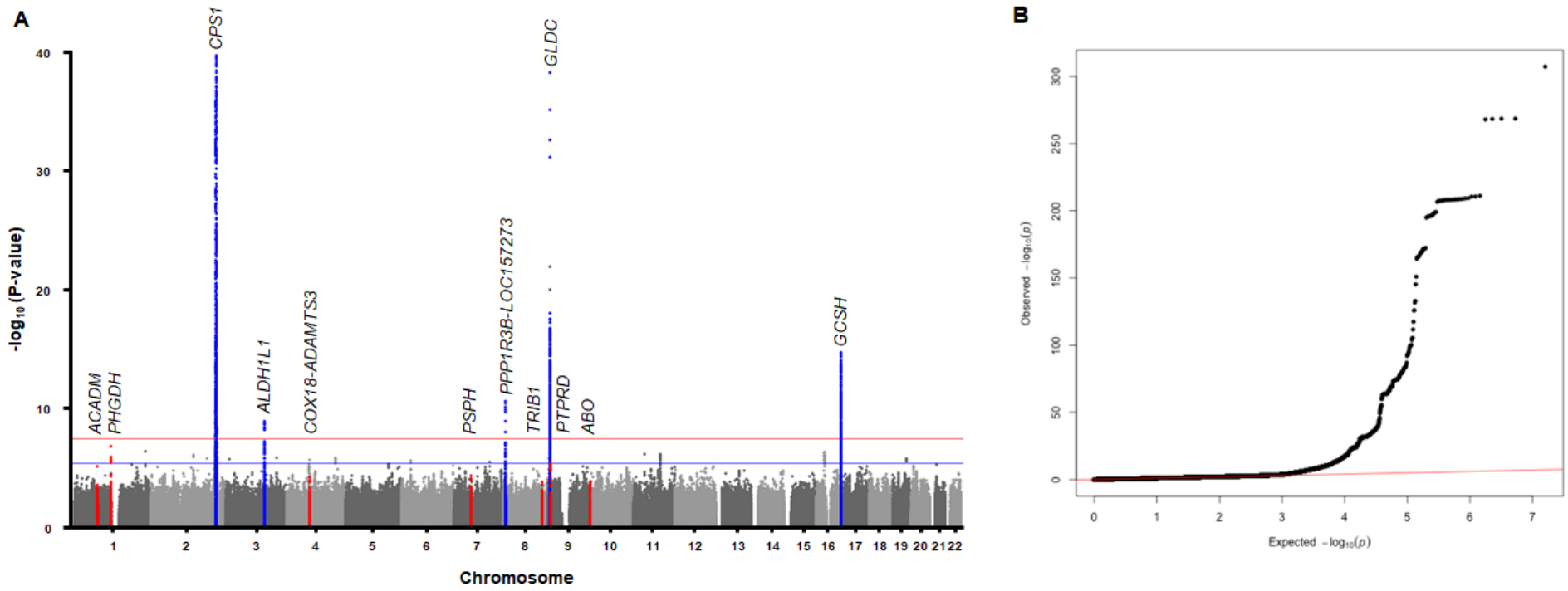


Figure S3. Results of GWAS meta-analysis for circulating glycine levels in women. (**A**) The Manhattan plot shows five previously identified loci significantly associated with circulating glycine levels (blue dots) in a stratified GWAS analysis with 10,886 women. Red dots indicate association signals for the seven novel identified in our meta-analysis with all 30,118 subjects, all of which were only suggestively associated in women. Genome-wide thresholds for significant ($P=5.0x10^{-8}$) and suggestive ($P=5.0x10^{-6}$) association are indicated by the horizontal red and dark blue lines, respectively. P-values are truncated at −log10 (P)=40. (**B**) The Q-Q plot shows the observed versus the expected p-values from the meta-analyses for glycine levels in women. These analyses yielded a genomic inflation factor (λ) of 1.002, indicating that the GWAS meta-analyses were not confounded by underlying population stratification.

