Genetic variants in *ALDH1L1* and *GLDC* influence the serine-to-glycine ratio in Hispanic children

Sergey A Krupenko,[1] Shelley A Cole,[2] Ruixue Hou,[1] Karin Haack,[2] Sandra Laston,[3,4] Nitesh R Mehta,[5,6] Anthony G Comuzzie,[7] Nancy F Butte,[5,6] and V Saroja Voruganti[1]

[1]Department of Nutrition and Nutrition Research Institute, University of North Carolina at Chapel Hill, Kannapolis, NC, USA; [2]Population Health Program, Texas Biomedical Research Institute, San Antonio, TX, USA; [3]Department of Human Genetics, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX, USA; [4]South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX, USA; [5]Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; [6]USDA/ARS Children Nutrition Research Center, Baylor College of Medicine, Houston, TX, USA; and [7]The Obesity Society, Silver Spring, MD, USA

ABSTRACT

Background: Glycine is a proteogenic amino acid that is required for numerous metabolic pathways, including purine, creatine, heme, and glutathione biosynthesis. Glycine formation from serine, catalyzed by serine hydroxy methyltransferase, is the major source of this amino acid in humans. Our previous studies in a mouse model have shown a crucial role for the 10-formyltetrahydrofolate dehydrogenase enzyme in serine-to-glycine conversion.

Objectives: We sought to determine the genomic influence on the serine-glycine ratio in 803 Hispanic children from 319 families of the Viva La Familia cohort.

Methods: We performed a genome-wide association analysis for plasma serine, glycine, and the serine-glycine ratio in Sequential Oligogenic Linkage Analysis Routines while accounting for relationships among family members.

Results: All 3 parameters were significantly heritable ($h^2 = 0.22$–0.78; $P < 0.004$). The strongest associations for the serine-glycine ratio were with single nucleotide polymorphisms (SNPs) in aldehyde dehydrogenase 1 family member L1 (*ALDH1L1*) and glycine decarboxylase (*GLDC*) and for glycine with *GLDC* ($P < 3.5 \times 10^{-8}$; effect sizes, 0.03–0.07). No significant associations were found for serine. We also conducted a targeted genetic analysis with *ALDH1L1* exonic SNPs and found significant associations between the serine-glycine ratio and rs2886059 ($\beta = 0.68$; SE, 0.25; $P = 0.006$) and rs3796191 ($\beta = 0.25$; SE, 0.08; $P = 0.003$) and between glycine and rs3796191 ($\beta = -0.08$; SE, 0.02; $P = 0.0004$). These exonic SNPs were further associated with metabolic disease risk factors, mainly adiposity measures ($P < 0.006$). Significant genetic and phenotypic correlations were found for glycine and the serine-glycine ratio with metabolic disease risk factors, including adiposity, insulin sensitivity, and inflammation-related phenotypes [estimate of genetic correlation $= -0.37$ to 0.35 ($P < 0.03$); estimate of phenotypic correlation $= -0.19$ to 0.13 ($P < 0.006$)]. The significant genetic correlations indicate shared genetic effects among glycine, the serine-glycine ratio, and adiposity and insulin sensitivity phenotypes.

Conclusions: Our study suggests that *ALDH1L1* and *GLDC* SNPs influence the serine-to-glycine ratio and metabolic disease risk. *Am J Clin Nutr* 2022;116:500–510.

Keywords: aldehyde dehydrogenase 1 family member l1, glycine decarboxylase, serine hydroxymethyltransferase, genome-wide association, heritability, adiposity, insulin sensitivity, inflammation, metabolic diseases

Introduction

Glycine, a common proteogenic amino acid, is required for numerous metabolic pathways, including purine, creatine, heme, and glutathione biosynthesis (1). Glycine conjugation with mitochondrial acyl-CoAs is also an essential metabolic pathway

The development of Sequential Oligogenic Linkage Analysis Routines (SOLAR) was supported by grant NIH-MH59490. The current study is supported by US NIH grant R01 DK126666. The Viva La Familia Study was supported by NIH-DK080457 and the USDA/Agricultural Research Service (corporative agreement 625–51000–053).

The contents of this publication do not necessarily reflect the views or policies of the USDA, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

Supplemental Figure 1 is available from the "Supplementary data" link in the online posting of the article and from the same link in the online table of contents at https://academic.oup.com/ajcn/.

Present address for RH: Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

Address correspondence to VSV (e-mail address: saroja@unc.edu).

Abbreviations used: ALDH1L1, 10-formyltetrahydrofolate dehydrogenase; *ALDH1L1*, aldehyde dehydrogenase 1 family member L1; BMIZ, BMI *z*-score; CRP, C-reactive protein; GCS, glycine cleavage system; GLDC, glycine decarboxylase; GWA, genome-wide association; MECR, mitochondrial trans-2-enoyl-coA reductase; MGA, measured genotype analysis; *PTPRU*, protein tyrosine phosphatase receptor type U; QUICKI, Quantitative Insulin-Sensitivity Check Index; SNP, single nucleotide polymorphism; SOLAR, Sequential Oligogenic Linkage Analysis Routines; SRSF4, serine- and arginine-rich splicing factor 4; VFS, Viva La Familia Study.

Received September 20, 2021. Accepted for publication April 21, 2022.

First published online April 23, 2022; doi: https://doi.org/10.1093/ajcn/nqac091.

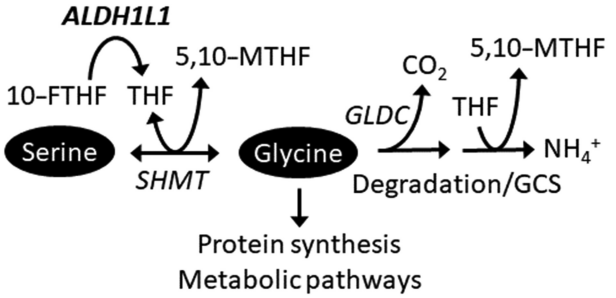

FIGURE 1 Schematic depicting glycine-relevant pathways. Glycine is synthesized from serine and is used for protein biosynthesis, conversion to other metabolites, or degraded through the GCS. Abbreviations: 5,10-MTHF, 5,10-methylene-tetrahydrofolate; 10-FTHF, 10-formyl-tetrahydrofolate; ALDH1L1, 10-formyltetrahydrofolate dehydrogenase; *ALDH1L1*, aldehyde dehydrogenase 1 family member L1; GCS, glycine cleavage system; GLDC, glycine decarboxylase; SHMT, serine hydroxy methyltransferase; THF, tetrahydrofolate.

to maintain adequate levels of free CoA (2), and several glycine conjugates, as well as glycine itself, have signaling functions (2, 3). Glycine deficiency is thus likely to affect numerous metabolic pathways (4–6). Overall, prolonged glycine insufficiency may have a negative effect on health (4). Indeed, studies have linked changes in serum glycine concentrations to pathophysiological conditions (4, 7–12). Specifically, decreased serum levels of glycine were shown to correlate with gestational diabetes mellitus (13), type 2 diabetes, obesity, nonalcoholic fatty liver disease (4, 12, 14, 15), and cardiovascular disease (11, 16, 17).

Changes in serum glycine were also noted in cancer patients, with numerous studies reporting that elevated serum glycine levels are associated with several types of cancer (18–26).

Humans obtain only a fraction of total body glycine from the diet, with up to 80% being synthesized in different tissues (5, 27). In humans, the main contributor to the glycine pool is glycolytic metabolism of glucose (4, 7). In this pathway, serine is derived from the glycolytic intermediate 3-phosphoglycerate and is then converted to glycine in the reaction, which requires the coenzyme tetrahydrofolate (7, 28). The source of tetrahydrofolate in humans is dietary folate, which comes in the form of either folic acid (fortified food or multivitamin supplements) or the naturally occurring form, 5-methyltetrahydrofolate. Therefore, folate availability and metabolic flux through folate-requiring pathways are also expected to affect the available glycine pool (28, 29). For example, it has been shown that enhanced glycine production in rapidly proliferating cancer cells is a consequence of upregulated mitochondrial folate pathways (8). Circulating concentrations of glycine are also regulated by its degradation through the glycine cleavage system (GCS), a mitochondrial complex of 4 enzymes, 1 of which requires tetrahydrofolate (**Figure 1**) (30). Knockout of the enzyme catalyzing the first step in this pathway, glycine decarboxylase (GLDC), led to increased glycine in plasma and urine and the development of neural tube defects in GLDC-deficient mice (31).

