scientific reports

Check for updates

OPEN

Genome-wide association study of plasma amino acids and Mendelian randomization for cardiometabolic traits

Ryota Toki[1], Sotaro Fushiki[1,2], Shun Kojima[1,2], Yoichi Sutoh[3,4], Yayoi Otsuka-Yamasaki[3,4], Sei Harada[1], Miho Iida[1], Aya Hirata[1], Naoko Miyagawa[1], Minako Matsumoto[1], Shun Edagawa[1], Atsuko Miyake[1], Kazuyo Kuwabara[1], Akiyoshi Hirayama[5], Masahiro Sugimoto[5], Asako Sato[5], Kaori Amano[5], Tomoyoshi Soga[5], Masaru Tomita[5], Kazuharu Arakawa[5], Kengo Kinoshita[6,7,8,9], Mika Sakurai-Yageta[6,8], Gen Tamiya[6,10,11], Hideki Ohmomo[3,4], Atsushi Shimizu[3,4], Tomonori Okamura[1] & Toru Takebayashi[1]✉

Plasma amino acids (AAs) have emerged as promising biomarkers for metabolic disorders, yet their causality remains unclear. We aimed to investigate the genetic determinants of AA levels in a cohort of 10,333 individuals and their causal effects on cardiometabolic traits using Mendelian randomization (MR). Plasma levels of 20 AAs were quantified using capillary electrophoresis mass spectrometry. Genome-wide association studies were conducted using BOLT-LMM and heritability estimation via LDSC analysis. Causal effects of AAs on 11 cardiometabolic traits were examined using two-sample MR analyses. We identified 85 locus-metabolite associations across 43 genes for 18 AAs, including 44 novel loci linked to metabolic genes. Heritability for AAs was estimated at 16%. MR analysis demonstrated cystine to positively associate with systolic blood pressure (SBP) (β = 0.056, SE = 0.010), while serine indicated protective effects on SBP (β = − 0.040, SE = 0.011), diastolic BP (β = − 0.044, SE = 0.010), and coronary artery disease (odds ratio 0.888, SE = 0.028). We identified potentially novel genetic loci associated with AA levels and demonstrated robust causal associations between several AAs and cardiometabolic traits. These findings reinforce the importance of AAs as potential biomarkers and therapeutic targets in cardiometabolic health.

Keywords Amino acids, Cardiometabolic diseases, Genome-wide association study, Mendelian randomization, Metabolomics

Background

Plasma metabolome is one of the promising tools to facilitate the diagnosis, prognosis, and risk assessment of various diseases[1,2]. Among various metabolites, amino acids (AAs) are valuable biomarkers of hypertension, type 2 diabetes mellitus (T2DM), and other cardiometabolic diseases[3–5]. Considering the pivotal roles in physiological processes and their sensitivity to pathological changes, amino acids are expected to serve as important markers of disease prevention and treatment efficacy in precision medicine.

[1]Department of Preventive Medicine and Public Health, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo, Japan. [2]Konica Minolta, Inc, Chiyoda-ku, Tokyo, Japan. [3]Division of Bioinformation Analysis, Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Shiwa-gun, Iwate, Japan. [4]Division of Bioinformation Analysis, Institute for Biomedical Sciences, Iwate Medical University, Shiwa-gun, Iwate, Japan. [5]Institute for Advanced Biosciences, Keio University, Tsuruoka City, Yamagata, Japan. [6]Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan. [7]Graduate School of Information Sciences, Tohoku University, Sendai, Miyagi, Japan. [8]Advanced Research Center for Innovations in Next-Generation Medicine, Tohoku University, Sendai, Miyagi, Japan. [9]Institute of Development, Aging and Cancer, Tohoku University, Sendai, Miyagi, Japan. [10]Graduate School of Medicine, Tohoku University, Sendai, Miyagi, Japan. [11]Center for Advanced Intelligence Project, RIKEN, Wako, Saitama, Japan. ✉email: ttakebayashi@keio.jp

AAs are considered heritable and genetically driven[6,7] in addition to being reflected by environmental factors[8]. Several genome-wide association studies (GWASs) have identified genetic polymorphisms associated with AAs[9–11], but the majority were conducted in European populations, and those involving East Asian populations had limited sample sizes. As many differences are observed between ethnic groups, including susceptibility to diseases and responses to drugs, it is critical to investigate the genetic basis for such differences. Furthermore, recent studies indicate the existence of a shared etiology with a broad metabolic basis across diseases and conditions[1]. Thus, exploring the mechanisms behind metabolic dynamics and examining causality between metabolite levels and diseases is becoming increasingly important, utilizing Mendelian randomization (MR) analysis as one useful approach[12–14]. However, the causal relationship between AAs and cardiometabolic diseases has been controversial.

Therefore, the present study aimed to explore genetic polymorphisms associated with plasma AAs in East Asian populations[11,15–17] to determine whether these loci are shared among ethnic groups. Furthermore, we investigated the causal relationship between AAs and cardiometabolic traits, including body mass index (BMI), lipids, hypertension, and T2DM, via two-sample MR analysis using the summary statistics from the GWAS in this study and publicly available data.

Results

Genetic polymorphisms associated with plasma AAs

A total of 10,333 individuals (5,514 women, 59.6 ± 10.0 years old) were included in the analysis. The plasma AA levels in this cohort were similar to those previously reported in a Japanese healthy population[18] (Table 1). The sample sizes for each GWAS conducted after phenotype quality control (QC) for each AA are shown in Table S1.

GWAS analysis demonstrated 85 loci-metabolite associations across 43 genes for 18 AA plasma concentrations with p-values $< 5 \times 10^{-8}$ (Table S2). Of these, 44 loci-metabolite associations across 23 genes were potentially novel with a one-to-one association with AA levels, and eight of these genes were enzyme-coding genes (Fig. 1; Table 2). As a sensitivity analysis, we also excluded individuals who reported antihypertensive, lipid-lowering, or antidiabetic medications and observed no significant changes (data not shown).

Of the remaining loci-metabolite associations found in this study, 31 associations were replicated in the GWAS Catalog, which includes results from studies that involve multi-ethnic populations. Similarly, 25 associations were reproduced from Tohoku Medical Megabank (TMM) in Japan genotyped with the same array[19], two more from another Japanese study using a different array[11]; comparing the results from other populations, two were consistent with a study of Chinese populations[17], four with the Western populations[9], and none with the Middle East[20]. (Table S3). For the top variants that precisely matched previously reported SNPs (e.g., alanine–*GCKR*, glutamine–*LMO1*), we confirmed that both the lead peak positions and effect directions aligned with earlier studies. In other cases where the top variants did not exactly match, we were able to verify positional similarity.

Genetic loci associated with multiple AAs are listed in Fig. 2, along with a correlation matrix for each pair of plasma AAs. A total of ten genes, including *SESN2* and *PDE4DIP*, were associated with multiple AAs, suggesting their pleiotropic signature. Several genes exhibited different directions of association for multiple AAs, such as *CPS1*. Positive associations were observed for both glycine and serine, which were also positively correlated with each other, whereas negative associations were observed for threonine, histidine, and arginine. Although the associations with *CPS1* were in the opposite direction for glycine and arginine, they were positively correlated, indicating differences in the direction of associations and correlations among AAs even when associated with the same genetic loci.

The heritability for each AA estimated using linkage disequilibrium score regression (LDSC) found cystine to have the highest heritability (h^2 = 0.30 [SE = 0.086]), and serine had the lowest (h^2 = 0.015 [SE = 0.085]). When all 20 AAs were combined, the weighted heritability was 0.16 (SE = 0.018), with minimal heterogeneity (Cochran's Q = 16.3, I^2 = 0) (Table S4).

