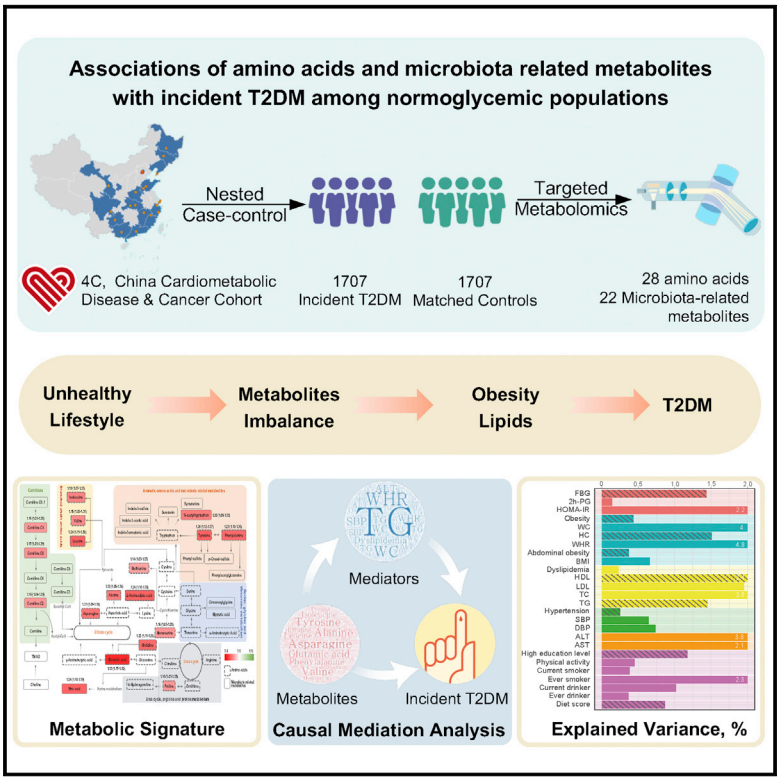
Cell Reports
Medicine

Article

Amino acids, microbiota-related metabolites, and the risk of incident diabetes among normoglycemic Chinese adults: Findings from the 4C study

Graphical abstract



Highlights

- Amino acids (AA) and microbiota-related metabolites (MRM) related to incident T2DM
- Systematic AA and MRM changes profile at normoglycemic stage before T2DM onset
- Obesity and lipids mediated linkages between AAs or MRMs and T2DM
- Risk factors explained limited metabolite variances among normoglycemic individuals

Authors

Shuangyuan Wang, Mian Li, Hong Lin, ..., Yufang Bi, Jieli Lu, for the 4C Study Group

Correspondence

gning@sibs.ac.cn (G.N.),
wqingw61@163.com (W.W.),
byf10784@rjh.com.cn (Y.B.),
jielilu@hotmail.com (J.L.)

In brief

Wang et al. report a systematic amino acids and microbiota-related metabolites change profile before T2DM onset. Causal mediation analyses indicate significant mediation linkages through obesity and lipids. Variances explained in serum metabolites are modestly limited in the comprehensive catalog of risk factor–metabolite–diabetes associations.

Wang et al., 2022, Cell Reports Medicine 3, 100727
September 20, 2022
https://doi.org/10.1016/j.xcrm.2022.100727

CellPress

Cell Reports Medicine

CellPress
OPEN ACCESS

Article

Amino acids, microbiota-related metabolites, and the risk of incident diabetes among normoglycemic Chinese adults: Findings from the 4C study

Shuangyuan Wang,[1,2,31] Mian Li,[1,2,31] Hong Lin,[1,2,31] Guixia Wang,[3,31] Yu Xu,[1,2,31] Xinjie Zhao,[4] Chunyan Hu,[1,2] Yi Zhang,[1,2] Ruizhi Zheng,[1,2] Ruying Hu,[5] Lixin Shi,[6] Rui Du,[1,2] Qing Su,[7] Jiqiu Wang,[1,2] Yuhong Chen,[1,2] Xuefeng Yu,[8] Li Yan,[9] Tiange Wang,[1,2] Zhiyun Zhao,[1,2] Ruixin Liu,[1,2] Xiaolin Wang,[4] Qi Li,[4] Guijun Qin,[10] Qin Wan,[11] Gang Chen,[12] Min Xu,[1,2] Meng Dai,[1,2] Di Zhang,[1,2] Xulei Tang,[13] Zhengnan Gao,[14] Feixia Shen,[15] Zuojie Luo,[16] Yingfen Qin,[16] Li Chen,[17]

(Author list continued on next page)

[1]Department of Endocrine and Metabolic Diseases, Shanghai Institute of Endocrine and Metabolic Diseases, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[2]Shanghai National Clinical Research Center for Endocrine and Metabolic Diseases, Key Laboratory for Endocrine and Metabolic Diseases of the National Health Commission of the PR China, Shanghai National Center for Translational Medicine, Ruijin Hospital, Shanghai Jiao-Tong University School of Medicine, 197 Ruijin 2nd Rd, Shanghai 200025, China
[3]The First Hospital of Jilin University, Changchun, China
[4]Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China
[5]Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China
[6]Affiliated Hospital of Guiyang Medical College, Guiyang, China
[7]Xinhua Hospital Affiliated to Shanghai Jiaotong University School of Medicine, Shanghai, China
[8]Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[9]Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China
[10]The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China
[11]The Affiliated Hospital of Southwest Medical University, Luzhou, China
[12]Fujian Provincial Hospital, Fujian Medical University, Fuzhou, China
[13]The First Hospital of Lanzhou University, Lanzhou, China
[14]Dalian Municipal Central Hospital, Dalian, China
[15]The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China
[16]The First Affiliated Hospital of Guangxi Medical University, Nanning, China

(Affiliations continued on next page)

SUMMARY

Although previous studies suggest that amino acids (AAs) and microbiota-related metabolites (MRMs) are associated with type 2 diabetes mellitus (T2DM), the results remain unclear among normoglycemic populations. We test 28 serum AAs and 22 MRMs in 3,414 subjects with incident diabetes and matched normoglycemic controls from the China Cardiometabolic Disease and Cancer Cohort (4C) Study. In fully adjusted logistic regression models, per SD increment of branched-chain AAs, aromatic AAs, asparagine, alanine, glutamic acid, homoserine, 2-aminoadipic acid, histidine, methionine, and proline are positively associated with incident T2DM. In the MRM panel, serum carnitines, N-acetyltryptophan, and uric acid are positively associated with incident T2DM. Causal mediation analyses indicate 34 significant causal mediation linkages, with 88.2% through obesity and lipids. Variances explained in the serum metabolites are modestly limited in the comprehensive catalog of risk factor–metabolite–diabetes associations. These findings reveal that systematic AAs and MRMs change profile before T2DM onset and support a potential role of metabolic alterations in the pathogenesis of diabetes.

INTRODUCTION

Diabetes has become a major influence on global health and economic burden, and also a main cause of death and disability worldwide.[1] Globally, the number of people with diabetes mellitus has quadrupled in the past 3 decades. Approximately 1 in 11 adults (463 million adults) worldwide had diabetes in 2019. The International Diabetes Federation estimates that there will

Yanan Huo,[18] Qiang Li,[19] Zhen Ye,[5] Yinfei Zhang,[20] Chao Liu,[21] Youmin Wang,[22] Shengli Wu,[23] Tao Yang,[24] Huacong Deng,[25] Jiajun Zhao,[26] Shenghan Lai,[27] Yiming Mu,[28] Lulu Chen,[29] Donghui Li,[30] Guowang Xu,[4] Guang Ning,[1,2,*] Weiqing Wang,[1,2,*] Yufang Bi,[1,2,*] and Jieli Lu[1,2,32,*] for the 4C Study Group