We have recently reported that another folate enzyme, cytosolic aldehyde dehydrogenase 1 family member L1 [*ALDH1L1*; 10-formyltetrahydrofolate dehydrogenase (ALDH1L1)] regulates the glycine metabolism in mouse liver (Figure 1) (29). Both male and female *Aldh1l1* knockout mice have lower levels of glycine and glycine conjugates in the liver (29). Polymorphisms in the human *ALDH1L1* gene are associated with an altered serine-to-glycine ratio in plasma (9, 32, 33). *ALDH1L1* also has a very high frequency of exonic single nucleotide polymorphisms (SNPs) (34). Of note, the frequencies of specific *ALDH1L1* SNPs are different between ethnic populations, causing markedly different haplotypes (34). Here, we analyzed genetic associations of plasma serine, glycine, and the serine-glycine ratio and their relationships with markers of metabolic health in a Hispanic children cohort, the Viva La Familia Study (VFS).

Methods

Study population

The VFS was designed to identify genetic variants that influence childhood obesity and its comorbidities in Hispanic children. The VFS study design, recruitment, and methodology, demographic, and phenotypic information have been described in detail elsewhere (35, 36). A total of 319 Hispanic families participated in the study. Each family was ascertained on an obese proband, defined as a 95th BMI percentile, between the ages of 4 and 19 years (**Supplemental Figure 1**). The relative pairs based on the analyzed phenotypes range from 2 identical sibling pairs, 722 to 903 siblings, 131 to 165 half siblings, 185 to 376 first cousins, and 6 to 8 half first cousins, totaling 1049 to 1454 relative pairs. All participants gave written informed consent or assent.

Ethics

The procedures followed were in accordance with the Helsinki Declaration of 1975, as revised in 1983. The protocols for human subject research were approved by the institutional review boards at the Baylor College of Medicine and Affiliated Hospitals and the Texas Biomedical Research Institute, and the data analysis was approved by the University of North Carolina at Chapel Hill.

Phenotyping

Anthropometric measurements were performed using standardized techniques (37). Fasting EDTA plasma samples were taken from 803 Hispanic children and stored at −80 °C. Analyses were performed on the plasma samples and replicates. Methods used to measure fasting blood biochemistries were described elsewhere (35, 38). In short, serum concentrations of glucose, urate, and lipids were measured using a clinical chemistry analyzer. An ELISA-based assay was used to measure serum concentrations of leptin, adiponectin, C-reactive protein (CRP), and intracellular adhesion molecule 1. Serum concentrations of folate, homocysteine, and eotaxin were measured via chemiluminescence using a Multiplex panel (Luminex Corp).

Metabolomic profiling

Nontargeted metabolomic profiling was performed using plasma samples from the Hispanic children and replicate samples from MTRX3, a large, extensively characterized pool of human plasma maintained by Metabolon. Details of the platforms used are given in Butte et al. (39). Briefly, 3 independent platforms were used: ultra-HPLC-tandem MS optimized for basic species, ultra-HPLC-tandem MS optimized for acidic species,

and GC-MS. The final data comprised 304 metabolites, after excluding metabolites that were missing for more than 40% of participants. The original raw ion intensities for all metabolites were normalized to account for the instrument interday tuning differences. Largely, each compound was corrected in run-day blocks by registering the medians to equal 1.00 and normalizing each data point proportionately (block correction). In certain conditions, to account for differences in metabolite levels due to variation in the amount of material present in each sample, biochemical data were normalized to addition factors. Values that were missing or below the limits of detection were imputed with the compound minimum values. Serine and glycine were measured as part of the metabolomic profile. The serine-to-glycine ratios were generated using corresponding values from the metabolomic data. All metabolite values are scaled, and imputed values of their intensity are presented as means ± SDs.

SNP genotyping

The Illumina HumanOmni1-Quad v1.0 BeadChip marker assays were used to genotype 1.1 million SNPs in 815 children enrolled in the VFS (40, 41). Genotype calls were obtained after scanning on the Illumina BeadStation 500GX and were analyzed using the GenomeStudio software. The genotyping error rate was 2 per 100,000 genotypes (based on duplicates). The average call rate for all SNPs per individual sample was 97%. The SNP genotypes were checked for Mendelian consistency using the SimWalk2 program (42). The estimates of the allele frequencies and their SEs were obtained using the Sequential Oligogenic Linkage Analysis Routines (SOLAR, v.8.5.1) software (43).

Whole-exome sequencing and genome variant identification.

The custom NimbleGen VCRome 2.1 capture reagent was used to capture the entire exome for each DNA sample (44, 45). The reagent targets coding exons from the consensus coding sequence, Vega human genome annotations, and National Center for Biotechnology Information (NCBI) RNA reference sequences. Sequencing on the Illumina platform was conducted using standard protocols followed by capture enrichment. Illumina sequence analysis was conducted based on the Human Genome Sequencing Center's integrated Mercury pipeline, and quality control was performed using a custom pipeline (45, 46). We selected 5 haplotype-specific exonic variants in *ALDH1L1* from the whole-exome sequencing data for a targeted analysis.

Heritability analysis

A variance component decomposition method was used to estimate the heritability of plasma concentrations of serine and glycine and their ratio (43). To estimate the genetic contributions to the variation in these phenotypes, their heritability was estimated using the software SOLAR v.8.5.1. The total phenotypic variance can be partitioned into its genetic and environmental components. The fraction of total phenotypic variance (V_P) resulting from additive genetic effects (V_G) is called heritability, and is denoted as $h^2 = V_G/V_P$. All traits were adjusted for age, age^2, sex, age x sex, age^2 x sex, and BMI z-score (BMIZ).

Genetic correlations

Phenotypic and genetic correlations were calculated between serine, glycine, the serine-glycine ratio, and metabolic disease risk factors using the following model:

$$\rho_P = \rho_G\sqrt{h_1^2}\sqrt{h_2^2} + \mathrm{rho}_E\left(\sqrt{(1-h_1^2)}\sqrt{(1-h_2^2)}\right) \qquad (1)$$

Here, ${h_1}^2$ and ${h_2}^2$ are heritabilities of the 2 phenotypes being studied, and ρ_G, ρ_E, and ρ_P are the estimates of additive genetic, environmental, and phenotypic correlations between the traits, respectively (47). A model in which all parameters were estimated was compared with a model in which the genetic correlation was constrained to 0. To test for complete pleiotropy between the 2 traits, a model in which the genetic correlation was constrained to 1 was compared with a model in which all parameters were estimated. Twice the difference of the logarithm likelihood of the 2 models asymptotically yields a distribution of chi square with 1 df (47). Evidence of pleiotropy (a common set of genes influencing more than 1 trait) was indicated by a genetic correlation significantly different from 0.

Genome-wide association approach using measured genotype analysis.

A total of 899,892 SNPs passed quality control and were included in the genome-wide association (GWA) analysis. A measured genotype analysis (MGA) (48) was performed on the inverse, normal-transformed, residual traits (after regressing for the covariate effects mentioned above) to minimize the nonnormality distribution of the data using SOLAR. Each SNP genotype was converted in SOLAR to a covariate measure equal to 0, 1, or 2 copies of the minor alleles (or the weighted covariate based on imputation for missing genotypes). These covariates were also included in the variance-component mixed models for the MGA compared with null models that incorporated the random effect of kinship and fixed effects, such as age, sex, and their interaction (age x sex). For the initial GWA screen, we tested each SNP covariate independently as a 1-df likelihood ratio test. Empirical thresholds for genome-wide significant and suggestive values were given as $P < 1 \times 10^{-7}$ and $P < 1 \times 10^{-6}$, respectively. The details on determining these P-value thresholds have been previously published (40, 41). Specifically, these thresholds were set accounting for our family-based cohort with its unique pedigree structure. First, the effective number of SNPs was computed using the method of Moskvina and Schmidt (49) and accounting for the linkage disequilibrium between all genotyped SNPs. The average ratio of the SNP effective number to the actual number obtained from an analysis of 1989 nonoverlapping bins of SNPs was used to calculate the genome-wide effective number of tests and, thus, the significance threshold for the GWA. To test for population stratification, we performed a quantitative transmission disequilibrium test (implemented in SOLAR) (50). Family-based studies enhance statistical power for the discovery of gene functions because family members share a significant proportion of genes and tend to have homogenous environmental exposures (51). Therefore, the combination of carefully characterized pedigrees and strong associations of phenotypes with genetic markers represents a robust approach in functional genomics. We employed this

TABLE 1 Descriptive statistics of key phenotypes by age[1]

Phenotype	Mean (SD)	Range	≤7years ($n = 144$)	>7 and ≤12 years ($n = 321$)	>12 and <17 years ($n = 233$)	>17 years ($n = 105$)
BMI *z*-score	1.51 (1.0)	−0.3 to 4.48	1.27 (1.3)	1.62 (1.0)	1.58 (1.0)	1.33 (1.0)
BMI	25.28 (7.6)	13.3–61.9	18.66 (4.2)	24.59 (6.6)	28.55 (7.3)	29.55 (8.5)
Fat, %	0.33 (0.09)	0.11–0.53	0.28 (0.08)	0.35 (0.09)	0.34 (0.09)	0.32 (0.09)
Waist circumference, cm	76.01 (18)	44–147.5	56.89 (9.6)	74.52 (14.6)	85.52 (15.9)	86.62 (17.5)
Serine	1.02 (0.2)	0.03–2.09	1.12 (0.26)	0.99 (0.22)	0.99 (0.25)	1.03 (0.22)
Glycine	1.03 (0.3)	0.03–2.12	1.07 (0.31)	0.98 (0.24)	1.07 (0.27)	1.08 (0.26)
Serine-glycine ratio	1.02 (0.2)	0.44–1.95	1.07 (0.2)	1.05 (0.2)	0.95 (0.2)	0.98 (0.2)
Folate, ng/ml	29.34 (14.2)	5.7–82.4	37.61 (14.7)	34.22 (13.3)	24.65 (11.6)	(7.6)

[1]All values are presented as means (SD).

approach in the present study using the VFS cohort, which is composed of well-characterized pedigrees with proper quality control measures and links to maternal genotypes.