The causal relationship between AAs and cardiometabolic traits

The results of MR analyses are summarized in Fig. 3 and Table S5. For systolic blood pressure (SBP), cystine demonstrated a robust positive association (inverse variance weighting [IVW] β = 0.056, standard error [SE] = 0.010), supported by the weighted median (WM) method (β = 0.070, SE = 0.013). Conversely, serine was negatively associated with SBP (IVW β = −0.040, SE = 0.011; WM β = −0.034, SE = 0.013). For diastolic blood pressure (DBP), serine also showed a significant negative association (IVW β = −0.044, SE = 0.010; WM β = −0.042, SE = 0.013). Alanine was positively associated with low-density lipoprotein cholesterol (LDL-C) (IVW β = 0.105, SE = 0.025; WM β = 0.110, SE = 0.031). For T2DM, threonine demonstrated a positive association in the primary analysis (IVW OR = 1.165, SE = 0.061), but did not remain after multiple testing correction. WM analysis yielded a similar association (OR = 1.112, SE = 0.041). For Hypertension, cystine showed a negative association (IVW OR = 0.665, SE = 0.185), although it did not meet statistical significance after false discovery rate (FDR) adjustment.

Although three exposure—outcome pairs initially showed MR–Egger intercept p-values below 0.05, none remained significant after multiple testing corrections, indicating minimal horizontal pleiotropy. Steiger filtering confirmed that genetic instruments were primarily influencing exposures rather than outcomes in all analyses. Furthermore, excluding outlier SNPs using the MR-PRESSO method did not alter the direction or magnitude of the associations. A detailed summary of the results, including crude P-values, FDR-adjusted P-values, and pleiotropy testing, is provided in Table S5.

	N (%)		Missing value, *N*
Total	10,333		–
Women	5514 (53%)		–
Current smoker	1548 (15%)		19
Current drinker	5227 (51%)		6
CVD history	344 (3.3%)		–
Cancer history	795 (7.7%)		26
Medication history			
Hypertension medication users	2950 (28.5%)		–
Hyperlipidemia medication users	1947 (18.8%)		–
Hyperglycemic medication users	655 (6.3%)		–
	Mean (SD)	Range	
Age, years	59.6 (10.0)	34.0–75.0	–
Body mass index, kg/m^2	23.3 (3.3)	12.7–45.9	16
Systolic blood pressure, mmHg	128.7 (18.5)	75.0–258.0	3
Diastolic blood pressure, mmHg	75.9 (11.5)	42.0–140.0	3
Plasma glucose, mg/dL	99.8 (17.3)	22.0–410.0	7
HbA1c, %	5.7 (0.6)	3.6–14.5	10
	Median (IQR)		
Total cholesterol, mg/dL	204.0 (183.0, 227.0)	68.0–660.0	2
HDL cholesterol, mg/dL	66.0 (55.0, 78.0)	21.0–224.0	1
LDL cholesterol, mg/dL	116.0 (96.6, 136.8)	14.0–295.0	4
Triglyceride, mg/dL	89.0 (65.0, 128.0)	12.0–1879.0	2
Amino acids, μM	Mean (SD)		
Alanine	353.0 (88.1)	67.8–2049.7	–
Arginine	85.9 (17.0)	26.7–406.5	–
Asparagine	47.8 (8.2)	2.1–234.8	–
Aspartic Acid	3.9 (6.7)	0.4–552.2	–
Cystine	57.3 (10.3)	22.4–260.2	–
Glutamic Acid	35.8 (18.9)	3.1–719.5	–
Glutamine	619.8 (81.6)	50.8–2217.3	–
Glycine	233.0 (68.7)	80.5–1108.4	–
Histidine	76.9 (15.5)	23.2–700.1	–
Isoleucine	59.0 (15.4)	21.6–428.2	–
Leucine	115.7 (25.7)	50.7–717.9	–
Lysine	186.1 (34.5)	65.7–949.9	–
Methionine	21.1 (6.3)	0.9–94.3	–
Phenylalanine	55.4 (10.0)	32.4–304.2	–
Proline	135.8 (45.1)	49.3–807.9	–
Serine	118.3 (31.3)	33.3–1954.6	–
Threonine	132.8 (31.1)	54.9–766.8	–
Tryptophan	49.9 (9.3)	18.1–222.8	–
Tyrosine	62.8 (13.3)	21.0–448.1	–
Valine	218.6 (42.6)	106.8–1246.7	–

Table 1. Clinical characteristics and plasma amino acid levels of the study participants. CVD, cardiovascular disease; HDL, high-density lipoprotein; IQR, interquartile range; LDL, low-density lipoprotein; SD, standard deviation.

Discussion

Using plasma AA concentrations from more than 10,000 individuals from Japan, our GWAS analysis indicated 85 loci-metabolite associations with 18 AAs. Fourty four loci–metabolites across 23 genes, all of which were previously unreported, demonstrated one-to-one associations with the phenotype, with 8 of these genes classified as enzyme-coding genes, and 27 were replicates of polymorphisms reported in previous studies of East Asian populations with smaller sample size. Heritability estimates revealed that approximately 16% of the variability in AA levels was attributable to genetic factors, underscoring the substantial genetic contribution to AA metabolism. Furthermore, our MR analysis provided evidence for causal associations of AAs with cardiometabolic traits, particularly cystine and serine with blood pressure, and alanine with LDL-C. These findings highlight the genetic underpinnings of AA metabolism and their potential influence on health outcomes.