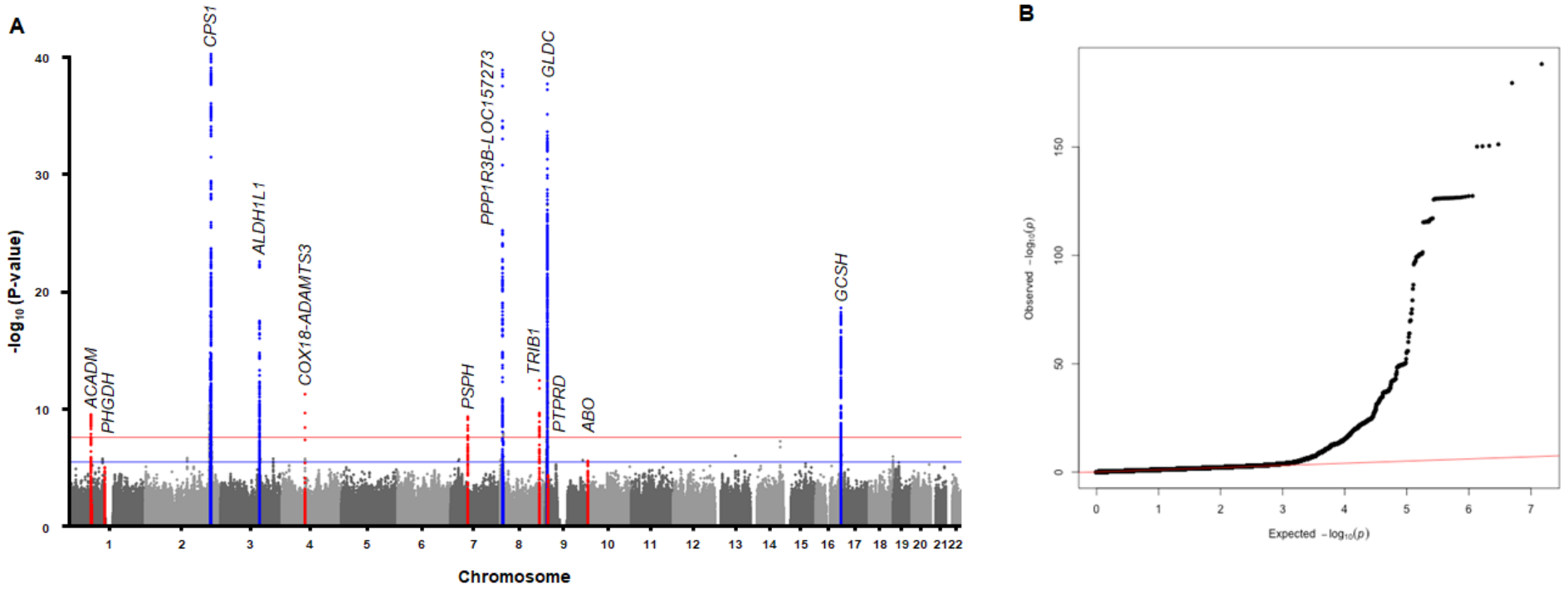


Figure S4. Results of GWAS meta-analysis for circulating glycine levels in men. (**A**) The Manhattan plot shows nine loci significantly associated with circulating glycine levels in a stratified GWAS analysis with 19,004 men. The five loci identified in previous studies are indicated by blue dots. The red dots indicate association signals at the seven novel identified by our meta-analysis with all 30,118 subjects, of which four were also significant in only men. Genome-wide thresholds for significant ($P=5.0x10^{-8}$) and suggestive ($P=5.0x10^{-6}$) association are indicated by the horizontal red and dark blue lines, respectively. P-values are truncated at −log10 (P)=40. (**B**) The Q-Q plot shows the observed versus the expected p-values from the meta-analyses for glycine levels in men. These analyses yielded a genomic inflation factor (λ) of 1.035, indicating that the GWAS meta-analyses were not confounded by underlying population stratification.