Targeted association analysis: ALDH1L1 exonic variants

We selected 5 haplotype-specific exonic variants in *ALDH1L1* from the whole-exome sequencing data. Like GWA studies, a regression-based approach, MGA, was used to analyze the associations of exonic variants with serine, glycine, the serine-glycine ratio, and other metabolic risk factors, including folate. The significance level was calculated for each phenotype using the Bonferroni correction. The heritability, genetic correlations, and targeted and GWA analyses were conducted using the SOLAR software, v. 8.5.1.

Results

Descriptive statistics

The metabolomic profiling was performed on 803 children (405 boys, 398 girls) from 319 families, with ages ranging between 4 and 19 years. Plasma concentrations of serine, glycine, and folate and anthropometric parameters showed high variability between individuals (**Table 1**). The ratio between serine and glycine varied considerably as well, although to a lesser extent than the concentrations of serine and glycine themselves (Table 1). When classified into different age groups, BMIs and fat percentages were associated with increasing age, as expected. In contrast, folate concentrations decreased with increasing age. Serine, glycine, and the serine-glycine ratio did not show changes associated with age. Although there was a slight decreasing trend between 7 and 12 years of ages, it was not significant and did not seem to follow any set pattern (Table 1). Of note, an early study indicated that in adults age did not affect the whole-body glycine metabolism, which was mainly dependent on protein intake (52). Our findings extend this trend (the age-independence of serum glycine) to younger age groups (Table 1).

Heritability estimates

Significant additive genetic effects or heritabilities were obtained for the metabolites and body composition measures ($P < 0.05$), ranging from 0.21 to 0.78. When classified into 2 age groups, the heritabilities ranged from 0.38 to 0.54 in those ≤12 years and from 0.47 to 0.72 in those >12 years for anthropometric traits, and from 0.23 to 0.80 in those ≤12 years and 0.24 and 1.00 in those >12 years for serine, glycine, the serine-glycine ratio, and folate (**Table 2**).

GWA analysis

We conducted a GWA analysis to identify significant variants affecting folate, serine, glycine, and the serine-glycine ratio. No significant associations were obtained for folate or serine, but 1 SNP in *GLDC*, rs1658943, was significantly associated with glycine. Key genetic associations were identified for the serine-glycine ratio, with 14 SNPs in *ALDH1L1*, *GLDC*, serine- and arginine-rich splicing factor 4 (*SRSF4*), protein tyrosine phosphatase receptor type U (*PTPRU*), and mitochondrial trans-2-enoyl-coA reductase (*MECR*; $P < 4.8 \times 10^{-7}$; **Table 3**;

TABLE 2 Heritability (h^2) estimates of key phenotypes by age

Phenotype	h^2 (SE)	*P* value	<12 years ($n = 465$)	*P* value	>12 years ($n = 338$)	*P* value
BMI *z*-score	0.21 (0.08)	0.003	0.38 (0.12)	0.0003	0.53 (0.16)	0.0008
BMI	0.33 (0.09)	4×10^{-5}	0.54 (0.13)	0.00002	0.47 (0.17)	0.004
Fat, %	0.41 (0.08)	4.1×10^{-10}	0.48 (0.12)	0.00002	0.72 (0.14)	0.000001
Waist circumference, cm	0.32 (0.09)	2.5×10^{-5}	0.49 (0.13)	0.00005	0.52 (0.16)	0.0008
Serine	0.28 (0.08)	0.00006	0.23 (0.13)	0.034	0.24 (0.19)	0.09
Glycine	0.22 (0.09)	0.004	0.33 (0.13)	0.006	0.46 (0.18)	0.006
Serine-glycine ratio	0.53 (0.09)	1.4×10^{-10}	0.73 (0.13)	1.9×10^{-8}	0.39 (0.17)	0.009
Folate, ng/ml	0.78 (0.08)	1.9×10^{-22}	0.80 (0.10)	3.4×10^{-13}	1.00 (0.50)	2.9×10^{-16}

TABLE 3 Genome-wide association analysis of glycine and serine-glycine ratio[1]

Trait	SNP	Effect size	β coefficient (SE)	*P* (SNP)	MAF	Gene[2]	Location	Chromosome (position)
Glycine	rs1658943	0.04	0.37 (0.07)	7.2×10^{-8}	0.20	*GLDC*	—	9 (6676953)
Serine-glycine ratio	rs13070856	0.07	−0.42 (0.07)	5.69×10^{-10}	0.20	*ALDH1L1*	intron	3 (126190572)
	rs9851577	0.07	−0.38 (0.06)	2.23×10^{-9}	0.27	*ALDH1L1*	intron	3 (126189467)
	rs12636371	0.06	−0.35 (0.06)	9.35×10^{-9}	0.28	*ALDH1L1*	intron	3 (126175942)
	rs13060596	0.05	−0.37 (0.07)	2.17×10^{-8}	0.24	*ALDH1L1*	intron	3 (16178695)
	rs1658943	0.05	−0.38 (0.07)	3.45×10^{-8}	0.20	*GLDC*	—	9 (6676953)
	rs2452785	0.05	0.30 (0.06)	4.74×10^{-8}	0.46	*MECR*	intron	1 (29207089)
	rs2230678	0.05	0.30 (0.06)	5.38×10^{-8}	0.46	*SRSF4*	missense	1 (29148829)
	rs2230677	0.05	0.30 (0.06)	5.38×10^{-8}	0.46	*SRSF4*	missense	1 (29148882)
	rs4521172	0.05	−0.31 (0.06)	1.57×10^{-8}	0.34	*ALDH1L1*	intron	3 (126164668)
	rs7046643	0.03	−0.31 (0.06)	1.80×10^{-8}	0.35	*GLDC*	—	9 (6682871)
	rs12404526	0.04	0.30 (0.06)	2.71×10^{-7}	0.34	*PTPRU*	intron	1 (29260191)
	—	0.04	0.30 (0.06)	2.88×10^{-7}	0.36	*SRSF4*	—	1 (29474907)
	rs4646701	0.05	−0.30 (0.06)	3.01×10^{-7}	0.41	*ALDH1L1*	intron	3 (126160613)
	rs2077523	0.05	−0.30 (0.06)	4.78×10^{-7}	0.33	*ALDH1L1*	intron	3 (126186546)

[1]Abbreviations: *ALDH1L1*, aldehyde dehydrogenase 1 family member L1; *GLDC*, glycine decarboxylase; MAF, minor allele frequency; *MECR*, mitochondrial trans-2-enoyl-CoA reductase; *PTPRU*, protein tyrosine phosphatase receptor type U; SNP, single nucleotide polymorphism; *SRSF4*, serine- and arginine-rich splicing factor 4.

[2]Regression-based genome-wide association analysis ($n = 803$).

Figure 2). The minor allele frequencies of these SNPs ranged between 0.20 and 0.46, and the effect sizes ranged from 0.03 to 0.07. For all *ALDH1L1* and *GLDC* SNPs, minor alleles were associated with lower serine-glycine ratios, whereas for *MECR*, *SRSF4*, and *PTPRU* SNPs, the minor alleles were associated with higher serine-glycine ratios. Every copy of the minor allele of rs1658943 was associated with an increase of 0.37 units of serum glycine, whereas increases in every copy of the minor allele of *ALDH1L1* SNPs were associated with 0.30- to 0.42-unit decreases in serine-glycine ratio units. We conducted an additional analysis to determine the interactions between SNPs and their associations with glycine and the serine-glycine ratio. We found a nominal association between the interaction of rs1658943 of *GLDC* with rs13070856 of *ALDH1L1* and serum glycine concentrations ($\beta = 0.08$; SE, 0.03; $P = 0.008$).