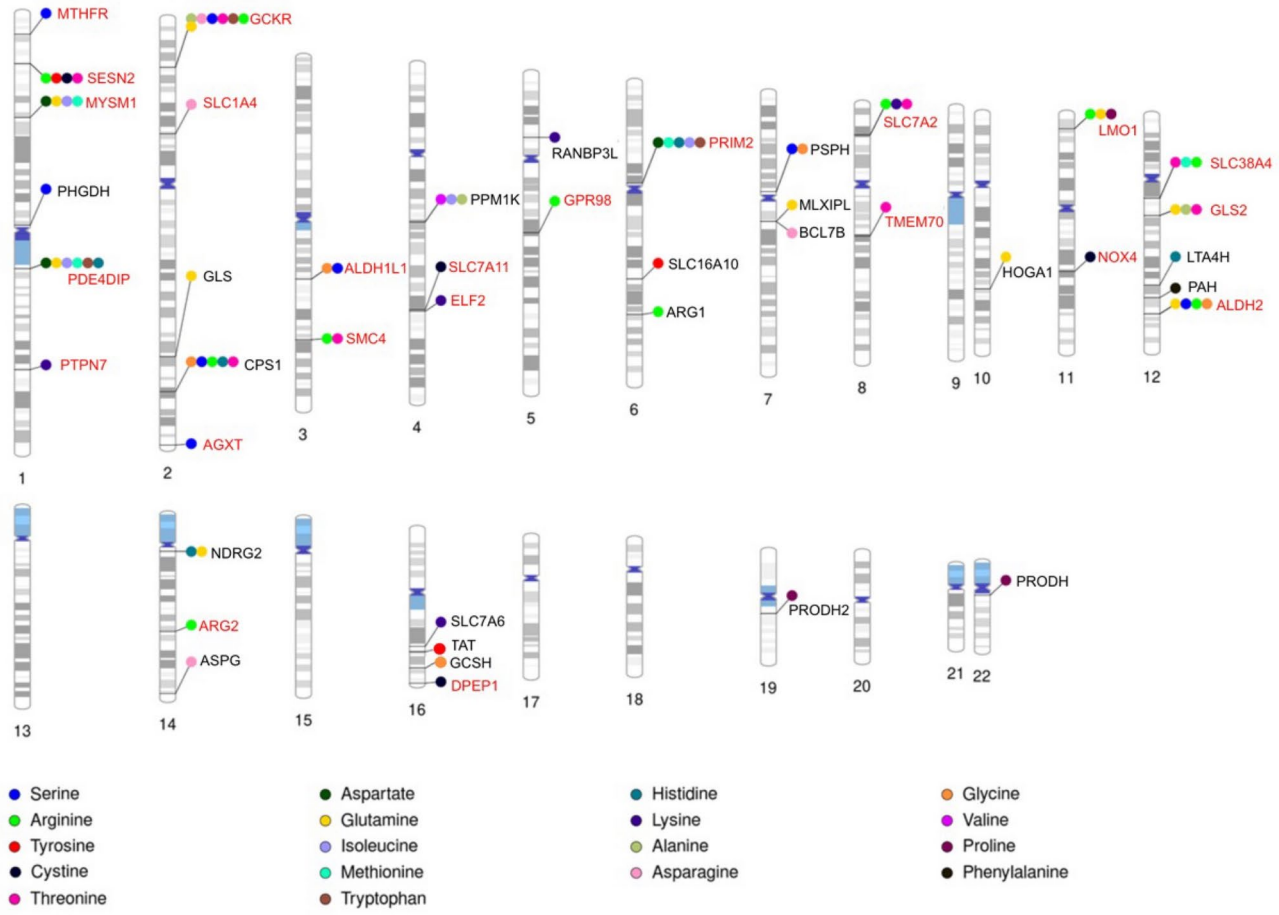


Fig. 1. Chromosomal map of the genomic risk loci associated with the 18 amino acids levels. PhenoGram (http://visualization.ritchielab.org/phenograms/plot) was used to create the map, and the novel genetic loci associated with amino acids reported in this study are depicted in red.

Comparison with the GWAS Catalog data revealed commonalities across ethnic groups, particularly for genes encoding enzymes related to AA metabolism, such as *PAH*, *PRODH*, and *PPM1 K*[21–23]. Although four loci-AA associations reported in the GWAS Catalog were not observed in this study, they were located at nearby loci through prudent analysis. Considering that they occurred during mapping of the causal loci, we believe that all peaks could be reproduced in this study, and thus, the identification of several new SNPs could be groundbreaking. Heritability estimates derived from LDSC revealed that approximately 16% of the variability in AA levels could be attributed to genetic factors. This level of heritability was comparable to estimates for certain AAs reported in Western populations[24,25], emphasizing the shared genetic influences across diverse ethnic groups.

The annotation results indicated that the potentially novel loci were not only associated with genes encoding enzymes that metabolize AAs, but also those encoding regulators, transporters, and transcription factors. Table S6 summarizes the list of loci associated with glycine, serine, and arginine, which were particularly interesting to interpret. *AGXT*, which encodes alanine–glyoxylate aminotransferase, was newly identified as having a significantly association with serine. Serine, together with glycine, are both metabolites of the AA catabolic pathway and involved in the citric acid circuit through pyruvate production[26]. They are regulated by their mutual conversion by serine hydroxylmethyltransferase with tetrahydrofolate as the substrate (Figure S1A, B). An association with glycine and serine was also detected for SNP in *ACAD10* on chromosome 12, which was attributable to LD, as the locus is in close proximity to *ALDH2*. In fact, the peak was attenuated and lost statistical significance in the GWAS results after adjustment for alcohol consumption as a covariate. This study was the first to report the association of *ARG2* with arginine. Arginine is hydrolyzed by arginase, releasing urea to produce ornithine (Figure S1-C). Arginase has two isoforms, type I and type II; the former is highly expressed in the liver and is involved in ammonia detoxification, while the latter supplies ornithine for the synthesis of polyamines, proline, and glutamate[27]. Our study not only reproduced the association of arginine with *ARG1*, but also found an association with *ARG2*, which had not been previously reported.

This study also provided unique suggestions for SNPs with possible pleiotropic associations to multiple AAs. Particularly, the directionality of the quantitative correlations and gene–trait relationships presented in Fig. 2 might indicate a true genetic common background independent of quantitative correlations in some of

Gene	Amino Acids	Top Lead SNP	REF	ALT	Chr	Position (GRCh37)	BETA	SE	FC	*P*-value	AAF	Annotation	Description	Category
MTHFR	Serine	rs1801133	G	A	1	11,856,378	−0.123	0.014	0.884	1.50E−17	0.369	Missense	Methylenetetrahydrofolate reductase	Enzyme
PTPN7	Lysine	rs3935010	C	T	1	202,138,516	0.115	0.020	1.122	1.50E−08	0.141	Non coding Transcript variant	Protein tyrosine phosphatase non-receptor type 7	Regulator
PDE4DIP	Aspartate	rs10714022	C	T	1	144,864,861	−0.384	0.031	0.681	6.40E−35	0.172	Intron	Phosphodiesterase 4D interacting protein	Regulator
	Glutamine	rs10714022	C	T	1	144,864,861	−0.304	0.031	0.738	2.40E−22	0.172	Intron		
	Isoleucine	rs10714022	C	T	1	144,864,861	0.266	0.031	1.305	1.60E−17	0.172	Intron		
	Methionine	rs10714022	C	T	1	144,864,861	0.392	0.031	1.479	4.80E−36	0.173	Intron		
	Tryptophan	rs10714022	C	T	1	144,864,861	0.249	0.031	1.283	1.30E−15	0.172	Intron		
	Histidine	rs587617572	C	T	1	144,873,410	0.255	0.031	1.291	4.00E−16	0.105	Intron		
SESN2	Threonine	rs34315986	C	G	1	28,586,376	0.155	0.026	1.168	2.80E−09	0.079	Synonymous	Sestrin 2	Regulator
	Arginine	rs74896528	C	T	1	28,598,287	0.216	0.028	1.241	2.50E−14	0.067	Missense		
	Tyrosine	rs74896528	C	T	1	28,598,287	0.198	0.028	1.219	2.40E−12	0.067	Missense		
	Cystine	rs74896528	C	T	1	28,598,287	−0.229	0.028	0.796	6.30E−16	0.067	Missense		
MYSM1	Aspartate	rs79758152	T	C	1	59,096,372	0.129	0.021	1.138	2.80E−10	0.163	Intergenic	Myb like, SWIRM and MPN domains 1	DNA-related
	Glutamine	rs79758152	T	C	1	59,096,372	0.139	0.021	1.149	1.30E−11	0.163	Intergenic		
	Isoleucine	rs79758152	T	C	1	59,096,372	−0.158	0.021	0.854	1.40E−14	0.163	Intergenic		
	Methionine	rs79758152	T	C	1	59,096,372	−0.188	0.021	0.829	7.30E−20	0.162	Intergenic		
GCKR	Tryptophan	rs1260326	T	C	2	27,730,940	−0.106	0.014	0.900	7.60E−14	0.431	Missense	Glucokinase regulator	Regulator
	Arginine	rs2911711	T	A	2	27,750,546	0.087	0.014	1.091	1.00E−09	0.425	Intergenic		
SLC1 A4	Asparagine	rs72538440	G	GC	2	65,225,088	−0.112	0.015	0.894	2.80E−13	0.703	Intron	Solute carrier family 1 member 4	Transporter
AGXT	Serine	rs117043148	C	T	2	241,810,154	0.181	0.027	1.199	3.20E−11	0.071	Intron	Alanine–glyoxylate aminotransferase	Enzyme
SMC4	Arginine	rs3851365	C	A	3	160,191,236	−0.106	0.018	0.900	5.20E−09	0.180	Intron	Structural maintenance of chromosomes 4	DNA-related
	Threonine	rs3851365	C	A	3	160,191,236	−0.112	0.018	0.894	5.80E−10	0.180	Intron		
ALDH1L1	Serine	rs35331719	C	A	3	125,907,392	−0.108	0.016	0.897	4.30E−12	0.281	Intron	Aldehyde dehydrogenase 1 family member L1	Enzyme
SLC7 A11	Cystine	rs7674749	T	A	4	139,149,032	0.127	0.014	1.135	4.30E−19	0.571	Intron	Solute carrier family 7 member 11	Transporter
ELF2	Lysine	rs11939912	G	A	4	139,938,087	−0.096	0.016	0.908	3.60E−09	0.251	Intron	E74 like ETS transcription factor 2	Transcription Factor
GPR98	Arginine	rs2367181	G	A	5	90,215,587	0.098	0.014	1.103	2.80E−12	0.439	Intron	–	-
PRIM2	Aspartate	rs12197119	T	A	6	57,281,739	0.158	0.018	1.171	1.40E−17	0.310	Intron	DNA primase subunit 2	DNA-related
	Methionine	rs12197119	T	A	6	57,281,739	−0.200	0.019	0.818	3.30E−27	0.310	Intron		
	Histidine	rs12197119	T	A	6	57,281,739	−0.151	0.019	0.860	3.70E−16	0.310	Intron		
	Isoleucine	rs77817532	A	T	6	57,294,331	0.176	0.023	1.192	2.00E−14	0.875	Intron		
	Tryptophan	rs77817532	A	T	6	57,294,331	0.145	0.023	1.156	2.60E−10	0.875	Intron		
SLC7 A2	Threonine	rs2720559	G	A	8	17,418,800	0.177	0.027	1.193	7.80E−11	0.073	Intron	Solute carrier family 7 member 2	Transporter
TMEM70	Threonine	rs62509311	A	T	8	74,907,295	0.150	0.020	1.162	4.80E−14	0.142	Intron	transmembrane protein 70	Other
LMO1	Arginine	rs2168101	C	A	11	8,255,408	−0.091	0.017	0.913	4.70E−08	0.223	Intron	LIM domain only 1	Other
	Proline	rs7952320	G	C	11	8,250,143	0.101	0.017	1.106	1.60E−09	0.779	Intron		
NOX4	Cystine	rs2289125	A	C	11	89,224,453	0.274	0.014	1.315	4.60E−85	0.540	5 Prime UTR	NADPH oxidase 4	Enzyme