[17]Qilu Hospital of Shandong University, Jinan, China
[18]Jiangxi Provincial People's Hospital Affiliated to Nanchang University, Nanchang, China
[19]The Second Affiliated Hospital of Harbin Medical University, Harbin, China
[20]Central Hospital of Shanghai Jiading District, Shanghai, China
[21]Jiangsu Province Hospital on Integration of Chinese and Western Medicine, Nanjing, China
[22]The First Affiliated Hospital of Anhui Medical University, Hefei, China
[23]Karamay Municipal People's Hospital, Xinjiang, China
[24]The First Affiliated Hospital of Nanjing Medical University, Nanjing, China
[25]The First Affiliated Hospital of Chongqing Medical University, Chongqing, China
[26]Shandong Provincial Hospital Affiliated to Shandong University, Jinan, China
[27]Johns Hopkins University School of Medicine, Baltimore, MD, USA
[28]Chinese People's Liberation Army General Hospital, Beijing, China
[29]Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[30]Department of Gastrointestinal Medical Oncology, University of Texas MD Anderson Cancer Center, Houston, TX, USA
[31]These authors contributed equally
[32]Lead contact
*Correspondence: gning@sibs.ac.cn (G.N.), wqingw61@163.com (W.W.), byf10784@rjh.com.cn (Y.B.), jielilu@hotmail.com (J.L.)
https://doi.org/10.1016/j.xcrm.2022.100727

be 578 million adults with diabetes by 2030, and 700 million by 2045.[2] China currently has the largest number of patients with diabetes in the world, with 129.8 million (12.8%) adults with diabetes and 356.9 million (35.2%) with prediabetes in 2017.[3] Health care cost for diabetes is a significant burden on the economy, as the total diabetes-related health cost reached 109 billion international dollars in 2019.[2]

Screening and early prediction of hyperglycemia are particularly important in the prevention and treatment of diabetes and its complications.[4,5] Advances in metabolomics allowing studies of small molecules, metabolites, have opened a new research venture for biomarker studies and precision medicine.[6] The human blood metabolome provides a comprehensive readout of human physiology obtained through assessment of hundreds of small circulating molecules, which reflect the influences and interactions of genetics, lifestyle, environment, medical treatment, and microbial activity.[6]

A growing number of cross-sectional or prospective studies have shown that amino acids (AAs) are likely to play an important role in the development of type 2 diabetes mellitus (T2DM).[7] In 2011, using the method of targeted liquid chromatography-tandem mass spectrometry (LC-MS/MS), branched-chain amino acids (BCAAs) and aromatic AAs (AAAs) were reported as predictors of the future development of T2DM in the Framingham Offspring Study,[8] and were later verified in other European[9–11] and Chinese populations.[12–14]

Recent studies have highlighted the significance of the microbiome in human health and diseases. Changes in the metabolites produced by microbiota, e.g., trimethlylamine oxide (TMAO), has been implicated in diabetes but the results are conflicting.[15–17] Diabetes has also been positively associated with several carnitine-related metabolites,[18] long chain acylcarnitines,[19] tryptophan, kynurenine-pathway metabolites (kynurenine, kynurenate, xanthurenate, and quinolinate), and indolelactate.[20] On the other hand, indolepropionate was inversely associated with T2DM risk.[20]

However, most of these studies were conducted in European populations and very few studies on microbiota-related metabolites (MRMs) and risk of diabetes have been conducted in prospective cohorts of Chinese populations. In addition, many of the preceding studies were conducted among mixed populations with normal glucose regulation (NGR) and impaired glucose regulation at baseline, which make it difficult to dissect whether the metabolic changes are the cause or the consequence of early dysglycemia. Furthermore, few metabolomic studies have detailed information on modifiable clinical and risk factors to establish actionable insights for the prevention and management of T2DM.

To overcome the above limitations and to shed light on the shared etiology and drivers of T2DM, using a nested case-control design, we tested 28 AA species and 22 MRM species in human serum using ultra performance liquid chromatography coupled to tandem mass spectrometry (UPLC-MS/MS), which confers accurate quantitation. We established a comprehensive catalog of risk factor–metabolite–diabetes associations among individuals with NGR from a nationwide prospective Chinese cohort. We sought to characterize metabolites and metabolic pathways that are most strongly associated with the onset of T2DM and associations of the identified metabolic targets with modifiable clinical and other risk factors.

RESULTS

Study population

We conducted a nested case-control study of 1,707 matched case-control pairs within the China Cardiometabolic Disease and Cancer Cohort (4C) Study, a nationwide, population-based, prospective cohort study with up to 5 years of follow-up.[5,21,22] During 2011–2012, 193,846 individuals (age >40 years) were enrolled from local resident registration systems of 20 communities from various geographic regions in China. During a median follow-up of 3.03 (interquartile range 2.87–3.24) years, 170,240 participants (87.8%) were followed up. Among 54,807 subjects defined as NGR based on 75 g oral glucose tolerance tests (OGTT) at baseline, 1,864 developed diabetes.[5,21,22] After excluding 157 individuals with missing serum samples, a final

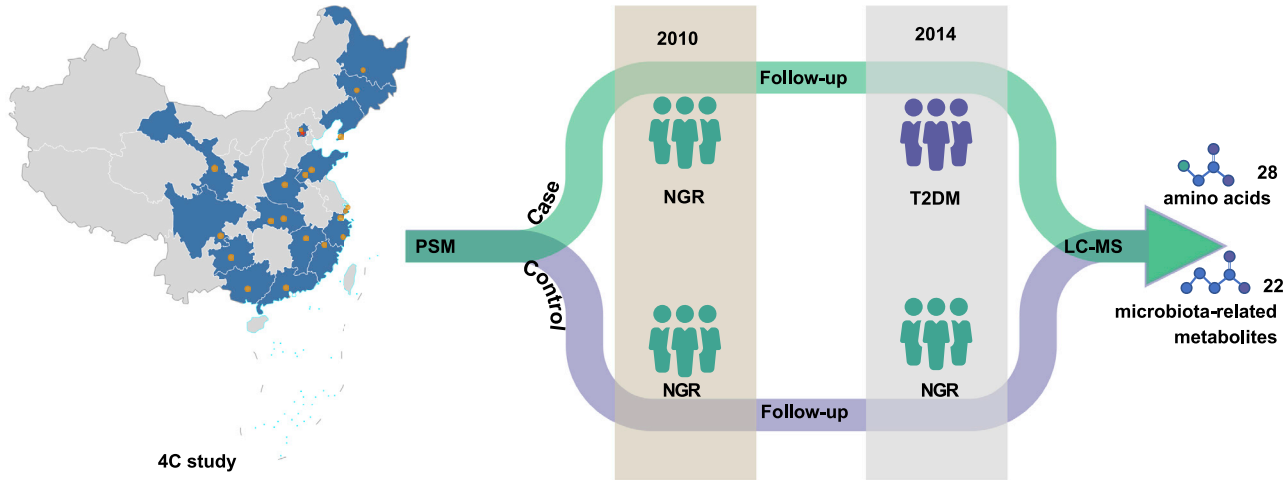


Figure 1. Flowchart of the nested case-control study

Cases (n = 1,707) were incident diabetes subjects with normal glucose regulation (NGR) at baseline and type 2 diabetes mellitus (T2DM) at follow-up. Propensity score matching (PSM) matched controls (n = 1,707) were NGR at both baseline and follow-up. Targeted metabolomics based on liquid chromatography-mass spectrometry (LC-MS) quantified 28 amino acids and 22 microbiota-related metabolites in baseline serum samples.

number of 1,707 incident diabetes cases were included (Figure 1). The control group of 1,707 NGR individuals at baseline was selected using propensity score matching (PSM) with a logistic model that included age, gender, BMI, and fasting plasma glucose (FPG).[8,21,23] The study protocol was approved by the Institutional Review Board of Ruijin Hospital affiliated to the Shanghai Jiao Tong University School of Medicine. Informed consent was obtained from each study participant.

Baseline characteristics

In this nested case-control study of 1,707 incident diabetes case subjects and 1,707 NGR individuals, 1,386 (40.6%) were men. The mean ± SD age of the whole study population was 57.56 ± 8.87 years (Table S1). In addition to age, gender, BMI, and FPG matched under PSM, other baseline characteristics, including smoking status, alcohol intake, physical activity, diet habit, and education status, were also well-matched between case subjects and control subjects. Case subjects showed higher levels of 2-h post prandial glucose, triglycerides (TG), total cholesterol (TC), aspartate aminotransferase (AST), alanine aminotransferase (ALT), and Homeostatic Model Assessment for Insulin Resistance (HOMA-IR) than control subjects. Family history of diabetes was reported in 11.19% of cases and 6.74% of controls. No difference was observed for low-density lipoprotein (LDL) cholesterol between the two groups. The baseline characteristics of the study participants have previously been described in detail.[21]

Distribution of AAs and MRMs

Fasting serum samples collected at baseline were analyzed for AAs and MRMs using the UPLC-MS/MS as described previously.[24,25] A total of 28 AA species were investigated. The greatest contributors were lysine, alanine, and glutamine, whereas 2-aminoadipic acid, homoserine, and γ-aminobutyric acid were present in relatively low concentrations (Figure 2A). Case subjects had higher levels of lysine, alanine, proline, valine, tryptophan, glutamic acid, leucine, ornithine, tyrosine, phenylalanine, histidine, a-aminobutyric acid, isoleucine, aspartic acid, methionine, 2-aminoadipic acid, and γ-aminobutyric acid than control subjects (all $p < 0.05$) (Figure 2A, Table S2).