Targeted genetic analysis

We further investigated the effects of 5 *ALDH1L1* exonic SNPs (rs1127717, rs2276724, rs2886059, rs3796191, and rs4646750) on the variations in plasma concentrations of folate, serine, and glycine and the serine-glycine ratio in a targeted analysis approach (**Table 4**). The minor allele frequencies of these SNPs ranged between 5% and 14%. Our results show significant associations between rs3796191 and glycine ($P = 0.0004$) and

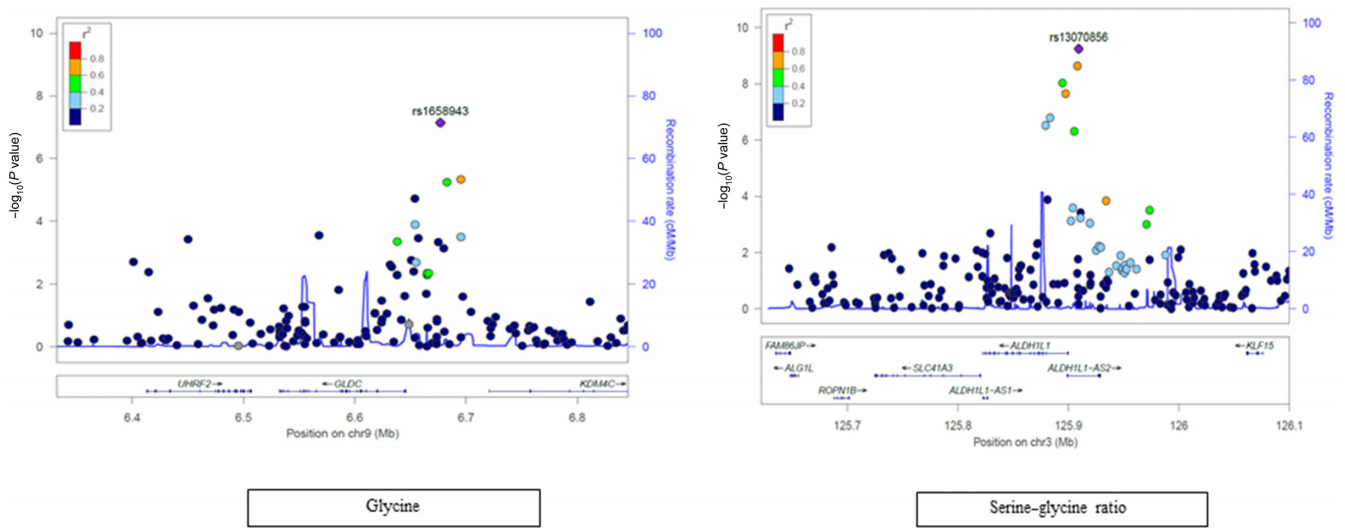

FIGURE 2 The genome-wide association of glycine and the serine-glycine ratio identifies significant associations with *GLDC* and *ALDH1L1* SNPs ($n = 803$): A locus zoom plot. Abbreviations: *ALDH1L1*, aldehyde dehydrogenase 1 family member L1; GLDC, glycine decarboxylase; SNP, single nucleotide polymorphism.

TABLE 4 *ALDH1L1* exonic SNPs association with glycine and the serine-glycine ratio[1]

Genotypes			Glycine		Serine-glycine ratio	
SNPs	Minor/major allele	Minor allele frequency	Beta (SE)	*P* value	Beta (SE)	*P* value
rs1127717	G/A	0.12	0.05 (0.09)	0.60	− 0.03 (0.09)	0.75
rs2276724	G/A	0.10	− 0.11 (0.09)	0.24	− 0.50 (0.26)	0.05
rs3796191	G/A	0.14	− 0.08 (0.02)	0.0004[2]	0.25 (0.08)	0.003[2]
rs4646750	G/A	0.05	− 0.06 (0.04)	0.06	0.16 (0.14)	0.24
rs2886059	A/C	0.12	− 0.07 (0.03)	0.02	0.68 (0.25)	0.006[2]

[1]Abbreviation: SNP, single nucleotide polymorphism.
[2]Associations significant after adjustments for multiple testing ($n = 803$), at a P value < 0.013.

the serine-glycine ratio ($P = 0.003$) and between rs2886059 and the serine-glycine ratio ($P = 0.006$; Table 4). Children with minor alleles of rs3796191 had lower concentrations of glycine and higher serine-glycine ratios as compared to children with the major allele. No significant associations were observed for these SNPs with serine or folate. The genotype-specific means of glycine and the serine-glycine ratio are depicted in **Figure 3**. For rs3796191, every addition of a minor allele was associated with a 0.08-unit decrease in the glycine concentration. Similarly, every addition of a minor allele was associated with a 0.25-unit increase in the serine-glycine ratio. Children with minor alleles of rs2886059 had higher serine-glycine ratios. Every addition of a minor allele was associated with a 0.68-unit increase in the serine-glycine ratio.

Associations of ALDH1L1 SNPs with markers of metabolic health

We performed an additional analysis of the VFS cohort and found associations of exonic SNPs in *ALDH1L1* with several markers of metabolic stress indicative of obesity (BMIZ, fat percentage, waist circumference, leptin); inflammation (adiponectin, IL-6, IL-1β, TNFα, CRP, eotaxin); type 2 diabetes [insulin, glucose, insulin resistance as estimated by HOMA-IR, Quantitative Insulin-Sensitivity Check Index (QUICKI)]; hyperlipidemia; HDL, LDL, and total cholesterol; triglycerides; hypertension (systolic and diastolic blood pressure); and altered urate concentrations. **Table 5** shows all associations with P values < 0.06. The SNP rs2276724 was associated with waist circumference, a surrogate measure of abdominal adiposity. Carriers of minor alleles tended to have a lower waist circumference. Likewise, allele-specific results of rs1127717 show that carriers of minor alleles had higher plasma concentrations of IL-6 and homocysteine and lower concentrations of urate and HDL cholesterol.

Genetic and phenotypic correlations of glycine and the serine-glycine ratio with metabolic disease risk factors

We conducted a genetic and phenotypic correlation analysis (**Table 6**) to determine whether the serine- glycine ratio was associated with metabolic stress risk factors. Both glycine and the serine-glycine ratio were genetically correlated with adiposity measures, mainly waist circumference (a surrogate measure of central or abdominal obesity) and insulin resistance or sensitivity measures (HOMA-IR and QUICKI). However, the

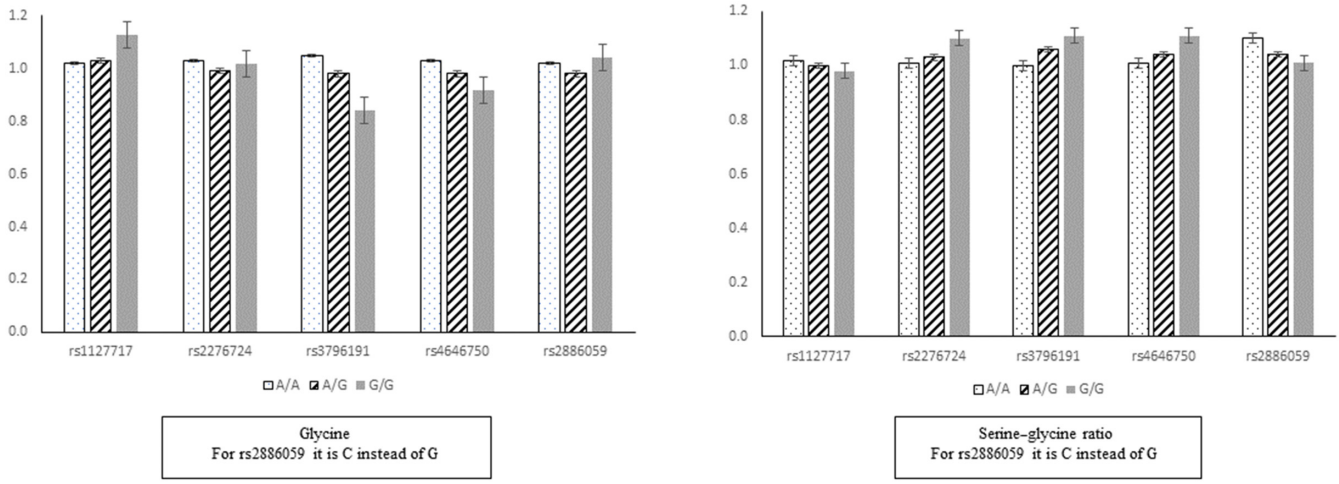

FIGURE 3 Genotype-specific means of glycine and the serine-glycine ratio in Hispanic children ($n = 803$). For rs3796191, every addition of a minor allele was associated with a 0.08-unit decrease in the glycine concentration. Similarly, every addition of a minor allele was associated with a 0.25-unit increase in the serine-glycine ratio. Children with minor alleles of rs2886059 had a higher serine-glycine ratio. Every addition of a minor allele was associated with a 0.68-unit increase in the serine-glycine ratio.