Table 2. Summary of potentially novel loci associated with amino acids in this study. AAF, alternate allele frequency; ALT, alternative allele; Chr, chromosome; FC, fold change; REF, reference allele; SE, standard error; SNP, single-nucleotide polymorphism.

Genetic loci																				
Genes	rsID	Gly	Ser	Asp	Glu	Met	Thr	Asn	Ala	Pro	His	Phe	Tyr	Trp	Ile	Val	Leu	Cys	Gln	Lys	Arg
CPS1	rs1047891	0.83	0.34				-0.14				-0.14										-0.21
GCKR	rs1260326		0.16				0.15	0.08	-0.11					-0.11							
GCKR	rs2911711																		0.10		0.09
LMO1	rs2168101																		-0.11		-0.09
MYSM1	rs79758152			0.13		-0.19									-0.16				0.14		
NDRG2	rs1998848										0.23								0.11		
PDE4DIP	rs10714022			-0.38		0.39								0.25	0.27				-0.30		
PPM1K	rs10014755														0.08	0.10					
PRIM2	rs12197119			0.16		-0.20					-0.15										
PRIM2	rs77817532													0.15	0.18						
SESN2	rs74896528												0.20					-0.23			0.22
SMC4	rs3851365						-0.11														-0.11

Amino Acids	Gly	Ser	Asp	Glu	Met	Thr	Asn	Ala	Pro	His	Phe	Tyr	Trp	Ile	Val	Leu	Cys	Gln	Lys	Arg
Gly	1.00	0.50	0.01	-0.23	0.07	0.16	0.24	0.03	-0.03	-0.04	-0.08	-0.09	-0.06	-0.16	-0.15	-0.16	-0.01	0.28	0.06	0.10
Ser	0.50	1.00	0.08	-0.08	0.17	0.36	0.36	0.05	0.05	0.10	0.06	0.09	0.02	-0.04	-0.02	-0.03	0.06	0.32	0.18	0.17
Asp	0.01	0.08	1.00	0.43	0.08	0.11	0.05	0.15	0.16	0.08	0.16	0.19	0.10	0.17	0.22	0.22	0.10	0.08	0.10	0.03
Glu	-0.23	-0.08	0.43	1.00	0.18	0.12	-0.02	0.34	0.37	0.19	0.33	0.38	0.33	0.43	0.46	0.47	0.26	-0.07	0.23	0.04
Met	0.07	0.17	0.08	0.18	1.00	0.42	0.44	0.41	0.36	0.40	0.45	0.48	0.43	0.43	0.35	0.42	0.11	0.21	0.41	0.33
Thr	0.16	0.36	0.11	0.12	0.42	1.00	0.55	0.38	0.39	0.39	0.30	0.41	0.32	0.32	0.29	0.30	0.11	0.33	0.39	0.45
Asn	0.24	0.36	0.05	-0.02	0.44	0.55	1.00	0.37	0.29	0.49	0.39	0.39	0.36	0.30	0.24	0.32	0.06	0.41	0.46	0.46
Ala	0.03	0.05	0.15	0.34	0.41	0.38	0.37	1.00	0.60	0.42	0.42	0.53	0.50	0.54	0.51	0.52	0.14	0.19	0.40	0.31
Pro	-0.03	0.05	0.16	0.37	0.36	0.39	0.29	0.60	1.00	0.34	0.37	0.42	0.41	0.56	0.50	0.52	0.13	0.18	0.31	0.30
His	-0.04	0.10	0.08	0.19	0.40	0.39	0.49	0.42	0.34	1.00	0.44	0.43	0.46	0.42	0.41	0.48	0.06	0.26	0.51	0.35
Phe	-0.08	0.06	0.16	0.33	0.45	0.30	0.39	0.42	0.37	0.44	1.00	0.63	0.52	0.58	0.56	0.64	0.28	0.22	0.46	0.30
Tyr	-0.09	0.09	0.19	0.38	0.48	0.41	0.39	0.53	0.42	0.43	0.63	1.00	0.57	0.54	0.53	0.57	0.24	0.26	0.46	0.34
Trp	-0.06	0.02	0.10	0.33	0.43	0.32	0.36	0.50	0.41	0.46	0.52	0.57	1.00	0.57	0.54	0.58	0.10	0.12	0.45	0.30
Ile	-0.16	-0.04	0.17	0.43	0.43	0.32	0.30	0.54	0.56	0.42	0.58	0.54	0.57	1.00	0.84	0.88	0.18	0.16	0.47	0.30
Val	-0.15	-0.02	0.22	0.46	0.35	0.29	0.24	0.51	0.50	0.41	0.56	0.53	0.54	0.84	1.00	0.90	0.26	0.22	0.50	0.25
Leu	-0.16	-0.03	0.22	0.47	0.42	0.30	0.32	0.52	0.52	0.48	0.64	0.57	0.58	0.88	0.90	1.00	0.25	0.23	0.54	0.29
Cys	-0.01	0.06	0.10	0.26	0.11	0.11	0.06	0.14	0.13	0.06	0.28	0.24	0.10	0.18	0.26	0.25	1.00	0.23	0.23	0.17
Gln	0.28	0.32	0.08	-0.07	0.21	0.33	0.41	0.19	0.18	0.26	0.22	0.26	0.12	0.16	0.22	0.23	0.23	1.00	0.36	0.39
Lys	0.06	0.18	0.10	0.23	0.41	0.39	0.46	0.40	0.31	0.51	0.46	0.46	0.45	0.47	0.50	0.54	0.23	0.36	1.00	0.53
Arg	0.10	0.17	0.03	0.04	0.33	0.45	0.46	0.31	0.30	0.35	0.30	0.34	0.30	0.30	0.25	0.29	0.17	0.39	0.53	1.00