In MRMs, uric acid comprised the greatest concentration, with the median concentration of 398.01μmol/L for the entire study population. The median concentration of plasma uric acid was higher in the diabetes cases (414.62 μmol/L) than in the control subjects (379.36 μmol/L, $p < 0.001$). The levels of carnitine C0, choline, and TMAO were quite high, while the concentrations of carnitine C5, carnitine C6, and N-acetyltryptophan were relatively low among the MRMs. Cases showed significantly higher concentrations of carnitine C0, choline, carnitine C2, phenyl sulfate, kynurenine, carnitine C3, carnitine C4, carnitine C8, carnitine C5, carnitine C6, and N-acetyltryptophan, and lower levels of p-Cresol sulfate, hippuric acid, indole-3-propionic acid, and cinnamoylglycine compared with controls (all $p < 0.05$) (Figure 2B, Table S2).

Associations of metabolites with major clinical parameters

Spearman correlation analysis revealed that baseline serum AAs and MRMs were correlated with a range of biochemical measurements and metabolic parameters (Figure 2C). For example, BCAAs (leucine, isoleucine, and valine), AAAs (tryptophan, phenylalanine, and tyrosine), acyl-carnitines (C0, C3, C4, and C5), alanine, glutamic acid, lysine, 2-aminoadipic acid, cystine, proline, and kynurenine were mostly positively associated with insulin resistance (HOMA-IR), body composition (waist circumference [WC], BMI, and waist-to-hip ratio [WHR]), liver function (ALT, AST), blood pressure (diastolic blood pressure [DBP], systolic blood pressure [SBP]), and lipid metabolism (TC, TG, and LDL), but inversely associated with high-density lipoprotein (HDL). Arginine, aspartic acid, carnitine C6, carnitine C8, and choline were positively associated with body composition, liver function, blood pressure, and lipid metabolism (TC, TG, LDL, and HDL). MRMs including indole-3-propionic acid, p-cresol sulfate, hippuric acid, serotonin, cinnamoylglycine,

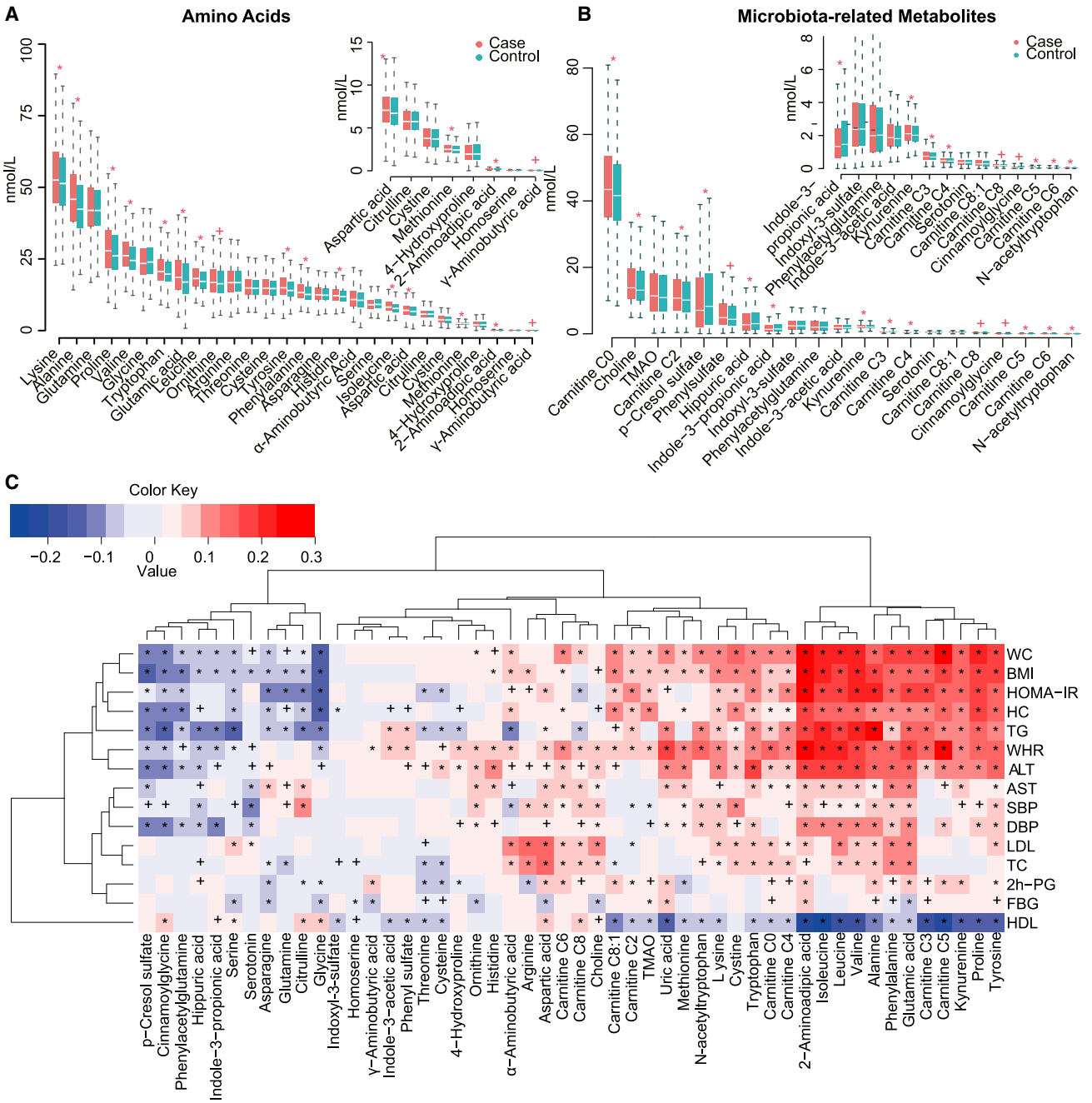


Figure 2. Serum amino acids and microbiota-related metabolites distribution among case subjects (n = 1,707) and control subjects (n = 1,707) and their correlation with clinical parameters

(A and B) Serum amino acid and microbiota-related metabolite levels in case and control groups. Results are shown as boxes denoting the interquartile range between the first and third quartiles. The line within the boxes denotes the median. Paired Wilcoxon rank sum test, *p < 0.01, +p < 0.05.

(C) Spearman's correlation analysis of the association of metabolites with the main clinical parameters at baseline. *p < 0.01, +p < 0.05. The color key represents the regression coefficients of the independent variables.

and phenylacetylglutamine were mainly negatively associated with insulin resistance, body composition, liver function, and TG, and positively associated with HDL, but carnitine C2, carnitine C8:1, TMAO, uric acid, and N-acetyltryptophan presented opposite associations with these clinical characteristics.

Association between metabolites and incident diabetes

The odds ratios (Ors) (95% confidence intervals [CIs]) per SD increment of T2DM risk for each of the AAs and MRMs are presented in Tables 1 and 2. In multivariable-adjusted logistic regression model including age, gender, BMI, smoking status,

Table 1. Association between amino acids per SD increment and risk of incident diabetes