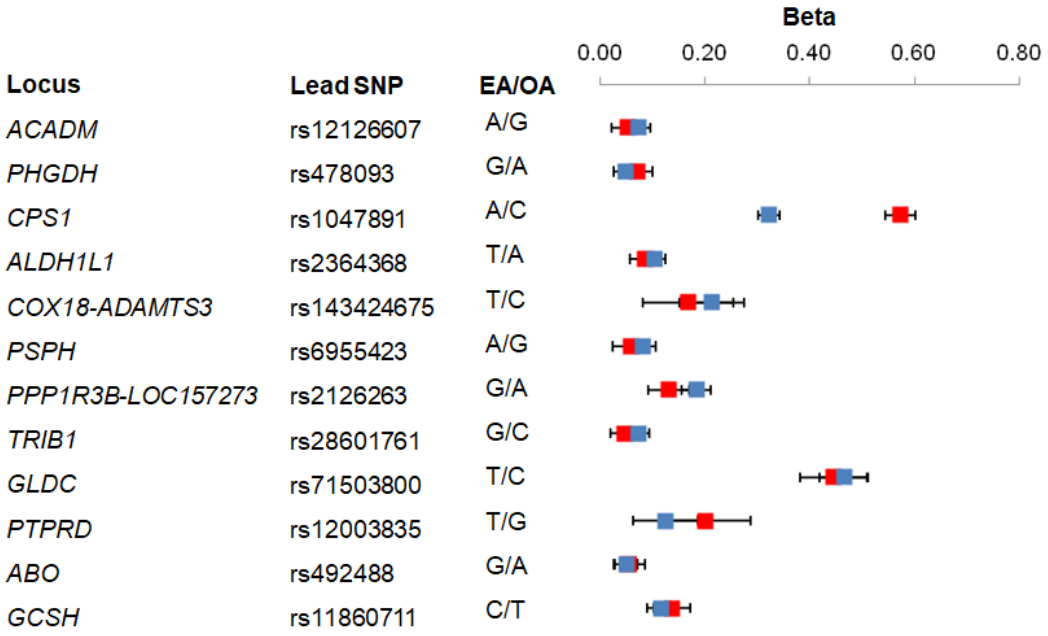


Figure S5. Sex-stratified results for 12 loci identified for circulating glycine levels. Effect sizes for the lead SNPs at the 12 loci identified for circulating glycine levels are shown in men (blue) and women (red) separately. With the exception of *CPS1*, which is associated with approximately two-fold higher glycine levels in women compared to men, effect sizes at the 11 other loci were similar in males and females. EA, effect allele; OA, other allele.

Supplemental References:

1. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, Feldstein AE, Britt EB, Fu X, Chung YM, Wu Y, Schauer P, Smith JD, Allayee H, Tang WH, DiDonato JA, Lusis AJ and Hazen SL. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*. 2011;472:57-63.
2. Koeth RA, Wang Z, Levison BS, Buffa JA, Org E, Sheehy BT, Britt EB, Fu X, Wu Y, Li L, Smith JD, DiDonato JA, Chen J, Li H, Wu GD, Lewis JD, Warrier M, Brown JM, Krauss RM, Tang WH, Bushman FD, Lusis AJ and Hazen SL. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat Med*. 2013;19:576-85.
3. Tang WH, Wang Z, Levison BS, Koeth RA, Britt EB, Fu X, Wu Y and Hazen SL. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med*. 2013;368:1575-84.
4. Hartiala JA, Tang WH, Wang Z, Crow AL, Stewart AF, Roberts R, McPherson R, Erdmann J, Willenborg C, Hazen SL and Allayee H. Genome-wide association study and targeted metabolomics identifies sex-specific association of CPS1 with coronary artery disease. *Nat Commun*. 2016;7:10558.
5. Vartiainen E, Laatikainen T, Peltonen M, Juolevi A, Mannisto S, Sundvall J, Jousilahti P, Salomaa V, Valsta L and Puska P. Thirty-five-year trends in cardiovascular risk factors in Finland. *Int J Epidemiol*. 2010;39:504-18.
6. Kettunen J, Demirkan A, Wurtz P, Draisma HH, Haller T, Rawal R, Vaarhorst A, Kangas AJ, Lyytikainen LP, Pirinen M, Pool R, Sarin AP, Soininen P, Tukiainen T, Wang Q, Tiainen M, Tynkkynen T, Amin N, Zeller T, Beekman M, Deelen J, van Dijk KW, Esko T, Hottenga JJ, van Leeuwen EM, Lehtimaki T, Mihailov E, Rose RJ, de Craen AJ, Gieger C, Kahonen M, Perola M, Blankenberg S, Savolainen MJ, Verhoeven A, Viikari J, Willemsen G, Boomsma DI, van Duijn CM, Eriksson J, Jula A, Jarvelin MR, Kaprio J, Metspalu A, Raitakari O, Salomaa V, Slagboom PE, Waldenberger M, Ripatti S and Ala-Korpela M. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun*. 2016;7:11122.
7. Raitakari OT, Juonala M, Ronnemaa T, Keltikangas-Jarvinen L, Rasanen L, Pietikainen M, Hutri-Kahonen N, Taittonen L, Jokinen E, Marniemi J, Jula A, Telama R, Kahonen M, Lehtimaki T, Akerblom HK and Viikari JS. Cohort profile: the cardiovascular risk in Young Finns Study. *Int J Epidemiol*. 2008;37:1220-6.
8. Smith EN, Chen W, Kahonen M, Kettunen J, Lehtimaki T, Peltonen L, Raitakari OT, Salem RM, Schork NJ, Shaw M, Srinivasan SR, Topol EJ, Viikari JS, Berenson GS and Murray SS. Longitudinal genome-wide association of cardiovascular disease risk factors in the Bogalusa heart study. *PLoS Genet*. 2010;6:e1001094.
9. Rantakallio P. The longitudinal study of the northern Finland birth cohort of 1966. *Paediatr Perinat Epidemiol*. 1988;2:59-88.
10. Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, Brodsky J, Jones CG, Zaitlen NA, Varilo T, Kaakinen M, Sovio U, Ruokonen A, Laitinen J, Jakkula E, Coin L, Hoggart C, Collins A, Turunen H, Gabriel S, Elliot P, McCarthy MI, Daly MJ, Jarvelin MR, Freimer NB and Peltonen L. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet*. 2009;41:35-46.

11. Laakso M, Kuusisto J, Stancakova A, Kuulasmaa T, Pajukanta P, Lusis AJ, Collins FS, Mohlke KL and Boehnke M. The Metabolic Syndrome in Men study: a resource for studies of metabolic and cardiovascular diseases. *J Lipid Res*. 2017;58:481-493.
12. Teslovich TM, Kim DS, Yin X, Stancakova A, Jackson AU, Wielscher M, Naj A, Perry JRB, Huyghe JR, Stringham HM, Davis JP, Raulerson CK, Welch RP, Fuchsberger C, Locke AE, Sim X, Chines PS, Narisu N, Kangas AJ, Soininen P, Genetics of Obesity-Related Liver Disease Consortium TAsDGCTDGR, Meta a, Ala-Korpela M, Gudnason V, Musani SK, Jarvelin MR, Schellenberg GD, Speliotes EK, Kuusisto J, Collins FS, Boehnke M, Laakso M and Mohlke KL. Identification of seven novel loci associated with amino acid levels using single-variant and gene-based tests in 8545 Finnish men from the METSIM study. *Hum Mol Genet*. 2018;27:1664-1674.