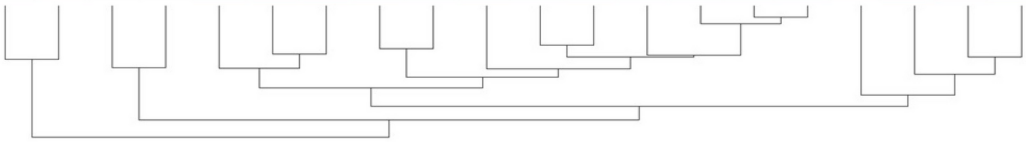


Fig. 2. List of pleiotropic loci associated with multiple amino acids and Spearman's correlation coefficients between amino acid concentrations. The values in the upper table in the figure are coefficient βs (estimated from the GWAS summary), and those in the lower table are Spearman's correlation coefficients. The dendrogram is based on the Euclidean distance calculated on the basis of Spearman's correlation coefficient. Positive associations (β, Spearman's correlation coefficient > 0) are highlighted in red, and negative associations (β, Spearman's correlation coefficient < 0) are highlighted in blue. Ala, alanine; Arg, arginine; Asn, asparagine; Asp, aspartic acid; Gln, glutamine; Glu, glutamic acid; Gly, glycine; His, histidine; Ile, isoleucine; Leu, leucine; Lys, lysine; Met, methionine; Phe, phenylalanine; Pro, proline; Ser, serine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; Val, valine.

these genes. In addition to *CPS1* described above, *PDE4DIP*, associated with five traits including serine and methionine, is another candidate gene exhibiting pleiotropy. The reports that both *CPS1* and *PDE4DIP*are associated with cardiometabolic disease[28,29] support our results that they are pleiotropic, affecting multiple amino acids, and may reflect the shared etiology of cardiometabolic disease. In addition, our study is the first to report the relationship between *PDE4DIP* and amino acids, and this relationship may be applicable to further studies of *PDE4DIP* and cardiometabolic health[29].

The two-sample MR analysis conducted in this study revealed causal relationships between plasma AA levels and several cardiometabolic traits. Alanine showed robust positive associations with LDL-C, triglyceride (TG), and hyperlipidemia, supporting its role in lipid metabolism and gluconeogenesis pathways, potentially contributing to dyslipidemia risk. In contrast, serine exhibited inverse relationships with SBP and DBP and a significant negative association with coronary artery disease (CAD), suggesting its role in mitigating cardiovascular risk via antioxidative and anti-inflammatory pathways[30]. Cystine displayed a strong positive association with SBP but showed a weak negative relationship with hypertension, although the association became insignificant after FDR correction. This inconsistency between continuous and categorical outcomes may reflect threshold-dependent effects or differing physiological mechanisms. Additionally, cystine showed a modest negative association with T2DM, possibly related to its role in redox regulation through its precursor relationship to cysteine and glutathione[31]. Threonine was modestly positively associated with T2DM, consistent with prior research linking it to insulin resistance[32]. Although the association lost significance after multiple testing corrections, it highlights a possible role in glucose metabolism that warrants further investigation. These findings, thus, underscore the importance of AAs as biomarkers and potential therapeutic targets for cardiometabolic diseases.

The well-known causal relationship between BCAAs and T2DM[33–35] was not replicated in this study. Weak instrumentation bias most likely caused incorrect estimation of many results, since the results with small *P*-values demonstrated the same direction as in previous studies. One possible explanation for the insignificant association between BCAAs and T2DM is that, unlike studies that increase power using genetic risk scores or meta-analyses, we used an SNP-by-SNP MR approach that was then combined. Although this method allows for more detailed checks for pleiotropy, it may lower overall power—especially with a moderate sample size—and could have masked weaker signals between BCAAs and T2DM.

The strengths of this study are that the associations shown in the GWAS Catalog and other GWAS data resources were well reproduced using sufficient sample size, genotyping, and whole-genome imputation using arrays specific to this population, as well as AA quantification using a reliable metabolomics platform[36,37]. The large sample size in this study, about five to ten times larger than that reported by previous Japanese studies[15], led to high detection power. Using an imputation panel specific to the Japanese population in this study might have also contributed to detecting novel loci. By comparing findings across diverse ethnic populations and leveraging metabolic pathway analyses, this study demonstrated a strong foundation of reproducibility and novelty. Despite being conducted in a single ethnic group, the similarity of plasma AA levels to those in prior studies suggested broad generalizability. Future research, including functional validation of the identified novel loci, would be critical in elucidating the biological mechanisms underlying these associations and their implications for cardiometabolic health.

Despite the large sample size and multiple analytical methods, this study has some limitations. First, because all participants were Japanese, our findings may not fully generalize to other populations. Although comparisons with prior GWAS studies suggest broadly plausible biology across ancestries, replication in multi-ethnic cohorts or meta-analyses would further solidify causality. Second, while two-sample Mendelian randomization supports the observed relationships, pleiotropy and unmeasured confounding cannot be definitively excluded. Here, we relied chiefly on univariable MR and several sensitivity analyses to capture a wide range of outcomes, but additional work employing multivariable MR[3] and genetic correlation analyses may clarify the metabolic pathways and pinpoint which amino acids exert genuine causal effects on cardiometabolic traits.

In conclusion, this study, leveraging data from over 10,000 individuals, sheds light on the genetic determinants of plasma AA levels and their causal roles in cardiometabolic health. Potentially novel genetic loci associated with AA metabolism were identified, along with robust evidence supporting causal relationships between specific AAs and cardiometabolic traits. Notably, cystine, serine, and alanine emerged as key players in blood pressure and lipid regulation, highlighting their potential as biomarkers or therapeutic targets. While the associations with T2DM and hypertension require further validation, this study demonstrates the utility of large-scale, metabolomic GWAS in advancing our understanding of the genetic and metabolic underpinnings of cardiometabolic diseases.

Methods

Study design and participants

This study used data from the baseline survey of Tsuruoka Metabolomics Cohort Study (TMCS), an ongoing prospective cohort study in Japan that aimed to identify metabolomic biomarkers for health and disorders associated with genetic and environmental factors. The survey was conducted between April 2012 and March 2015, enrolling 11,002 individuals aged 35–74 years living or working in Tsuruoka City who underwent annual health examinations in municipalities or workplaces. Details of TMCS have been previously published[37–39]. This study received approval from the Medical Ethics Committee of Keio University School of Medicine (approval number: 20110264), and written informed consent was obtained from all participants. The study protocol conformed to the ethical guidelines of the Declaration of Helsinki of 1975 and its subsequent amendments.

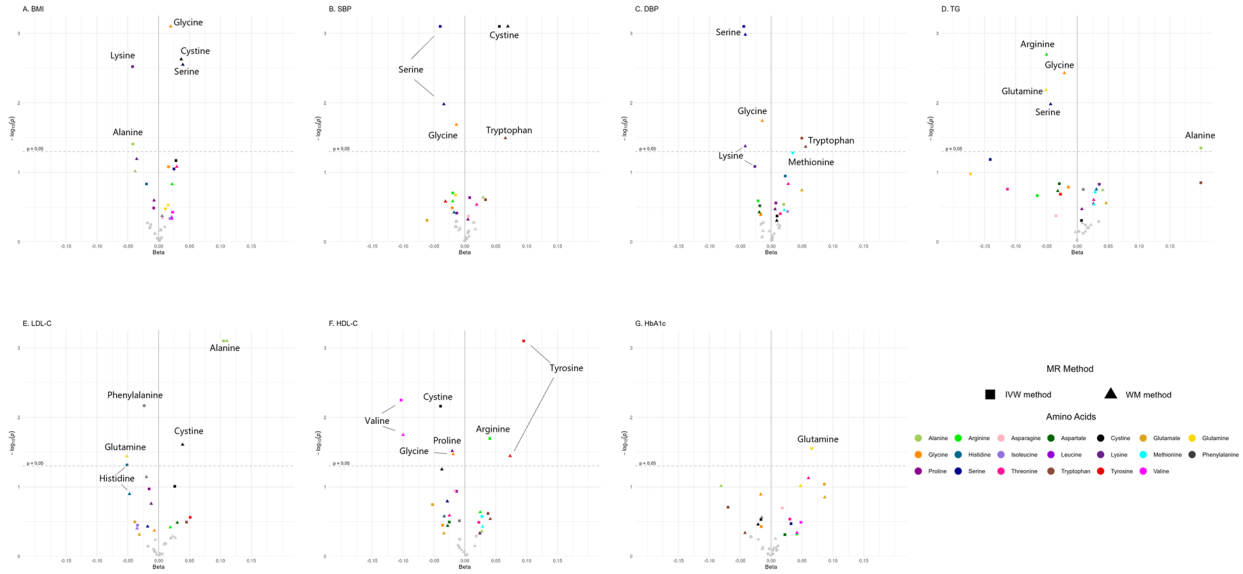
Data collection, AA quantification, genotyping, and imputation