	Multivariable-adjusted[a]		+ Diet score adjusted		+ALT, AST adjusted		+2h-PG adjusted		+HOMA-IR adjusted		+ Fully adjusted[b]	
	OR (95% CI)	p value	OR (95% CI)	p value	OR (95% CI)	p value	OR (95% CI)	p value	OR (95% CI)	p value	OR (95% CI)	p value
Ala-Asp-Glu metabolism												
Aspartic acid	1.02 (0.93–1.11)	0.7000	1.04 (0.95–1.14)	0.3663	1.02 (0.93–1.12)	0.6062	1 (0.91–1.09)	0.9465	1 (0.92–1.09)	0.9596	1.02 (0.92–1.14)	0.6607
Glutamine	1.03 (0.96–1.12)	0.3940	1.07 (0.98–1.17)	0.1379	1.05 (0.97–1.14)	0.2241	1.04 (0.96–1.12)	0.3767	1.04 (0.96–1.13)	0.3546	1.11 (1.01–1.23)	0.0376
γ-Aminobutyric acid	1.07 (0.99–1.15)	0.0808	1.09 (1.00–1.19)	0.0421	1.07 (0.99–1.16)	0.0834	1.07 (0.99–1.15)	0.0914	1.07 (0.99–1.15)	0.0927	1.10 (1.00–1.21)	0.0520
α-Aminobutyric acid	1.08 (1.00–1.17)	0.0462	1.09 (1.00–1.2)	0.0481	1.08 (1.00–1.17)	0.0537	1.08 (1–1.17)	0.0454	1.07 (0.99–1.16)	0.1014	1.09 (0.99–1.21)	0.0857
Asparagine	1.08 (1.00–1.17)	0.0569	1.13 (1.03–1.24)	0.0071*	1.09 (1.00–1.18)	0.0387	1.08 (1–1.17)	0.0428	1.11 (1.02–1.2)	0.0172*	1.21 (1.09–1.34)	0.0002**
Alanine	1.25 (1.15–1.35)	<0.0001**	1.24 (1.13–1.36)	<0.0001**	1.22 (1.12–1.33)	<0.0001**	1.24 (1.14–1.34)	<0.0001**	1.22 (1.13–1.33)	<0.0001**	1.2 (1.08–1.34)	0.0005**
Glutamic acid	1.3 (1.19–1.43)	<0.0001**	1.34 (1.2–1.49)	<0.0001**	1.3 (1.18–1.43)	<0.0001**	1.28 (1.17–1.4)	<0.0001**	1.3 (1.18–1.43)	<0.0001**	1.33 (1.17–1.5)	<0.0001**
Aromatic amino acid												
Tryptophan	1.09 (1–1.19)	0.0536	1.1 (0.99–1.21)	0.0722	1.07 (0.98–1.17)	0.1560	1.09 (1–1.19)	0.0536	1.05 (0.97–1.15)	0.2458	1.07 (0.96–1.2)	0.1928
Phenylalanine	1.21 (1.11–1.31)	<0.0001**	1.25 (1.14–1.37)	<0.0001**	1.19 (1.09–1.3)	0.0001**	1.2 (1.11–1.31)	<0.0001**	1.18 (1.08–1.29)	0.0002**	1.23 (1.1–1.36)	0.0002**
Tyrosine	1.23 (1.14–1.33)	<0.0001**	1.22 (1.12–1.34)	<0.0001**	1.22 (1.12–1.33)	<0.0001**	1.22 (1.13–1.32)	<0.0001**	1.22 (1.12–1.33)	<0.0001**	1.24 (1.12–1.37)	<0.0001**
Branch chain amino acid												
Valine	1.22 (1.12–1.32)	<0.0001**	1.2 (1.09–1.31)	0.0002**	1.20 (1.1–1.31)	<0.0001**	1.22 (1.12–1.33)	<0.0001**	1.19 (1.09–1.29)	0.0001**	1.18 (1.06–1.31)	0.0028**
Isoleucine	1.22 (1.12–1.33)	<0.0001**	1.2 (1.08–1.32)	0.0004**	1.20 (1.09–1.32)	0.0001**	1.22 (1.12–1.33)	<0.0001**	1.19 (1.09–1.31)	0.0001**	1.19 (1.07–1.33)	0.0020**
Leucine	1.26 (1.16–1.37)	<0.0001**	1.27 (1.15–1.4)	<0.0001**	1.24 (1.13–1.36)	<0.0001**	1.26 (1.16–1.37)	<0.0001**	1.22 (1.12–1.34)	<0.0001**	1.24 (1.11–1.39)	0.0001**
Gly-Ser-Thr metabolism												
Glycine	0.97 (0.9–1.05)	0.4611	1.03 (0.94–1.12)	0.5862	0.97 (0.9–1.06)	0.5222	0.98 (0.91–1.06)	0.6432	0.99 (0.91–1.07)	0.7925	1.07 (0.97–1.19)	0.1859
Serine	0.99 (0.91–1.07)	0.7654	1.03 (0.94–1.13)	0.5271	1.00 (0.91–1.09)	0.9324	1 (0.92–1.08)	0.9339	1.01 (0.93–1.1)	0.8184	1.11 (0.99–1.23)	0.0647
Threonine	1.04 (0.96–1.12)	0.3547	1.05 (0.96–1.15)	0.3129	1.04 (0.95–1.13)	0.3884	1.05 (0.97–1.14)	0.2220	1.03 (0.95–1.12)	0.4090	1.08 (0.98–1.2)	0.1305
Homoserine	1.07 (0.99–1.15)	0.1050	1.1 (1.01–1.2)	0.0302	1.08 (0.99–1.17)	0.0824	1.07 (0.99–1.16)	0.0859	1.08 (0.99–1.17)	0.0722	1.16 (1.05–1.28)	0.0032**
His and lys metabolism												
Lysine	1.09 (1.01–1.19)	0.0309	1.11 (1.01–1.21)	0.0321	1.08 (0.99–1.18)	0.0657	1.09 (1.01–1.19)	0.0346	1.08 (0.99–1.18)	0.0687	1.09 (0.98–1.21)	0.1095
Histidine	1.14 (1.06–1.23)	0.0010**	1.18 (1.08–1.29)	0.0003**	1.15 (1.05–1.25)	0.0013**	1.14 (1.06–1.24)	0.0009**	1.15 (1.06–1.25)	0.0007**	1.23 (1.11–1.36)	0.0001**
2-Aminoadipic acid	1.24 (1.14–1.35)	<0.0001**	1.26 (1.15–1.39)	<0.0001**	1.23 (1.13–1.35)	<0.0001**	1.23 (1.13–1.34)	<0.0001**	1.21 (1.11–1.32)	<0.0001**	1.24 (1.1–1.38)	0.0002**
Cys and met metabolism												
Cysteine	1.04 (0.96–1.13)	0.3237	1.05 (0.96–1.15)	0.2624	1.04 (0.96–1.13)	0.3305	1.05 (0.97–1.14)	0.2127	1.04 (0.96–1.13)	0.3772	1.09 (0.98–1.21)	0.1034
Cystine	1.06 (0.98–1.15)	0.1232	1.05 (0.97–1.15)	0.2454	1.04 (0.95–1.13)	0.4012	1.06 (0.98–1.15)	0.1364	1.07 (0.99–1.17)	0.0830	1.02 (0.93–1.13)	0.6489
Methionine	1.13 (1.04–1.22)	0.0035**	1.15 (1.05–1.26)	0.0030**	1.1 (1.01–1.2)	0.0218	1.13 (1.04–1.22)	0.0038**	1.11 (1.02–1.21)	0.0126*	1.14 (1.03–1.27)	0.0106*
Arg and pro metabolism												
Arginine	0.96 (0.88–1.04)	0.2709	1.00 (0.91–1.09)	0.9374	0.95 (0.88–1.04)	0.2591	0.96 (0.89–1.04)	0.3213	0.93 (0.86–1.01)	0.0945	0.99 (0.89–1.10)	0.8373
4-Hydroxyproline	0.98 (0.91–1.06)	0.6621	1.00 (0.92–1.09)	0.9902	0.96 (0.88–1.04)	0.3209	0.99 (0.92–1.08)	0.8929	0.96 (0.89–1.04)	0.3537	0.99 (0.90–1.10)	0.8644
Citrulline	1.02 (0.94–1.1)	0.6090	1.05 (0.96–1.15)	0.2656	1.03 (0.95–1.12)	0.5053	1.04 (0.96–1.13)	0.2903	1.02 (0.94–1.11)	0.5769	1.11 (1.00–1.23)	0.0505
Ornithine	1.03 (0.95–1.11)	0.4551	1.03 (0.94–1.14)	0.5220	1.02 (0.94–1.11)	0.5794	1.03 (0.95–1.12)	0.4171	1.04 (0.96–1.13)	0.3212	1.04 (0.94–1.16)	0.4590
Proline	1.13 (1.05–1.22)	0.0015**	1.13 (1.04–1.23)	0.0054*	1.12 (1.04–1.22)	0.0050*	1.14 (1.05–1.23)	0.0012**	1.12 (1.03–1.21)	0.0061*	1.16 (1.05–1.28)	0.0037**

[a]Adjusted for age, gender, BMI, smoking status, alcohol intake, physical activity, education status, diabetes family history, systolic blood pressure, fasting blood glucose, triglycerides, low density lipoprotein cholesterol, and high density lipoprotein cholesterol.

[b]Adjustment of multivariable-adjusted model plus diet, alanine aminotransferase (ALT), aspartate aminotransferase (AST), 2-h postload plasma glucose (2h-PG), and Homeostatic Model Assessment for Insulin Resistance (HOMA-IR). *False discovery rate (FDR) <0.05, **FDR <0.01.