TABLE 5 *ALDH1L1* exonic SNPs associations with metabolic risk factors[1]

Phenotype	SNP	β coeff (SE)	P value[3]
BMI z-score	rs2886059	0.41 (0.21)	0.054
BMI, kg/m^2	rs2276724	− 4.9 (1.5)	0.0009[2]
	rs2886059	4.1 (1.4)	0.003[2]
Waist circumference, cm	rs2276724	− 8.02 (3.0)	0.006[2]
	rs2886059	6.67 (2.9)	0.021
Urate, mg/dl	rs1127717	− 0.21 (0.08)	0.011[2]
	rs2276724	− 0.36 (0.13)	0.028
HDL cholesterol, mg/dl	rs1127717	− 2.23 (0.9)	0.003[2]
IL-6	rs1127717	0.48 (0.2)	0.004[2]
Total homocysteine	rs1127717	0.24 (0.08)	0.004[2]
Eotaxin	rs2276724	3.78 (1.4)	0.044
	rs4646750	− 4.18 (1.9)	0.031

[1] Alleles were coded such that a major/major allele was 0, a major/minor allele was 1, and a minor/minor allele was 2. Only associations with P values < 0.06 are shown here. Abbreviations: SNP, single nucleotide polymorphism.

[2] P values significant at <0.013 after adjusting for multiple tests ($n = 782$).

genetic correlations (shared genetic effects) were stronger with glycine than the serine-glycine ratio. A phenotypic correlations analysis showed positive correlations of the serine-glycine ratio with adiposity measures (BMIZ, fat mass, fat percentage, fat-free mass, leptin, and waist circumference), insulin resistance–related measures (glucose, insulin, HOMA-IR), blood pressure, LDL cholesterol, and urate, while QUICKI values (insulin sensitivity measure) were inversely correlated with the ratio. As expected, all these results were inverse with respect to glycine concentrations.

Discussion

The most important finding of this study was the strong association of *ALDH1L1* SNPs with the serine-glycine ratio in a Hispanic population. Additionally, *GLDC* SNPs were associated with glycine and the serine-glycine ratio. Recent studies have linked plasma glycine to disease conditions and proposed it as an early marker of certain pathophysiological conditions (9, 14, 17, 53, 54). Glycine is a nonessential amino acid that can be synthetized by several pathways, most notably by the folate-dependent conversion of serine (1, 4, 5, 7). Nonetheless, studies indicate that up to 80% of this amino acid in humans could originate from endogenous biosynthesis (4, 5). Since the main pathway of glycine biosynthesis is from serine (4, 7), we evaluated the serine-to-glycine ratio as potential marker of metabolic health. Some literature suggests that this ratio could be an indicator of metabolic health (9, 14, 17, 53, 54). Thus, our primary goal in this study was to evaluate the effects of SNPs influencing the serine-glycine ratio along with serine and glycine. Our previous study in a mouse model had shown that ALDH1L1 is important to the folate-dependent serine-glycine conversion, and we hypothesized that *ALDH1L1* SNPs would have strong influences on concentrations of serine and glycine.

ALDH1L1 is a common folate enzyme highly expressed in the liver, but also in several other tissues. It is best known as a pan-astrocyte marker, and also is among the most underexpressed proteins in human cancers [reviewed in Krupenko and Krupenko (55)]. While the role of ALDH1L1 in the neural system is not completely clear, the protein downregulation in cancers was linked to its function in depleting 10-formyl-tetrahydrofolate, the substrate for the de novo purine biosynthesis. Indeed, rapidly proliferating cancer cells are strongly dependent on an abundant supply of nucleotides, and this explains the silencing of ALDH1L1, which could impede purine biosynthesis by competing for the same substrate (55). However, in nonproliferating tissues, the ALDH1L1-catalyzed reaction is important to supply tetrahydrofolate for pathways that flux 1-carbon groups into the reduced folate pool (29, 56). The key reaction in these pathways is the conversion of serine to glycine (30). Therefore, alterations in ALDH1L1 concentrations or activity can govern glycine generation and, thus, influence its concentrations in the body. While the ALDH1L1 protein level in cells is regulated through transcriptional control (57, 58) or altered degradation (56), SNPs in the gene could also affect its expression. Furthermore, exonic SNPs could also affect the ALDH1L1 function by altering the enzyme activity or stability (34). Of note, SNPs in the *ALDH1L1* gene have been investigated as risk factors for several diseases, most notably cancer [reviewed in Krupenko and Horita (34)]. These SNPs were also linked to spina bifida (59), type 2 diabetes (9), and stroke (60).

Our study indicates that several SNPs in *ALDH1L1* are associated with altered serine-glycine ratios. For most of these SNPs, minor alleles were associated with lower serine-glycine ratios. Few genomic studies have shown association of *ALDH1L1* SNPs with glycine and the serine-glycine ratio. In a study involving 4 European cohorts, *ALDH1L1* rs1107366 was strongly associated with the serine-glycine ratio (9), though in our study, this SNP was nominally associated with glycine ($P = 0.04$) and was not associated with the serine-glycine ratio. The analysis of a large European cohort in the study by Jia et al. (10) revealed associations between circulating glycine concentrations and *ALDH1L1* SNPs. The same study also found associations between glycine concentrations and *GLDC* SNPs, although the genetic variants for both *ALDH1L1* and *GLDC* reported by Jia et al. (10) were different than the SNPs in our study. As of now, most of the studies analyzing genetic influences on glycine levels have been conducted in Europeans or in Asians (9, 10, 16).

The associations between *GLDC* SNPs and altered glycine and the serine-glycine ratio in plasma are not surprising, since this enzyme is a part of the mitochondrial complex GCS, which is the main pathway of glycine degradation in mammals (31, 61). This complex includes 4 enzymes, each catalyzing a distinctive step in this process (30). GLDC catalyzes the first and the rate-limiting step in the glycine cleavage. Previous studies linked rare mutations in the *GLDC* gene to nonketotic hyperglycinemia, conditions characterized by a large accumulation of glycine in body fluids (62). The effects of common SNPs in *GLDC* on glycine levels, and associated physiological conditions, are less clear, though such SNPs have been investigated (61). For example, genome-wide studies analyzing glycine levels have shown significant associations with *GLDC* SNPs (9, 10, 33). Of note, activities of ALDH1L1 and GCS not only reside in different cellular compartments (cytoplasm compared with mitochondria) but should also oppositely affect glycine levels, facilitating glycine production (ALDH1L1) or degradation (GLDC). Furthermore, our additional analysis of interactions between SNPs showed significant associations between serum

TABLE 6 Genotypic and phenotypic correlations between glycine, the serine-glycine ratio, and metabolic disease risk factors ($n = 782$)[1]