Details have been previously reported[37,39]. All data and samples were collected at the designated health examination sites, and demographic information was collected through a standardized self-administered questionnaire. Fasting blood samples were drawn in the morning and stored at 4 °C immediately after collection, and metabolite extraction was completed within 6 h of collection to minimize the effects of metabolic reactions.

The levels of 20 common AAs (alanine, arginine, asparagine, aspartic acid, cystine, glutamic acid, glutamine, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, and valine) were measured by capillary electrophoresis time-of-flight mass spectrometry.

DNA samples were stored under strict temperature conditions before and after extraction. The Applied Biosystems™ Axiom™ Japonica Array® NEO (Life Technologies Corporation, Thermo Fisher Scientific, Waltham, MA, USA) was used for genotyping[40]. After QC procedures, prephasing was conducted using SHAPEIT2[41] followed by whole-genome imputation using IMPUTE4[42] with Japanese-specific 3.5 KJPNv2[43] and 1 KGP Phase 3[44] as reference panels, respectively. After imputation, samples with low imputation quality were excluded from the QC analysis. All of the subsequent QC processes were performed using PLINK[45]. For sample QC, those with a call rate < 0.95, PI-HAT > 0.1875 non-Japanese, and log10-transformed AA concentration ≥ 4 standard deviations from the mean were excluded, and the number of samples that passed QC for each trait is shown in Table S1. For variant QC, we started with 656,188 SNPs directly genotyped from the array. We removed 32,965 SNPs with call rate < 0.99, 10,684 SNPs with minor allele frequency < 0.01, and 20,168 SNPs with Hardy–Weinberg $P < 1 \times 10^{-5}$, noting some overlap across these filters. After imputation, 3735 SNPs failing the post-imputation Hardy–Weinberg threshold ($P < 1 \times 10^{-6}$) were excluded. As a result, 11,584,837 variants passed the final criteria (MAF ≥ 0.01, imputation information ≥ 0.7, and call rate ≥ 0.98) and were used in downstream analyses.

A. Cardiometabolic Traits (continuous)



B. Cardiometabolic Traits (categorical)

Genome-wide association

GWAS was performed using BOLT-LMM v2.3.6, a well-established software that can employ and analyze linear mixed-effects models suitable for population stratification[46]. Metabolite concentrations were log10 transformed, followed by rank-based inverse standard transformation. Optimal parameters were chosen using a Bayesian

◀**Fig. 3**. Volcano plots of Mendelian randomizations illustrating causal associations between amino acids and cardiometabolic traits. (**A**) Cardiometabolic traits as continous variables (BMI, SBP, DBP, TG, LDL-C, HDL-C, and HbA1c) (**B**) Cardiometabolic traits as categorical variables (Hypertension, Hyperlipidemia, T2DM, and CAD). The x-axis represents either the beta estimate (β) or the odds ratio (OR), and the y-axis represents the negative logarithm of the P-value. Each point represents a single SNP associated with each amino acid indicated by color coding. Associations were estimated using two MR methods: inverse variance weighting (IVW) and weighted median (WM), and points were indicated in squares and triangles, respectively. Solid vertical lines indicate null effect (β = 0 for continuous traits, OR = 1 for categorical traits), and dashed horizontal lines represent the threshold of P-value at 0.05 based on crude P-values. Points shaded in light gray represent associations with P-values > 0.5. Annotations highlight amino acids with strong associations based on crude P-values, effect size magnitude, and consistency between methods. BMI, body mass index; CAD, coronary artery disease; DBP, diastolic blood pressure; HbA1c, hemoglobin A1c; HDL-C, high-density lipoprotein-cholesterol; LDL-C, low density lipoprotein-cholesterol; SBP, systolic blood pressure; T2DM, type 2 diabetes; TG, triglyceride.

mixture model with a sparse prior. In this linear mixed-effects approach, sex, age, and the first 20 principal components (PC1–20) were specified as fixed effects, while a random effect—based on a genetic relationship matrix—captured polygenic structure. A residual term accounted for the remaining variation. Consequently, the observed phenotype was modeled as the sum of the fixed effects, the random effect, and the residual. The trait information was the standardized concentrations of 20 AAs in plasma samples. The P-value threshold was set to 5.0×10^{-8} for genome-wide significance. As a sensitivity check for potential confounding, we also repeated the GWAS after excluding individuals who reported antihypertensive, lipid-lowering, or antidiabetic medication use.

Annotations

FUMA GWAS was used for the fine-mapping of GWAS summary data[47]. The settings were as follows:

- Reference panel: 1 KG EAS (1 KGP Phase 3).
- Reference panel for linkage disequilibrium (LD) analysis: 1 KG EAS.
- LD relationship threshold for defining individual significant SNPs: $r^2 = 0.6$.
- LD relationship threshold for defining lead SNPs: $r^2 = 0.1$.

Results were compared with previously reported variants using database information from the Japanese Multi Omics Reference Panel (jMorp, Japan) created by TMM, Japan and previous reports from Japanese cohorts[15]. Subsequently, previous studies on non-Japanese populations were assessed, followed by extensive comparisons of our results to those reported in the GWAS Catalog (EMBL-EBI and NHGRI, USA). The effects of the expression quantitative trait loci and LD were examined to see if they were not caused by neighboring genes (Fig. 4). The GTEx portal (GTEx v.8) was used to evaluate expression quantitative trait loci, and dbSNP and Metascape (v3.5.20240101)[48] were used for variant information. Kyoto Encyclopedia of Genes and Genomes was employed to assess metabolic pathways to annotate phenotypes, and Spearman's correlation coefficient was calculated to determine the quantitative correlation among AAs using R version 4.1.2.

Heritability analysis

Obtained GWAS summaries were used to estimate the heritability for each trait using LDSC; h2 was calculated for each AA using LD scores calculated from the 1 KGP Phase 3 East Asian population[44] using LDSC. Each calculated h^2 was weighted using its SE to assess overall trends. Cochran's Q statistic and the I^2 statistic were calculated to examine the variation in effect sizes for each AA for heterogeneity assessment.

Mendelian randomization

We conducted two-sample Mendelian randomization (MR) using summary statistics from our in-house GWAS of plasma amino acids (exposures) and publicly available outcome GWAS data (Table S7) obtained through the IEU OpenGWAS project. Variants were selected if they showed F-statistics ≥ 10 (indicating strong instruments) and $P < 1 \times 10^{-6}$ for association with each amino acid, and we enforced independence by excluding SNPs in linkage disequilibrium ($R^2 < 0.01$ within 500 kb). Effect alleles were then harmonized between the exposure and outcome datasets, retaining only SNPs present in both. All analyses were implemented using the "TwoSampleMR" package[49] in R. Causal estimates were obtained with IVW and WM methods[12]. Horizontal pleiotropy was assessed via MR–Egger (intercept test)[50] and MR–PRESSO (10,000 bootstraps) to detect and remove outlier SNPs. Additionally, Steiger filtering was applied to confirm that each instrumental variable (IV) explained more variance in the amino acid exposure than in the outcome, thereby reducing the risk of reverse causation. Multiple comparisons were controlled at a false discovery rate of 5% using the Benjamini–Hochberg method. Together, these steps strengthen the validity of our MR assumptions and allow for robust inferences regarding the causal effects of amino acids on cardiometabolic traits.