CellPress
OPEN ACCESS

Table 2. Association between gut microbiota metabolites per SD increment and risk of incident diabetes

	Multivariable-adjusted[a]		+ Diet Score adjusted		+ALT, AST adjusted		+2h-PG adjusted		+HOMA-IR adjusted		+ Fully adjusted[b]	
	OR (95% CI)	p value	OR (95% CI)	p value	OR (95% CI)	p value	OR (95% CI)	p value	OR (95% CI)	p value	OR (95% CI)	p value
Carnitine metabolism												
Carnitine C0	1.1 (1.02–1.19)	0.0121*	1.16 (1.06–1.26)	0.0016**	1.09 (1.01–1.18)	0.0308	1.11 (1.02–1.19)	0.0100*	1.04 (0.96–1.12)	0.3836	1.08 (0.98–1.2)	0.1352
Carnitine C2	1.13 (1.04–1.23)	0.0028*	1.16 (1.05–1.27)	0.0022**	1.13 (1.04–1.24)	0.0056*	1.14 (1.05–1.24)	0.0024**	1.12 (1.03–1.22)	0.0076*	1.16 (1.04–1.29)	0.0061*
Carnitine C3	1.17 (1.07–1.27)	0.0002**	1.16 (1.05–1.27)	0.0022**	1.15 (1.06–1.26)	0.0012**	1.17 (1.08–1.27)	0.0002**	1.13 (1.04–1.23)	0.0049*	1.13 (1.01–1.25)	0.0286
Carnitine C4	1.07 (0.99–1.15)	0.0787	1.1 (1.01–1.2)	0.0281	1.06 (0.98–1.14)	0.1576	1.07 (0.99–1.15)	0.0802	1.03 (0.96–1.11)	0.4053	1.07 (0.97–1.17)	0.1988
Carnitine C5	1.14 (1.05–1.23)	0.0013**	1.17 (1.07–1.29)	0.0008**	1.12 (1.03–1.22)	0.0063*	1.13 (1.05–1.23)	0.0023**	1.12 (1.03–1.21)	0.0078*	1.14 (1.02–1.26)	0.0191
Carnitine C6	1.14 (1.06–1.23)	0.0003**	1.21 (1.11–1.32)	<0.0001**	1.14 (1.06–1.24)	0.0009**	1.13 (1.04–1.21)	0.0020**	1.12 (1.04–1.21)	0.0026*	1.17 (1.05–1.29)	0.0031*
Carnitine C8	1.11 (1.03–1.2)	0.0054*	1.18 (1.08–1.29)	0.0002**	1.12 (1.03–1.21)	0.0064*	1.09 (1.01–1.18)	0.0200*	1.1 (1.02–1.19)	0.0165*	1.16 (1.05–1.28)	0.0040*
Carnitine C8:1	0.99 (0.9–1.08)	0.7604	1.06 (0.96–1.19)	0.2515	0.99 (0.9–1.09)	0.8103	0.98 (0.9–1.07)	0.6662	1.01 (0.92–1.12)	0.7959	1.1 (0.97–1.25)	0.1409
Choline	1.12 (1.02–1.23)	0.0151*	1.12 (1.01–1.24)	0.0361	1.12 (1.01–1.23)	0.0248	1.12 (1.02–1.23)	0.0144*	1.13 (1.02–1.24)	0.0152*	1.13 (1.01–1.27)	0.0409
TMAO	1.05 (0.97–1.14)	0.1987	1.07 (0.98–1.17)	0.1172	1.04 (0.96–1.13)	0.3118	1.06 (0.98–1.15)	0.1368	1.03 (0.95–1.12)	0.4135	1.08 (0.98–1.19)	0.1346
Gut amino acids metabolites												
Indole-3-propionic acid	0.9 (0.84–0.97)	0.0065*	0.93 (0.85–1.01)	0.0849	0.89 (0.83–0.97)	0.0044*	0.91 (0.84–0.98)	0.0105*	0.93 (0.86–1)	0.0439	0.92 (0.84–1.02)	0.1100
p-Cresol sulfate	0.92 (0.86–1)	0.0386	0.93 (0.85–1.01)	0.0950	0.95 (0.87–1.02)	0.1649	0.93 (0.86–1.00)	0.0601	0.91 (0.84–0.99)	0.0198*	0.97 (0.88–1.07)	0.5291
Hippuric acid	0.93 (0.86–1)	0.0515	0.94 (0.86–1.02)	0.1603	0.93 (0.86–1)	0.0574	0.93 (0.86–1.00)	0.0433	0.93 (0.86–1.01)	0.0868	0.95 (0.87–1.05)	0.3423
Serotonin	0.94 (0.87–1.01)	0.0803	0.94 (0.86–1.02)	0.1289	0.93 (0.86–1.01)	0.0929	0.94 (0.87–1.01)	0.0797	0.94 (0.87–1.02)	0.1528	0.92 (0.83–1.03)	0.1383
Cinnamoylglycine	0.94 (0.87–1.02)	0.1332	0.97 (0.89–1.05)	0.4500	0.95 (0.87–1.03)	0.1914	0.94 (0.87–1.01)	0.1097	0.96 (0.88–1.04)	0.2852	0.98 (0.89–1.09)	0.7333
Phenylacetylglutamine	0.97 (0.9–1.05)	0.4382	0.98 (0.9–1.07)	0.6482	0.98 (0.9–1.06)	0.5596	0.98 (0.91–1.06)	0.5739	0.96 (0.89–1.04)	0.2936	1.02 (0.92–1.13)	0.7287
Indoxyl-3-sulfate	0.98 (0.91–1.05)	0.5022	1.01 (0.93–1.1)	0.8278	0.98 (0.91–1.06)	0.5963	0.98 (0.91–1.06)	0.6644	0.95 (0.88–1.02)	0.1703	1.03 (0.93–1.14)	0.5776
Indole-3-acetic acid	1.02 (0.95–1.09)	0.6396	1.01 (0.93–1.1)	0.8220	1.01 (0.93–1.09)	0.8421	1.02 (0.95–1.10)	0.5500	1.01 (0.94–1.09)	0.7216	1.01 (0.92–1.11)	0.8880
Phenyl sulfate	1.05 (0.98–1.13)	0.1906	1.08 (0.99–1.17)	0.0918	1.05 (0.97–1.14)	0.2086	1.06 (0.98–1.14)	0.1361	1.04 (0.96–1.13)	0.2920	1.11 (1.01–1.23)	0.0298
Kynurenine	1.09 (1.01–1.18)	0.0274	1.10 (1.01–1.2)	0.0366	1.08 (0.99–1.17)	0.0798	1.10 (1.01–1.19)	0.0223*	1.08 (1–1.18)	0.0486	1.11 (1.00–1.23)	0.0476
N-acetyltryptophan	1.19 (1.1–1.28)	<0.0001**	1.21 (1.1–1.33)	<0.0001**	1.17 (1.08–1.28)	0.0002**	1.19 (1.09–1.28)	<0.0001**	1.16 (1.07–1.26)	0.0003**	1.20 (1.08–1.33)	0.0005**
Uric acid	1.23 (1.13–1.34)	<0.0001**	1.3 (1.17–1.44)	<0.0001**	1.22 (1.11–1.33)	<0.0001**	1.23 (1.13–1.34)	<0.0001**	1.17 (1.07–1.28)	0.0007**	1.24 (1.1–1.39)	0.0004**

[a]Adjusted for age, gender, BMI, smoking status, alcohol intake, physical activity, education status, diabetes family history, systolic blood pressure, fasting blood glucose, triglycerides, low density lipoprotein cholesterol, high density lipoprotein cholesterol.