		Genetic correlations (ρ_G)		Phenotypic correlations (ρ_P)	
Phenotype 1	Phenotype 2	Correlation coefficient (SE)	*P* value	Correlation coefficient (SE)	*P* value
Serine-glycine	BMI *z*-score	0.34 (0.15)	0.041[2]	0.19 (0.02)	1.5×10^{-6} [2]
	Fat mass	0.27 (0.16)	0.10	0.19 (0.02)	2.2×10^{-6} [2]
	Fat free mass	0.13 (0.14)	0.37	0.11 (0.02)	0.006[2]
	Percent fat	0.25 (0.14)	0.10	0.19 (0.02)	1.5×10^{-6} [2]
	Waist circumference	0.39 (0.14)	0.009[2]	0.24 (0.02)	9.6×10^{-10} [2]
	Glucose	0.29 (0.16)	0.08	0.16 (0.02)	0.00001[2]
	Insulin	0.31 (0.16)	0.065	0.14 (0.02)	0.0003[2]
	HOMA-IR	0.35 (0.16)	0.031[2]	0.17 (0.02)	5.5×10^{-6} [2]
	QUICKI	−0.37 (0.16)	0.030[2]	−0.19 (0.03)	6.8×10^{-7} [2]
	Leptin	0.10 (0.16)	0.53	0.13 (0.01)	0.0008[2]
	Systolic blood pressure	0.18 (0.15)	0.25	0.130 (0.01)	0.0008[2]
	LDL cholesterol	0.08 (0.12)	0.52	0.08 (0.03)	0.040[2]
	Urate	0.24 (0.14)	0.09	0.11 (0.02)	0.004[2]
	CRP	0.16 (0.16)	0.32	0.16 (0.02)	5.0×10^{-6} [2]
Glycine	BMI *z*-score	−0.43 (0.18)	0.048[2]	−0.41 (0.003)	9.4×10^{-27} [2]
	Fat mass	−0.54 (0.17)	0.013[2]	−0.37 (0.03)	2.0×10^{-23} [2]
	Fat free mass	−0.28 (0.18)	0.15	−0.23 (0.03)	7.5×10^{-10} [2]
	Percent fat	−0.41 (0.16)	0.037[2]	−0.42 (0.04)	6.6×10^{-27} [2]
	Waist circumference	−0.63 (0.14)	0.001[2]	−0.43 (0.04)	2.3×10^{-29} [2]
	Glucose	−0.40 (0.21)	0.07	−0.15 (0.02)	0.000035[2]
	Insulin	−0.62 (0.18)	0.005[2]	−0.37 (0.04)	6.1×10^{-25}
	HOMA-IR	−0.60 (0.18)	0.005[2]	−0.36 (0.04)	3.0×10^{-24}
	QUICKI	0.75 (0.17)	0.0009[2]	0.42 (0.03)	1.5×10^{-31} [2]
	Leptin	−0.02 (0.21)	0.92	−0.35 (0.03)	1.2×10^{-19} [2]
	Adiponectin	0.10 (0.15)	0.49	0.19 (0.02)	6.7×10^{-7} [2]
	Systolic blood pressure	−0.18 (0.20)	0.40	−0.25 (0.02)	1.6×10^{-11} [2]
	Total cholesterol	−0.19 (0.15)	0.22	−0.17 (0.02)	3.8×10^{-6} [2]
	HDL cholesterol	0.04 (0.16)	0.81	0.094 (0.03)	0.010[2]
	LDL cholesterol	−0.14 (0.16)	0.41	−0.15 (0.02)	0.00005[2]
	Triglycerides	−0.33 (0.16)	0.051	−0.23 (0.02)	5.0×10^{-11} [2]
	Urate	−0.56 (0.17)	0.004[2]	−0.25 (0.04)	1.0×10^{-12} [2]
	CRP	−0.28 (0.20)	0.20	−0.28 (0.03)	1.7×10^{-15} [2]
	ICAM-1	−0.17 (0.17)	0.34	−0.19 (0.02)	3.0×10^{-6} [2]
	IL-6	0.31 (0.22)	0.16	−0.08 (0.02)	0.03[2]
	Total homocysteine	0.25 (0.15)	0.09	0.094 (0.03)	0.014[2]

[1]Abbreviations: ρ_G, estimate of genetic correlation; ρ_P, estimate of phenotypic correlation; CRP, C-reactive protein; ICAM-1, intracellular adhesion molecule 1; QUICKI, Quantitative Insulin-Sensitivity Check Index.

[2]Correlations that are genotypically or phenotypically significant.

glycine concentrations and the interaction between rs1658943 of GLDC and rs13070856 of ALDH1L1, indicating a functional link between pathways of glycine production and degradation. The other SNP that had significant associations with the serine-glycine ratio belonged to *MECR*. The *MECR* gene encodes an enzyme that catalyzes the last step of mitochondrial fatty acid synthesis, and its overexpression has been shown to increase the activity of peroxisome proliferator-activated receptor α, which further downregulates proinflammatory genes (63, 64). The *MECR* SNPs have been linked with airway obstruction (64) and neurodegenerative diseases (65).

Three SNPs in the *SRSF4* gene were also associated with the serine-glycine ratio. This gene codes for a serine/arginine-rich splicing factor, which is involved in alternative splicing of a wide variety of pre-mRNAs (66, 67). While the effects of splicing factor SNPs on metabolic processes can be very indirect, the list of genes in which splicing is regulated by SRSF4 in a murine model includes *Gldc* (67). Importantly, SNPs in the serine/arginine-rich protein family, including 1 SNP in *SRSF4*, are linked to pathophysiological conditions (66). Additionally, our analysis shows that in general, SNPs associated with altered glycine levels are also linked to an altered metabolic phenotype, suggesting their regulatory role and highlighting the necessity of evaluating such SNPs as risk factors for certain metabolic diseases. Of note, we have previously shown lower glycine and serine levels in children with obesity from the VFS cohort (39), and such phenomena have also been reported recently in the Uppsala Longitudinal Study of Childhood Obesity cohort (68). These findings have an important implication, since higher glycine levels are associated with increased insulin sensitivity (12). Furthermore, glycine concentrations and the serine-glycine ratio in our study correlated well with common markers of metabolic health, such as the levels of glucose and insulin, fat mass, fat percentage, and waist circumference. Likewise, our analysis of genetic variants linked *ALDH1L1* SNPs not only to glycine levels and the serine-glycine ratio but also to parameters

reflecting the metabolic phenotype, such as adiposity, indicating a very intricate relationship between diet, genotypes, and metabolic outcomes. Despite such complexity, circulating glycine levels, as well as the serine-glycine ratio, could be helpful markers to predict alterations in metabolic health.

In conclusion, glycine is involved in key metabolic processes, and alterations of its circulating levels could cause metabolic insufficiency or serve as a marker of certain metabolic diseases or conditions. Glycine levels are significantly influenced by genetic variations in enzymes of glycine-related pathways. Since frequencies of related SNPs vary significantly between ethnic populations, the glycine status could shift as well. Of note, studies correlating the glycine status with the risks of metabolic diseases have been mostly conducted in European populations, and there is a paucity of relevant genetic studies, as well as insufficient representation in genetic databases, of non-European populations. In this respect, our study provides important novel genetic associations for glycine and the serine-glycine ratio in Hispanic children. Our findings related to *ALDH1L1* and *GLDC* SNPs may have functional implications for serine-glycine conversion and the glycine metabolism, and provide a mechanistic basis linking these SNPs to downstream pathways and to the risks of metabolic diseases. It should be noted that the effects of specific SNPs on the glycine metabolism can likely be modified depending on the diet composition, primarily the folate and protein content. Towards this end, a randomized controlled intervention study involving 57 healthy women showed increased plasma glycine in the folic acid–supplemented group (69). Though several studies investigated the effect of the diet on serum glycine (70–73), the association between protein consumption and circulating glycine levels in humans is not clear due to additional contributing factors, such as the source of protein, the role of the microbiome in glycine bioavailability, and the compensatory effect of dietary betaine [reviewed in Alves et al. (4)]. Thus, future genetic and biomarker analyses should be considered in the context of specific diets to be in line with the precision nutrition approach.

We thank the participants of the Viva la Familia Study for their extraordinary cooperation and involvement in this project.

The authors' responsibilities were as follows—SAK, SAC, AGC, NFB, and VSV: conceived and designed the research; SAK and VSV conducted the research, wrote the paper, have primary responsibility for the final content; VSV and RH: analyzed data; KH, SL, and NRM: performed the experiments; and all authors: read and approved the final manuscript.

Author disclosures: The authors report no conflicts of interest.

Data Availability

Data described in the manuscript, the code book, and the analytic code will be made available upon request.