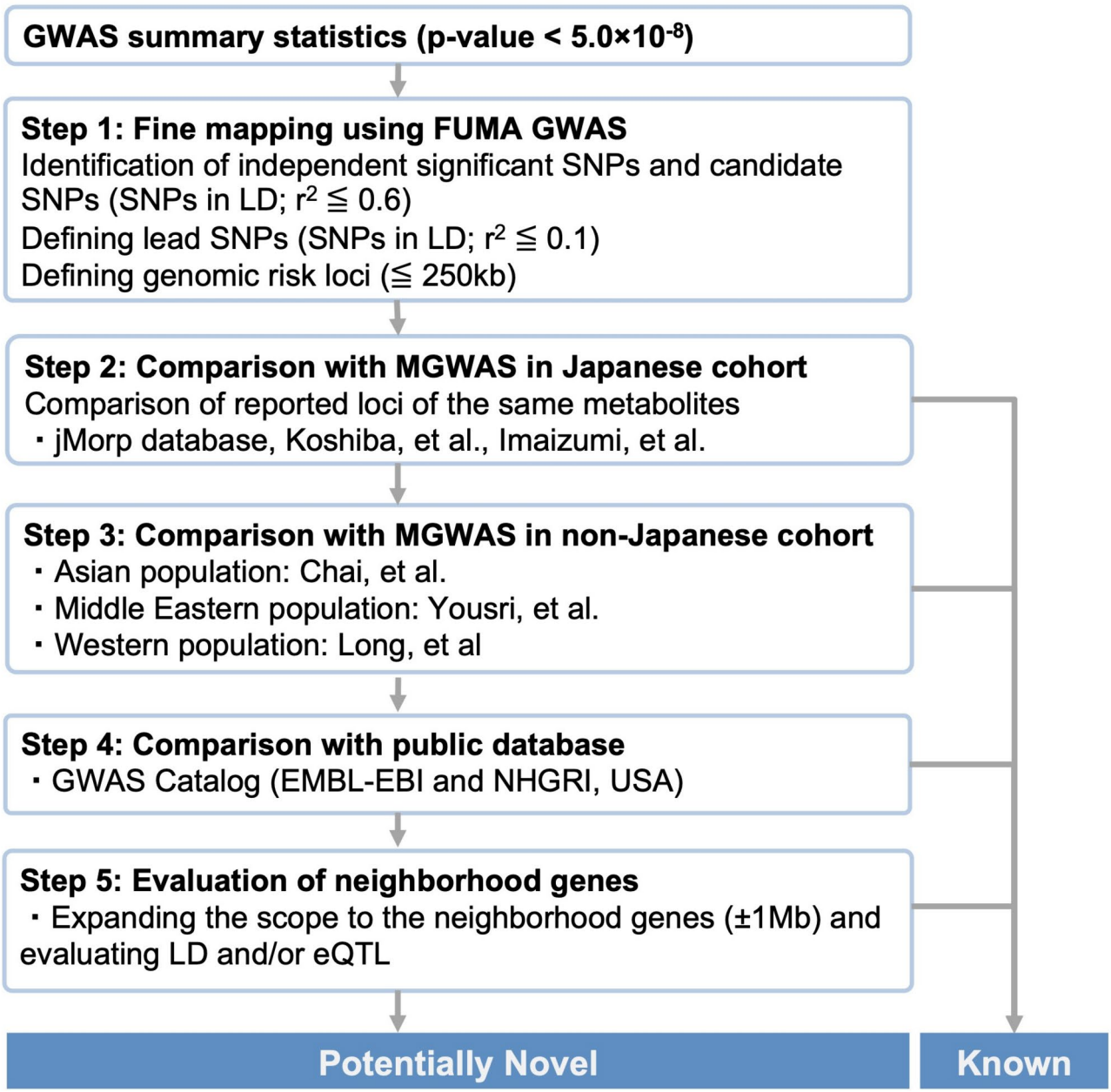


Fig. 4. Flowchart for comparing the GWAS summary data of this study with previous studies. eQTL, expression quantitative trait locus; GWAS, genome-wide association study; LD, link disequilibrium; MGWAS, metabolite genome-wide association study; SNP, single nucleotide polymorphism; TMM, Tohoku Medical Megabank.

Data availability

All relevant data are within the manuscript and its Supplementary Information. Other data could not be made publicly available because the study participants did not consent to make their information freely accessible. Based on these findings, the Ethics Committee for TMCS (which includes representatives of Tsuruoka citizens, the administration of Tsuruoka City, a lawyer, and expert advisers) strictly inhibits any public data sharing. However, the datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 9 December 2024; Accepted: 16 April 2025
Published online: 25 April 2025

References

1. Buergel, T. et al. Metabolomic profiles predict individual multidisease outcomes. *Nat. Med.* **28**, 2309–2320 (2022).