[b]Adjustment of multivariable-adjusted model plus diet, alanine aminotransferase (ALT), aspartate aminotransferase (AST), 2-h postload plasma glucose (2h-PG), and Homeostatic Model Assessment for Insulin Resistance (HOMA-IR). *False discovery rate (FDR) <0.05, **FDR <0.01.

drinking status, physical activity, education, family history of diabetes, SBP, FPG, HDL cholesterol, LDL cholesterol, TG, 11 AAs were significantly associated with incident T2DM ($p < 0.05$, and false discovery rate [FDR] <0.05). Briefly, BCAAs (leucine, isoleucine and valine), AAAs (phenylalanine and tyrosine), alanine, glutamic acid, histidine, 2-aminoadipic acid, methionine, and proline were positively associated with T2DM, with ORs ranging between 1.30 (glutamic acid; 95% CI 1.19–1.43) and 1.13 (proline, 95% CI 1.05–1.22, methionine 95% CI 1.04–1.22). Sensitivity analyses with adjustment for additional baseline risk factors, including diet score, liver enzyme levels, 2-h postload plasma glucose (2h-PG), and HOMA-IR, showed similar results. Interestingly, additional adjustment of diet score, per SD increment of asparagine was associated with 13% increased risk of incident diabetes (OR: 1.13, 95% CI: 1.03–1.24, $p = 0.0071$, FDR <0.05). In the fully adjusted model including all the confounding factors, including diet score, liver enzymes, 2h-PG, and HOMA-IR, per SD increment of homoserine was associated with increased risk of diabetes (OR: 1.16, 95% CI: 1.05–1.28, $p = 0.0032$, FDR <0.01). The ORs of other AAs did not show significant change, and a total of 13 AAs were positively associated with increased risk of diabetes ($p < 0.05$, FDR <0.05) (Table 1).

In the panel of MRMs, a total of 10 MRMs were positively associated with diabetes risk, including six carnitines, choline, indole-3-propionic acid, N-acetyltryptophan in tryptophan metabolism, and uric acid in multivariable-adjusted model ($p < 0.05$, and FDR <0.05) (Table 2). ORs ranged between 1.23 (uric acid, 95% CI 1.13–1.34) and 1.10 (carnitine C0, 95% CI 1.02–1.19). We did not observe significant association between TMAO and incident T2DM (OR: 1.05, 95% CI: 0.97–1.14). In the fully adjusted model including all the confounding factors, including diet score, liver enzymes, 2h-PG, and HOMA-IR, the association between carnitine C0, carnitine C3, carnitine C5, choline, and indole-3-propionic acid with risk of incident diabetes changed significantly. Five MRMs remained positively associated with incident T2DM, including serum carnitines C2, C6, C8, N-acetyltryptophan, and uric acid. ORs ranged between 1.24 (uric acid, 95% CI 1.10–1.39) and 1.16 (carnitine C2, 95% CI 1.04–1.29, carnitine C8, 95% CI 1.05–1.28).

In sensitivity analysis of random sampling, the OR trends of all 18 metabolites associated with incident T2DM were 100% verified in the 200 times of tests of randomly selecting 80% of all samples as a subset for each time. Among them, the ORs of 15 metabolites were further verified to be significant (with $p < 0.05$) in more than 90% of tests (Table S3). Results of 2-fold internal validation of the associations between the identified metabolites in the fully adjusted model and risk of T2DM are shown in Tables S4–S5.

Based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (www.genome.jp/kegg), we generated a plot of metabolic pathways in which the identified metabolites are involved. The ORs (95% CI) per SD increment of metabolites and T2DM risk for each of the 13 AAs and 5 MRMs that were significantly associated with T2DM from the fully adjusted models are presented (Figure S1).

The risk for developing T2DM across AA and MRM quartiles are calculated in fully adjusted logistic regression models, and the risk estimates are similar. In Table S6, 17 of 28 AAs were positively associated with risk of T2DM (p for trend <0.05, and FDR <0.05). Higher levels of serum BCAAs and AAAs were consistently associated with a higher risk of T2DM. The $OR_{Q4\ vs\ Q1}$ (the highest quartile Q4 versus the reference Q1) (95% CI) ranged between 2.35 (1.70–3.23) for phenylalanine, and 1.50 (1.12–2.20) for tryptophan. Table S7 shows the associations between MRM quartiles and the risk of T2DM. Only five of 22 MRMs remained positively associated with risk of T2DM, with the $OR_{Q4\ vs\ Q1}$ (95% CI) ranging from 1.77 (1.29–2.40) for uric acid and 1.39 (1.06–1.82) for phenyl sulfate ($p < 0.05$, and FDR <0.05).

Causal mediation analysis between metabolites and clinical risk factors contributed T2DM

We performed causal mediation analysis on the linkages between clinical risk factors (including obesity, lipid metabolism, blood pressure, and liver enzyme), the 18 metabolites that significantly associated with incident T2DM in the fully adjusted model, and T2DM (Tables S8 and S9). In total, 34 significant causal mediation linkages ($p_{mediation}$ <0.05 and Mediated proportion >10%) were observed, and mainly focused on mediation effect through obesity and lipids, with 88.2% (30 of 34) significant causal relationships. In particular, TG mediated 12% and WHR mediated 14% of the effects of 2-aminoadipic acid on T2DM (both $p_{mediation}$ <0.001) (Figure 3A). For associations between BCAAs and incident T2DM, TG mediated 12% to 18% ($p_{mediation}$ <0.001) and WHR mediated 12% to 14% ($p_{mediation} = 0.002$) of the effect (Figure 3B). For the MRM-T2DM associations, dyslipidemia, especially TG (Mediated proportion = 13%, $p_{mediation}$ <0.001) mediated the most effect of carnitine C2 on incident T2DM, while WHR (Mediated proportion = 12%–13%, $p_{mediation}$ <0.002) mediated the association of carnitine C6 and carnitine C8 with T2DM (Figure 3C). In addition, N-acetyltryptophan-T2DM association was mediated by TG and WHR with proportion of 12% ($p_{mediation} = 0.005$) (Figure 3D).

Variance in metabolite levels explained by each clinical risk factor

Next, we further assessed the variance in metabolite levels explained by each clinical risk factor and measurement at baseline (Figure S2, Supporting Information). Among them, prominent clinical risk – related metabolites with more than 2% of its variance explained by single clinical risk factor are highlighted in Figure 4. We observed associations of HOMA-IR with 2-aminoadipic acid; body composition with glutamine, phenylalanine, isoleucine, 2-aminoadipic acid, methionine, citrulline, proline, carnitine C5, cinnamoylglycine, and uric acid; liver function with methionine, tryptophan, and BCAAs (valine, leucine, and isoleucine); and smoking behavior with 2-aminoadipic acid. The strongest associations were detected between carnitine C5 with WHR. No metabolite was found associated with blood pressure. It is remarkable that 17.1% of 2-aminoadipic acid variance was explained by combined clinical factors including HOMA-IR (2.16%), body composition (5.1%), liver function (3.3%), smoking behavior (2.3%), and so on, and the strongest associations were detected between 2-aminoadipic acid with insulin resistance. More than 10% of BCAA variance (valine, leucine, and isoleucine) was explained by combined clinical

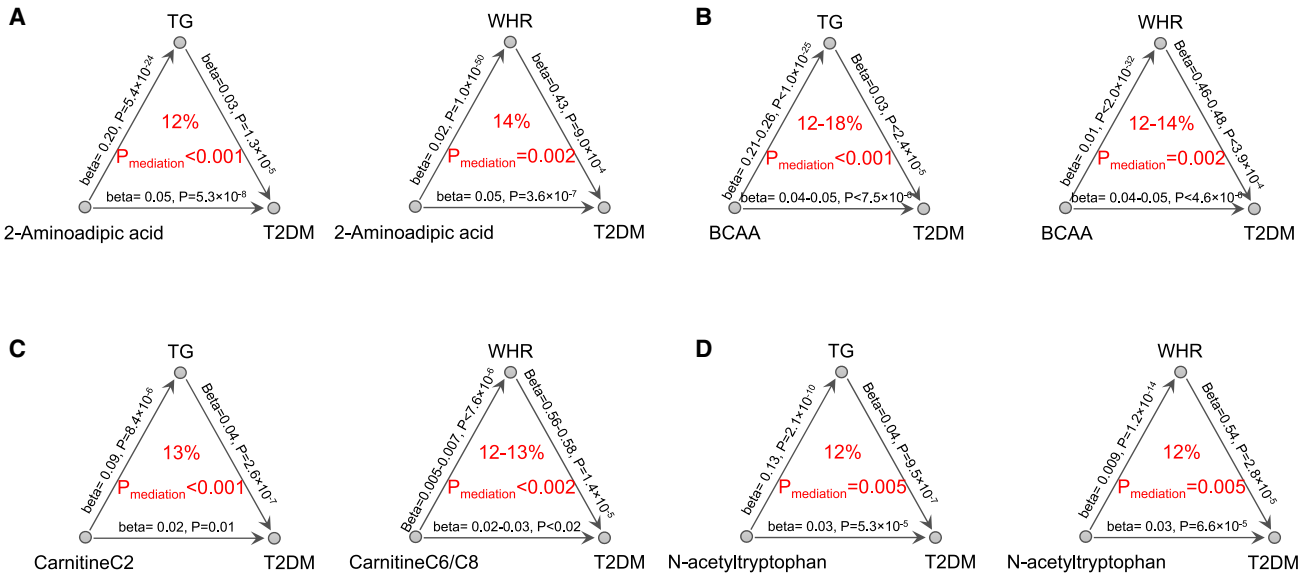


Figure 3. Causal mediation analysis among metabolites, clinical risk factors, and T2DM

(A) Mediation linkages between TG\WHR and 2-aminoadipic acid contributed T2DM.