References

1. Razak MA, Begum PS, Viswanath B, Rajagopal S. Multifarious beneficial effect of nonessential amino acid, glycine: a review. Oxid Med Cell Longev 2017;2017:1716701.
2. Badenhorst CP, van der Sluis R, Erasmus E, van Dijk AA. Glycine conjugation: importance in metabolism, the role of glycine N-acyltransferase, and factors that influence interindividual variation. Expert Opin Drug Metab Toxicol 2013;9(9):1139–53.
3. Bowery NG, Smart TG. GABA and glycine as neurotransmitters: a brief history. Br J Pharmacol 2006;147(S1):S109–19.
4. Alves A, Bassot A, Bulteau AL, Pirola L, Morio B. Glycine metabolism and its alterations in obesity and metabolic diseases. Nutrients 2019;11(6):1356.
5. Wang W, Wu Z, Dai Z, Yang Y, Wang J, Wu G. Glycine metabolism in animals and humans: implications for nutrition and health. Amino Acids 2013;45(3):463–77.
6. Shafiekhani M, Ommati MM, Azarpira N, Heidari R, Salarian AA. Glycine supplementation mitigates lead-induced renal injury in mice. J Exp Pharmacol 2019;11:15–22.
7. Adeva-Andany M, Souto-Adeva G, Ameneiros-Rodriguez E, Fernandez-Fernandez C, Donapetry-Garcia C, Dominguez-Montero A. Insulin resistance and glycine metabolism in humans. Amino Acids 2018;50(1):11–27.
8. Jain M, Nilsson R, Sharma S, Madhusudhan N, Kitami T, Souza AL, Kafri R, Kirschner MW, Clish CB, Mootha VK. Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. Science 2012;336(6084):1040–4.
9. Xie W, Wood AR, Lyssenko V, Weedon MN, Knowles JW, Alkayyali S, Assimes TL, Quertermous T, Abbasi F, Paananen J, et al. Genetic variants associated with glycine metabolism and their role in insulin sensitivity and type 2 diabetes. Diabetes 2013;62(6):2141–50.
10. Jia Q, Han Y, Huang P, Woodward NC, Gukasyan J, Kettunen J, Ala-Korpela M, Anufrieva O, Wang Q, Perola M, et al. Genetic determinants of circulating glycine levels and risk of coronary artery disease. J Am Heart Assoc 2019;8(10):e011922.
11. Ntzouvani A, Nomikos T, Panagiotakos D, Fragopoulou E, Pitsavos C, McCann A, Ueland PM, Antonopoulou S. Amino acid profile and metabolic syndrome in a male Mediterranean population: a cross-sectional study. Nutr Metab Cardiovasc Dis 2017;27(11):1021–30.
12. Yan-Do R, MacDonald PE. Impaired "glycine"-mia in type 2 diabetes and potential mechanisms contributing to glucose homeostasis. Endocrinology 2017;158(5):1064–73.
13. Nevalainen J, Sairanen M, Appelblom H, Gissler M, Timonen S, Ryynanen M. First-trimester maternal serum amino acids and acylcarnitines are significant predictors of gestational diabetes. Rev Diabet Stud 2016;13(4):236–45.
14. Gar C, Rottenkolber M, Prehn C, Adamski J, Seissler J, Lechner A. Serum and plasma amino acids as markers of prediabetes, insulin resistance, and incident diabetes. Crit Rev Clin Lab Sci 2018;55(1):21–32.
15. Klein MS, Shearer J. Metabolomics and type 2 diabetes: translating basic research into clinical application. J Diabetes Res 2016;2016:3898502.
16. Chang X, Wang L, Guan SP, Kennedy BK, Liu J, Khor CC, Low AF, Chan MY, Yuan JM, Koh WP, et al. The association of genetically determined serum glycine with cardiovascular risk in East Asians. Nutr Metab Cardiovasc Dis 2021;31(6):1840–4.
17. Ding Y, Svingen GF, Pedersen ER, Gregory JF, Ueland PM, Tell GS, Nygard OK. Plasma glycine and risk of acute myocardial infarction in patients with suspected stable angina pectoris. J Am Heart Assoc 2016;5(1):e002621.
18. Cadoni G, Giraldi L, Chiarla C, Gervasoni J, Persichilli S, Primiano A, Settimi S, Galli J, Paludetti G, Arzani D, et al. Prognostic role of serum amino acids in head and neck cancer. Dis Markers 2020;2020:2291759.
19. Zhou Q, Zhang LY, Xie C, Zhang ML, Wang YJ, Liu GH. Metabolomics as a potential method for predicting thyroid malignancy in children and adolescents. Pediatr Surg Int 2020;36(2):145–53.
20. Niziol J, Ossolinski K, Tripet BP, Copie V, Arendowski A, Ruman T. Nuclear magnetic resonance and surface-assisted laser desorption/ionization mass spectrometry-based serum metabolomics of kidney cancer. Anal Bioanal Chem 2020;412(23):5827–41.
21. Berker Y, Vandergrift LA, Wagner I, Su L, Kurth J, Schuler A, Dinges SS, Habbel P, Nowak J, Mark E, et al. Magnetic resonance spectroscopy-based metabolomic biomarkers for typing, staging, and survival estimation of early-stage human lung cancer. Sci Rep 2019;9(1):10319.
22. Ni J, Xu L, Li W, Zheng C, Wu L. Targeted metabolomics for serum amino acids and acylcarnitines in patients with lung cancer. Exp Ther Med 2019;18(1):188–98.
23. Amara CS, Ambati CR, Vantaku V, Badrajee Piyarathna DW, Donepudi SR, Ravi SS, Arnold JM, Putluri V, Chatta G, Guru KA, et al.