2. Aderemi, A. V., Ayeleso, A. O., Oyedapo, O. O. & Mukwevho, E. Metabolomics: A scoping review of its role as a tool for disease biomarker discovery in selected non-communicable diseases. *Metabolites* **11**, 418 (2021).
3. Lin, C. et al. The causal associations of Circulating amino acids with blood pressure: a Mendelian randomization study. *BMC Med.* **20**, 414 (2022).
4. Lieu, E. L., Nguyen, T., Rhyne, S. & Kim, J. Amino acids in cancer. *Exp. Mol. Med.* **52**, 15–30 (2020).
5. Vanweert, F., Schrauwen, P. & Phielix, E. Role of branched-chain amino acid metabolism in the pathogenesis of obesity and type 2 diabetes-related metabolic disturbances BCAA metabolism in type 2 diabetes. *Nutr. Diabetes*. **12**, 35 (2022).
6. Bar, N. et al. A reference map of potential determinants for the human serum metabolome. *Nature* **588**, 135–140 (2020).
7. Hagenbeek, F. A. et al. Heritability estimates for 361 blood metabolites across 40 genome-wide association studies. *Nat. Commun.* **11**, 39 (2020).
8. Winterton, A. et al. Associations of loneliness and social isolation with cardiovascular and metabolic health: a systematic review and meta-analysis protocol. *Syst. Rev.* **9**, 102 (2020).
9. Long, T. et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat. Genet.* **49**, 568–578 (2017).
10. Chen, Y. et al. Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human diseases. *Nat. Genet.* **55**, 44–53 (2023).
11. Imaizumi, A. et al. Genetic basis for plasma amino acid concentrations based on absolute quantification: a genome-wide association study in the Japanese population. *Eur. J. Hum. Genet.* **27**, 621–630 (2019).
12. Sanderson, E. et al. Mendelian randomization. *Nat. Rev. Methods Prim.* **2**, 1–21 (2022).
13. He, M. et al. Causal relationship between human blood metabolites and risk of ischemic stroke: a Mendelian randomization study. *Front. Genet.* **15**, 1333454 (2024).
14. Wang, Z. & Yang, Q. The causal relationship between human blood metabolites and the risk of visceral obesity: a Mendelian randomization analysis. *Lipids Health Dis.* **23**, 39 (2024).
15. Koshiba, S. et al. Identification of critical genetic variants associated with metabolic phenotypes of the Japanese population. *Commun. Biol.* **3**, 662 (2020).
16. Sakaue, S. et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53**, 1415–1424 (2021).
17. Chai, J. F. et al. Associations with metabolites in Chinese suggest new metabolic roles in Alzheimer's and Parkinson's diseases. *Hum. Mol. Genet.* **29**, 189–201 (2020).
18. Shimizu, F. et al. Plasma levels of various amino acids and their changes upon protein uptakes in Japanese young and old men and women. *Integr. Food Nutr. Metab.* **4** (2017).
19. Tadaka, S. et al. jMorp: Japanese multi-omics reference panel update report 2023. *Nucleic Acids Res.* **52**, D622–D632 (2024).
20. Yousri, N. A. et al. Whole-exome sequencing identifies common and rare variant metabolic QTLs in a middle Eastern population. *Nat. Commun.* **9**, 333 (2018).
21. Dobrowolski, S. F., Phua, Y. L., Vockley, J., Goetzman, E. & Blair, H. C. Phenylketonuria oxidative stress and energy dysregulation: emerging pathophysiological elements provide interventional opportunity. *Mol. Genet. Metab.* **136**, 111–117 (2022).
22. Bogner, A. N. & Tanner, J. J. Structure-affinity relationships of reversible proline analog inhibitors targeting proline dehydrogenase. *Org. Biomol. Chem.* **20**, 895–905 (2022).
23. Hamilton, A., Eliasson, L. & Knudsen, J. G. Amino acids and the changing face of the α-cell. *Peptides* **166**, 171039 (2023).
24. Tanha, H. M., Sathyanarayanan, A., International Headache Genetics Consortium & & Nyholt, D. R. Genetic overlap and causality between blood metabolites and migraine. *Am. J. Hum. Genet.* **108**, 2086–2098 (2021).
25. Liu, S. et al. Genome-wide association study of maternal plasma metabolites during pregnancy. *Cell. Genom.* **4**, 100657 (2024).
26. Tomé, D. Amino acid metabolism and signalling pathways: potential targets in the control of infection and immunity. *Nutr. Diabetes* **11**, 20 (2021).
27. Canè, S., Geiger, R. & Bronte, V. The roles of arginases and arginine in immunity. *Nat. Rev. Immunol.* https://doi.org/10.1038/s41577-024-01098-2 (2024).
28. Zhang, L., Zou, Y., Lu, Y., Li, Z. & Gao, F. Unraveling the therapeutic potential of carbamoyl phosphate synthetase 1 (CPS1) in human diseases. *Bioorg. Chem.* **130**, 106253 (2023).
29. Mani, A. PDE4DIP in health and diseases. *Cell. Signal.* **94**, 110322 (2022).
30. Wu, X. et al. Glycine-serine-threonine metabolic axis delays intervertebral disc degeneration through antioxidant effects: An imaging and metabonomics study. *Oxid. Med. Cell. Longev.* **2021**, 5579736 (2021).
31. Asantewaa, G. & Harris, I. S. Glutathione and its precursors in cancer. *Curr. Opin. Biotechnol.* **68**, 292–299 (2021).
32. Liu, Y. et al. Dietary intakes of methionine, threonine, lysine, arginine and histidine increased risk of type 2 diabetes in Chinese population: does the mediation effect of obesity exist? *BMC Public. Health* **23**, 1551 (2023).
33. White, P. J. & Newgard, C. B. Branched-chain amino acids in disease. *Science* **363**, 582–583 (2019).
34. Lotta, L. A. et al. Genetic predisposition to an impaired metabolism of the branched-chain amino acids and risk of type 2 diabetes: A Mendelian randomisation analysis. *PLoS Med.* **13**, e1002179 (2016).
35. Ramzan, I. et al. The association between Circulating branched chain amino acids and the Temporal risk of developing type 2 diabetes mellitus: A systematic review & meta-analysis. *Nutrients* **14**, 4411 (2022).
36. Harada, S. et al. Reliability of plasma Polar metabolite concentrations in a large-scale cohort study using capillary electrophoresis-mass spectrometry. *PLoS One* **13**, e0191230 (2018).
37. Iida, M. et al. Profiling of plasma metabolites in postmenopausal women with metabolic syndrome. *Menopause* **23**, 749–758 (2016).
38. Harada, S. et al. Study profile of the Tsuruoka metabolomics cohort study (TMCS). *J. Epidemiol. Adv.* JE20230192 (2024).
39. Sakurai-Yageta, M. et al. Japonica array NEO with increased genome-wide coverage and abundant disease risk SNPs. *J. Biochem.* **170**, 399–410 (2021).
40. Delaneau, O., Zagury, J. F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
41. Bycroft, C. et al. The UK biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
42. Tadaka, S. et al. 3.5KJPNv2: an allele frequency panel of 3552 Japanese individuals including the X chromosome. *Hum. Genome Var.* **6**, 1–9 (2019).
43. Byrska-Bishop, M. et al. High-coverage whole-genome sequencing of the expanded 1000 genomes project cohort including 602 trios. *Cell* **185**, 3426–3440e19 (2022).
44. Purcell, S. et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
45. Loh, P. R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed model association for biobank-scale data sets. *Nat. Genet.* **50**, 906–908 (2018).
46. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
47. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).
48. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, e34408 (2018).
49. Zhu, X. Mendelian randomization and pleiotropy analysis. *Quant. Biol.* **9**, 122–132 (2021).

50. Wolfe, D., Dudek, S., Ritchie, M. D. & Pendergrass, S. A. Visualizing genomic information across chromosomes with phenogram. *BioData Min.* **6**, 18 (2013).

Acknowledgements
We would like to thank the residents of Tsuruoka city for their interest in our study and the members of the TMCS team for their commitment to this project. We also would like to thank Japan Society for the Promotion of Science (Grant Number JPMJMS2023) for the research support.

Author contributions
Ryota Toki was in charge of conceptualisation, formal analysis, investigation, methodology, writing the original draft, reviewing, and editing. Sotaro Fushiki and Shun Kojima were in charge of formal analysis, visualisation, methodology. Yoichi Sutoh and Atsushi Shimizu were in charge of supervision, formal analysis, review and editing. Yayoi Otsuka-Yamasaki and Hideki Ohmomo were in charge of formal analysis, review and editing. Sei Harada, Miho Iida, Aya Hirata, Naoko Miyagawa, Minako Matsumoto, Shun Edagawa, Atsuko Miyake, Kazuyo Kuwabara, Akiyoshi Hirayama, Masahiro Sugimoto, Tomoyoshi Soga, Kengo Kinoshita, Mika Sakurai-Yageta, and Gen Tamiya were in charge of data curation, review and editing. Asako Sato and Kaori Amano were in charge of data curation. Masaru Tomita and Kazuharu Arakawa were in charge of funding acquisition, review and editing. Tomonori Okamura was in charge of supervision, review and editing. Toru Takebayashi was in charge of funding acquisition, project administration, supervision, review and editing.

Funding
This study has been funded by the Yamagata Prefectural Government (http://www.pref.yamagata.jp/), the city of Tsuruoka (https://www.city.tsuruoka.lg.jp/), and Japan Society for the Promotion of Science (http://www.jsps.go.jp/) under Grant Numbers JP24390168, JP15H04778, and JP25670303. This work was also supported by JST-OPERA Program Grant Number JPMJOP1842, Japan with joint research funds from Kinoca Minolta, Inc. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Declarations

Competing interests
Toru Takebayashi received research funds from the Yamagata Prefecture Government, City of Tsuruoka, Japan Society for the Promotion of Science under the Grant Numbers JP24390168, JP15H04778, JP25670303, and JST-OPERA Program Grant Number JPMJOP1842 with joint research funds from Konica Minolta, Inc. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.All other authors have no competing interests. All other authors have no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-98992-z.

Correspondence and requests for materials should be addressed to T.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.