(B) Mediation linkages between TG\WHR and BCAAs contributed T2DM.

(C) Mediation linkages between TG\WHR and carnitines contributed T2DM.

(D) Mediation linkages between TG\WHR and N-acetyltryptophan contributed T2DM. Data used for causal mediation analysis were measured from all 3,414 subjects.

factors including body composition, lipid metabolism, and liver function. On the other hand, some of identified antecedents of T2DM, such as serum levels of carnitine C2, carnitine C6, and carnitine C8, were rarely explained by traditional clinical measures (<2%).

DISCUSSION AND CONCLUSION

The current study is the largest and most comprehensive investigation on the associations between serum AAs and MRMs with risk of developing diabetes in a prospective cohort of normoglycemic Chinese individuals. We identified several metabolites and their mediating pathways that are significantly associated with risk of T2DM.

Numerous discovery-based metabolomics studies have suggested that certain AAs may be both markers and effectors of incident diabetes (Table S10). Consistent with previous findings,[8,11] the current study confirmed a predictive value of BCAAs (valine, leucine, and isoleucine) and AAAs (tyrosine and phenylalanine) for risk of diabetes in a Chinese population. We also confirmed that alanine, glutamic acid, homoserine, 2-aminoadipic acid, histidine, and proline could be good candidates as early biomarkers for incident T2DM.[26–30] Moreover, we observed that serum asparagine might be a potential biomarker for the development of T2DM among individuals with normoglycemia. In the Framingham Heart Study, per SD change of plasma asparagine was reported to be inversely associated with insulin and HOMA, whereas the correlation between glucose and asparagine was not significant ($\beta = 0.2$, $p > 0.05$).[31] Asparagine was reported as a protective biomarker of diabetes risk in a subset of 2,939 Atherosclerosis Risk in Communities (ARIC) study participants with metabolomics data and without prevalent diabetes,[32] whereas null relationship was reported in other studies.[33,34] Among 2,519 individuals with coronary artery disease but without diabetes, with a median follow-up of 10.3 years, asparagine was identified in an optimal model for predicting new-onset type 2 diabetes; however, the association was rendered statistically nonsignificant after adjustment for glucose and multiple comparisons.[33] In our study, serum asparagine was shown to be associated with increased risk of diabetes in the multivariable-adjusted model plus the adjustment of diet score, and in the fully adjusted model. The discrepancies between our study and the previous findings might be partially explained by the PSM matched prospective nested case-control study design, inclusion of NGR individuals only, and sensitivity analyses with adjustment for additional baseline confounding factors. Interestingly, mice hepatic knock-down of the asparagine synthetase gene resulted in a significant decrease in plasma glucose concentration, suggesting that asparagine synthetase plays a direct role in glucose metabolism. This enzyme converts aspartate and glutamine to asparagine and glutamate in an ATP-dependent reaction and can affect glucose degradation.[35]

Circulating tryptophan-related metabolites have been reported to be associated with obesity, insulin resistance, and diabetes, whereas the results were inconsistent (Table S10).[36–39] A population-based cohort study included 5,181 Finnish men from the cross-sectional Metabolic Syndrome in Men (METSIM) study indicated kynurenine and N-acetyltryptophan were related to increased risk of incident T2DM.[15] Recently, data from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), suggested that tryptophan, four kynurenine-pathway metabolites (kynurenine, kynurenate, xanthurenate, and quinolinate),

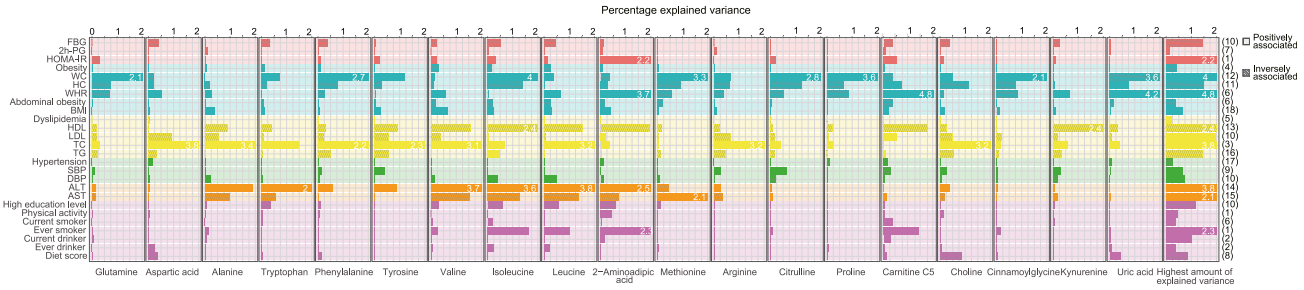


Figure 4. Variance explained in selected metabolites associated with T2DM

Different colors are used to distinguish different categories of risk factors. Red: glucose metabolism; blue: obesity; yellow: lipid metabolism; green: blood pressure; orange: liver enzymes; and purple: lifestyle factors. Solid colors indicate positive associations with metabolite levels, whereas shading indicates inverse associations. The column on the far right indicates the maximum amount of variance for any metabolite by each risk factor: (1) 2-aminoadipic acid; (2) α-aminobutyric acid; (3) aspartic acid; (4) carnitine C3; (5) carnitine C4; (6) carnitine C5; (7) carnitine C6; (8) choline; (9) citrulline; (10) glutamic acid; (11) glycine; (12) isoleucine; (13) kynurenine; (14) leucine; (15) methionine; (16) serine; (17) serotonin; (18) valine. Data used for variance analysis were measured from all 3,414 subjects.

and indolelactate were positively associated with T2DM risk, while indolepropionate was inversely associated with T2DM risk,[20] but null association has been reported in other studies.[37,39] The difference between the above results might be partially explained by the study design and the adjustment of confounding factors.

Likewise, previous findings on the association of TMAO and its nutrient precursors (choline and carnitine) with T2DM is also inconsistent.[40–44] The POUNDS Lost trial found TMAO and its precursors were related to lesser improvements in glycemia and insulin sensitivity.[41] TMAO has been related to increased risk of T2DM in a cross-sectional study of the Chinese population[42] and in a meta-analysis of 12 clinical studies[40]; however, null associations have been reported in one of two cohorts in another Chinese study[43] and a Norwegian study.[44] The current study has also shown a null association of TMAO and T2DM. Nevertheless, we detected significant associations of other choline and carnitine metabolites, including carnitine C0, C2, C3, C5, C6, and C8, with the risk of T2DM in multivariable-adjusted model. In fully adjusted model with additional adjustment of diet score, liver enzyme, 2h-PG, and HOMA-IR, higher carnitine C2, C3, C5, C6, and C8 was associated with incident diabetes. These discrepancies might be partially explained by the differences in the sample size, study populations, and study design (including only NGR at baseline in the current study versus both NGR and diabetes in the meta-analysis study) and the adjustment of covariates. For other MRMs, we confirmed the previous observations that serum uric acid was positively associated with risk of diabetes.[15]

Results from causal mediation analysis supplemented knowledge on the potential biological mechanism of metabolites influencing the onset of T2DM. Previous studies revealed that 2-aminoadipic acid, an oxidation intermediate of lysine degradation, was correlated with adipogenic differentiation. Circulating 2-aminoadipic acid was positively associated with fat mass, fat percent, WC, and TG, suggesting a contribution status to the early prevention of obesity-related metabolic disorders.[45] Our results provided evidence supporting the hypothesis that metabolites causally impact the onset of diabetes, and further emphasized that clinical phenotypes including dyslipidemia (especially TG metabolism disorder) and WHR might be important in regulating AA and MRM contributions to T2DM. Further *in vivo* and *in vitro* experiments are needed to provide concrete, clear, and definite associations.