Serum metabolic profiling identified a distinct metabolic signature in bladder cancer smokers: a key metabolic enzyme associated with patient survival. Cancer Epidemiol Biomarkers Prev 2019;28(4): 770–81.
24. Rohnisch HE, Kyro C, Olsen A, Thysell E, Hallmans G, Moazzami AA. Identification of metabolites associated with prostate cancer risk: a nested case-control study with long follow-up in the Northern Sweden Health and Disease Study. BMC Med 2020;18(1):187.
25. Yang B, Wang C, Xie Y, Xu L, Wu X, Wu D. Monitoring tyrosine kinase inhibitor therapeutic responses with a panel of metabolic biomarkers in chronic myeloid leukemia patients. Cancer Sci 2018;109(3):777–84.
26. Sirnio P, Vayrynen JP, Klintrup K, Makela J, Karhu T, Herzig KH, Minkkinen I, Makinen MJ, Karttunen TJ, Tuomisto A. Alterations in serum amino-acid profile in the progression of colorectal cancer: associations with systemic inflammation, tumour stage and patient survival. Br J Cancer 2019;120(2):238–46.
27. Zhong Z, Wheeler MD, Li X, Froh M, Schemmer P, Yin M, Bunzendaul H, Bradford B, Lemasters JJ. L-glycine: a novel antiinflammatory, immunomodulatory, and cytoprotective agent. Curr Opin Clin Nutr Metab Care 2003;6(2):229–40.
28. Ducker GS, Rabinowitz JD. One-carbon metabolism in health and disease. Cell Metab 2017;25(1):27–42.
29. Krupenko NI, Sharma J, Pediaditakis P, Fekry B, Helke KL, Du X, Sumner S, Krupenko SA. Cytosolic 10-formyltetrahydrofolate dehydrogenase regulates glycine metabolism in mouse liver. Sci Rep 2019;9(1):14937.
30. Tibbetts AS, Appling DR. Compartmentalization of mammalian folate-mediated one-carbon metabolism. Annu Rev Nutr 2010;30(1):57–81.
31. Pai YJ, Leung KY, Savery D, Hutchin T, Prunty H, Heales S, Brosnan ME, Brosnan JT, Copp AJ, Greene ND. Glycine decarboxylase deficiency causes neural tube defects and features of non-ketotic hyperglycinemia in mice. Nat Commun 2015;6(1):6388.
32. Dharuri H, Henneman P, Demirkan A, van Klinken JB, Mook-Kanamori DO, Wang-Sattler R, Gieger C, Adamski J, Hettne K, Roos M, et al. Automated workflow-based exploitation of pathway databases provides new insights into genetic associations of metabolite profiles. BMC Genom 2013;14(1):865.
33. Wittemans LBL, Lotta LA, Oliver-Williams C, Stewart ID, Surendran P, Karthikeyan S, Day FR, Koulman A, Imamura F, Zeng L, et al. Assessing the causal association of glycine with risk of cardio-metabolic diseases. Nat Commun 2019;10(1):1060.
34. Krupenko SA, Horita DA. The role of single-nucleotide polymorphisms in the function of candidate tumor suppressor ALDH1L1. Front Genet 2019;10:1013.
35. Butte NF, Cai G, Cole SA, Comuzzie AG. Viva la Familia Study: Genetic and environmental contributions to childhood obesity and its comorbidities in the Hispanic population. Am J Clin Nutr 2006;84(3):646–54; quiz 73–4.
36. Butte NF, Cai G, Cole SA, Wilson TA, Fisher JO, Zakeri IF, Ellis KJ, Comuzzie AG. Metabolic and behavioral predictors of weight gain in Hispanic children: the Viva la Familia Study. Am J Clin Nutr 2007;85(6):1478–85.
37. Lohman TG, Roche AF, Martorell R. Anthropometric standardization reference manual. Champaign (IL): Human Kinetics Books; 1988.
38. Voruganti VS, Goring HH, Diego VP, Cai G, Mehta NR, Haack K, Cole SA, Butte NF, Comuzzie AG. Genome-wide scan for serum ghrelin detects linkage on chromosome 1p36 in Hispanic children: results from the Viva La Familia Study. Pediatr Res 2007;62(4):445–50.
39. Butte NF, Liu Y, Zakeri IF, Mohney RP, Mehta N, Voruganti VS, Goring H, Cole SA, Comuzzie AG. Global metabolomic profiling targeting childhood obesity in the Hispanic population. Am J Clin Nutr 2015;102(2):256–67.
40. Comuzzie AG, Cole SA, Laston SL, Voruganti VS, Haack K, Gibbs RA, Butte NF. Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. PLoS One 2012;7(12):e51954.
41. Voruganti VS, Laston S, Haack K, Mehta NR, Cole SA, Butte NF, Comuzzie AG. Serum uric acid concentrations and SLC2A9 genetic variation in Hispanic children: the Viva La Familia Study. Am J Clin Nutr 2015;101(4):725–32.
42. Sobel E, Lange K. Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. Am J Hum Genet 1996;58(6):1323–37.
43. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 1998;62(5):1198–211.
44. Sabo A, Mishra P, Dugan-Perez S, Voruganti VS, Kent JW Jr., Kalra D, Cole SA, Comuzzie AG, Muzny DM, Gibbs RA, et al. Exome sequencing reveals novel genetic loci influencing obesity-related traits in Hispanic children. Obesity (Silver Spring) 2017;25(7): 1270–6.
45. Hou R, Cole SA, Graff M, Haack K, Laston S, Comuzzie AG, Mehta NR, Ryan K, Cousminer DL, Zemel BS, et al. Genetic variants affecting bone mineral density and bone mineral content at multiple skeletal sites in Hispanic children. Bone 2020;132:115175.
46. Yu B, Pulit SL, Hwang SJ, Brody JA, Amin N, Auer PL, Bis JC, Boerwinkle E, Burke GL, Chakravarti A, et al. Rare exome sequence variants in CLCN6 reduce blood pressure levels and hypertension risk. Circ Cardiov Genet 2016;9(1):64–70.
47. Hopper JL, Mathews JD. Extensions to multivariate normal models for pedigree analysis. Ann Hum Genet 1982;46(4):373–83.
48. Boerwinkle E, Chakraborty R, Sing CF. The use of measured genotype information in the analysis of quantitative phenotypes in man. I. Models and analytical methods. Ann Hum Genet 1986;50(2): 181–94.
49. Moskvina V, Schmidt KM. On multiple-testing correction in genome-wide association studies. Genet Epidemiol 2008;32(6): 567–73.
50. Havill LM, Dyer TD, Richardson DK, Mahaney MC, Blangero J. The quantitative trait linkage disequilibrium test: a more powerful alternative to the quantitative transmission disequilibrium test for use in the absence of population stratification. BMC Genet 2005;6(S1): S91.
51. Borecki IB, Province MA. Genetic and genomic discovery using family studies. Circulation 2008;118(10):1057–63.
52. Gersovitz M, Bier D, Matthews D, Udall J, Munro HN, Young VR. Dynamic aspects of whole body glycine metabolism: influence of protein intake in young adult and elderly males. Metabolism 1980;29(11):1087–94.
53. Takashina C, Tsujino I, Watanabe T, Sakaue S, Ikeda D, Yamada A, Sato T, Ohira H, Otsuka Y, Oyama-Manabe N, et al. Associations among the plasma amino acid profile, obesity, and glucose metabolism in Japanese adults with normal glucose tolerance. Nutr Metab (Lond) 2016;13(1): 5.
54. Kettunen J, Demirkan A, Wurtz P, Draisma HH, Haller T, Rawal R, Vaarhorst A, Kangas AJ, Lyytikainen LP, Pirinen M, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. Nat Commun 2016;7(1):11122.
55. Krupenko SA, Krupenko NI. Loss of ALDH1L1 folate enzyme confers a selective metabolic advantage for tumor progression. Chem Biol Interact 2019;302:149–55.
56. Khan QA, Pediaditakis P, Malakhau Y, Esmaeilniakooshkghazi A, Ashkavand Z, Sereda V, Krupenko NI, Krupenko SA. CHIP E3 ligase mediates proteasomal degradation of the proliferation regulatory protein ALDH1L1 during the transition of NIH3T3 fibroblasts from G0/G1 to S-phase. PLoS One 2018;13(7):e0199699.
57. Oleinik NV, Krupenko NI, Krupenko SA. Epigenetic silencing of ALDH1L1, a metabolic regulator of cellular proliferation, in cancers. Genes Cancer 2011;2(2):130–9.
58. Damal Villivalam S, Ebert SM, Lim HW, Kim J, You D, Jung BC, Palacios HH, Tcheau T, Adams CM, Kang S. A necessary role of DNMT3A in endurance exercise by suppressing ALDH1L1-mediated oxidative stress. EMBO J 2021;40(9):e106491.
59. Franke B, Vermeulen SH, Steegers-Theunissen RP, Coenen MJ, Schijvenaars MM, Scheffer H, den Heijer M, Blom HJ. An association study of 45 folate-related genes in spina bifida: involvement of cubilin (CUBN) and tRNA aspartic acid methyltransferase 1 (TRDMT1). Birth Defects Res A Clin Mol Teratol 2009;85(3):216–26.
60. Williams SR, Yang Q, Chen F, Liu X, Keene KL, Jacques P, Chen WM, Weinstein G, Hsu FC, Beiser A, et al. Genome-wide meta-analysis of homocysteine and methionine metabolism identifies five one carbon metabolism loci and a novel association of ALDH1L1 with ischemic stroke. PLoS Genet 2014;10(3):e1004214.
61. O'Reilly J, Pangilinan F, Hokamp K, Ueland PM, Brosnan JT, Brosnan ME, Brody LC, Molloy AM. The impact of common genetic variants in the mitochondrial glycine cleavage system on relevant metabolites. Mol Genet Metab Rep 2018;16:20–2.

62. Coughlin CR 2nd, Swanson MA, Kronquist K, Acquaviva C, Hutchin T, Rodriguez-Pombo P, Vaisanen ML, Spector E, Creadon-Swindell G, Bras-Goldberg AM, et al. The genetic basis of classic nonketotic hyperglycinemia due to mutations in GLDC and AMT. Genet Med 2017;19(1):104–11.
63. Kim DG, Yoo JC, Kim E, Lee YS, Yarishkin OV, Lee DY, Lee KH, Hong SG, Hwang EM, Park JY. A novel cytosolic isoform of mitochondrial trans-2-enoyl-CoA reductase enhances peroxisome proliferator-activated receptor α activity. Endocrinol Metab (Seoul) 2014;29(2):185–94.
64. van der Plaat DA, de Jong K, Lahousse L, Faiz A, Vonk JM, van Diemen CC, Nedeljkovic I, Amin N, Obeidat M, van Duijn CM, et al. The well-known gene HHIP and novel gene MECR are implicated in small airway obstruction. Am J Respir Crit Care Med 2016;194(10): 1299–302.
65. Nair RR, Koivisto H, Jokivarsi K, Miinalainen IJ, Autio KJ, Manninen A, Poutiainen P, Tanila H, Hiltunen JK, Kastaniotis AJ. Impaired mitochondrial fatty acid synthesis leads to neurodegeneration in mice. J Neurosci 2018;38(45):9781–800.
66. Tan W, Wang W, Ma Q. Physiological and pathological function of serine/arginine-rich splicing factor 4 and related diseases. Biomed Res Int 2018;2018:3819719.
67. Anko ML, Muller-McNicoll M, Brandl H, Curk T, Gorup C, Henry I, Ule J, Neugebauer KM. The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. Genome Biol 2012;13(3):R17.
68. Mullner E, Rohnisch HE, von Bromssen C, Moazzami AA. Metabolomics analysis reveals altered metabolites in lean compared with obese adolescents and additional metabolic shifts associated with hyperinsulinaemia and insulin resistance in obese adolescents: a cross-sectional study. Metabolomics 2021;17(1):11.
69. Hefni ME, Witthoft CM, Moazzami AA. Plasma metabolite profiles in healthy women differ after intervention with supplemental folic acid v. folate-rich foods. J Nutr Sci 2018;7:e32.
70. Iguacel I, Schmidt JA, Perez-Cornago A, Van Puyvelde H, Travis R, Stepien M, Scalbert A, Casagrande C, Weiderpass E, Riboli E, et al. Associations between dietary amino acid intakes and blood concentration levels. Clin Nutr 2021;40(6): 3772–9.
71. Wittenbecher C, Muhlenbruch K, Kroger J, Jacobs S, Kuxhaus O, Floegel A, Fritsche A, Pischon T, Prehn C, Adamski J, et al. Amino acids, lipid metabolites, and ferritin as potential mediators linking red meat consumption to type 2 diabetes. Am J Clin Nutr 2015;101(6):1241–50.
72. Sjoberg KA, Schmoll D, Piper MDW, Kiens B, Rose AJ. Effects of short-term dietary protein restriction on blood amino acid levels in young men. Nutrients 2020;12(8):2195.
73. Altorf-van der Kuil W, Brink EJ, Boetje M, Siebelink E, Bijlsma S, Engberink MF, van't Veer P, Tome D, Bakker SJ, van Baak MA, et al. Identification of biomarkers for intake of protein from meat, dairy products and grains: a controlled dietary intervention study. Br J Nutr 2013;110(5):810–22.