Moreover, we provided a comprehensive catalog of risk factor–metabolite-diabetes associations in a population with NGR, which helped us contextualize our findings and provide future directions in metabolomics studies. Through integration of the metabolome and phenome in the current study, we identified early metabolic changes related to the development of T2DM. Our systematic comparison allowed us to untangle associations between closely related molecules. In most cases, variance explained in the serum metabolites associated with diabetes by risk factors were modestly limited, which emphasized the irreplaceable role of metabolites in predicting the risk of diabetes even in individuals with NGR. On the other hand, several clinical risk-related metabolite discoveries have been made in our study, such as carnitine C5 and WHR, methionine, and liver enzymes. It was reported that 2-aminoadipic acid was associated with insulin resistance.[45,46] Our results further revealed that in addition to HOMA-IR and body composition, liver enzymes and smoking behavior also influenced 2-aminoadipic acid levels, indicating the important role of body weight management and lifestyle in 2-aminoadipic homeostasis. In the current study, serum short-chain carnitine levels are rarely explained by behavioral and metabolic risk factors, including the diet score. Previous studies reported that elevated acylcarnitine levels observed in patients with diabetes or insulin resistance are closely related to habitual diet, especially dietary protein intakes.[47] Using a targeted metabolomics approach, Floegel et al. demonstrated that 3.5% of acylcarnitine variation was explained by habitual diet in the EPIC-Potsdam cohort.[3,48] Healthy dietary score[49] collected in the current study is composed of vegetable, fish, soy product, and sugary drink intake, but not meat intake. That might be a reason why variance of carnitines could rarely be explained.

The mechanisms underlying the relationship between metabolites with glucose metabolism and insulin resistance have not been fully established. It was proposed that BCAAs could activate the mammalian target of rapamycin complex 1 (mTORC1) signaling, impair BCAA metabolism, and lead to the accumulation of toxic metabolites, which bring about mitochondrial

dysfunction in pancreatic islet β cells and the occurrence of insulin resistance and T2DM.[50,51] Tryptophan metabolism follows three major pathways, including the kynurenine pathway via indoleamine 2,3-dioxygenase 1 (IDO1), the serotonin production pathway via Trp hydroxylase 1 (TpH1), and the direct transformation of Trp into several molecules, including ligands of the aryl hydrocarbon receptor (AhR) by the gut microbiota.[36] The kynurenine pathway is the primary path of tryptophan catabolism in most mammalian cells, and kynurenines inhibit proinsulin synthesis from pancreatic islets or generate complexes with insulin decreasing its biological activity and promote the development of insulin resistance.[52] In our current study, in the fully adjusted model with all confounding factors, serum N-acetyltryptophan, but not tryptophan or kynurenine, was associated with increased risk of diabetes. N-acetyltryptophan was observed in human urine as an MRM in 2017.[53] Previous study identified it as a possible biomarker for colonization resistance[54]; however, little is known about its role in regulating the pathologic process of T2DM. Further studies are warranted to strengthen possible molecular mechanisms. Although our findings among a normoglycemic population may have strong biological plausibility, the current study is unable to determine whether the associations reflect causal relationships or are rather driven by early processes of disease development due to its observational nature. Mendelian randomization study on AAs and MRMs will be helpful to make causal inference.

In conclusion, our study included a large sample size and prospective study design, and hence, the ability to study the association with T2DM, the inclusion of populations from 20 communities from various geographic regions in China, the population being normoglycemic at baseline, and adjustment for many potential confounders. It is worth mentioning that the measurement of the glucose regulation status was based on the OGTT at both baseline and the follow-up visit, which make it possible for accurate evaluation of glucose dysregulation. In terms of the metabolomics analytical approaches, our work extensively covers 28 key AA classes that are essential in the overall homeostatic balance of AA metabolism and 22 MRMs.

Our foregoing data underscore the potential importance of AAs and MRMs early in the pathogenesis of diabetes and demonstrate the mediating pathways through the characterization of triangles of clinical risk factor–metabolite-diabetes link. Further studies are warranted to test whether serum AA and MRM measurements can be potential biomarkers used in patient screening or treatment monitoring and elucidate the biological mechanisms of metabolites in the onset and progression of T2DM. Our study may also provide insights for the diagnosis as well as therapeutic targets for T2DM.

LIMITATIONS OF THE STUDY

Our study, however, also has several limitations. First, the relatively short follow-up duration might limit the predictive potential of our identified metabolites panel to a narrow time window. Further studies with longer follow-up duration might be additionally informative. Second, information regarding the diet and medication affecting microbiota was not available and gut flora was not assessed in the current study. Third, metabolites were measured only at baseline fasting samples, and it would also be interesting to see whether metabolite composition continued to change with incident disease. Fourth, MRMs including short-chain fatty acids (SCFAs) and secondary bile acids were not involved in the present study because of data acquisition method limitation. Associations between bile acids and incident T2DM were characterized previously.[21] Further studies on circulating concentrations of SCFAs and T2DM would be worthy additions. Finally, the PSM-matched case-control design hampers the use of matching factors (age, gender, BMI, FPG) for risk prediction in the case-control samples. Although the current study was conducted in normoglycemic individuals at baseline, substantial metabolic differences were observed between case subjects and control subjects.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Details of blood sample and data collection
- METHOD DETAILS
  - Definition of diabetes
  - Amino acids and microbiota-related metabolites measurement
  - Amino acids
  - Microbiota-related metabolites
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2022.100727.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (grant nos. 91857205, 82088102, 81930021, 81970728, 81970691, 82170819, and 21904084), Shanghai Outstanding Academic Leaders Plan (grant nos. 20XD1422800), Shanghai Medical and Health Development Foundation (grant nos. DMRFP_I_01), Clinical Research Plan of SHDC (grant nos. SHDC2020CR3064B, SHDC2020CR1001A), Science and Technology Committee of Shanghai (grant nos. 20Y11905100 and 19411964200), and Clinical Research Project of Shanghai Municipal Health Commission (20214Y0002).

AUTHOR CONTRIBUTIONS

J.L., Y.B., G.N., and W.W. conceived and designed the study. G.N. supervised the study. J.L., M.L., W.S., and C.H. did the statistical analysis. All authors contributed to acquisition, analysis, or interpretation of data. J.L., Y.B., and M.L. drafted the manuscript. All authors revised the report and approved the final version before submission. J.L. and G.N. had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 21, 2022
Revised: June 16, 2022
Accepted: July 22, 2022
Published: August 22, 2022

REFERENCES

1. Global, regional, and National Life Expectancy, All-Cause Mortality, and Cause-specific Mortality for 249 Causes of Death, 1980-2015 (2016). A Systematic Analysis for the Global Burden of Disease Study 2015. Lancet *388*, 1459–1544. https://doi.org/10.1016/S0140-6736(16)31012-1.
2. International Diabetes Federation (2019). IDF Diabetes Atlas, 9th edition. https://diabetesatlas.org/upload/resources/material/20200302_133351_IDFATLAS9e-final-web.pdf.
3. Lu, J., Li, M., Xu, Y., Bi, Y., Qin, Y., Li, Q., Wang, T., Hu, R., Shi, L., Su, Q., et al. (2020). Early life famine exposure, ideal cardiovascular health metrics, and risk of incident diabetes: findings from the 4C study. Diabetes Care *43*, 1902–1909. https://doi.org/10.2337/dc19-2325.
4. Yu, M., Zhan, X., Yang, Z., and Huang, Y. (2021). Measuring the global, regional, and national burden of type 2 diabetes and the attributable risk factors in all 194 countries. J. of diabetes *13*, 613–639. https://doi.org/10.1111/1753-0407.13159.
5. Lu, J., He, J., Li, M., Tang, X., Hu, R., Shi, L., Su, Q., Peng, K., Xu, M., Xu, Y., et al. (2019). Predictive value of fasting glucose, postload glucose, and hemoglobin A1c on risk of diabetes and complications in Chinese adults. Diabetes Care *42*, 1539–1548. https://doi.org/10.2337/dc18-1390.
6. Chen, Z.-Z., and Gerszten, R.E. (2020). Metabolomics and proteomics in type 2 diabetes. Circ. Res. *126*, 1613–1627. https://doi.org/10.1161/CIRCRESAHA.120.315898.
7. Yang, Q., Vijayakumar, A., and Kahn, B.B. (2018). Metabolites as regulators of insulin sensitivity and metabolism. Nat. Rev. Mol. Cell Biol. *19*, 654–672. https://doi.org/10.1038/s41580-018-0044-8.
8. Wang, T.J., Larson, M.G., Vasan, R.S., Cheng, S., Rhee, E.P., McCabe, E., Lewis, G.D., Fox, C.S., Jacques, P.F., Fernandez, C., et al. (2011). Metabolite profiles and the risk of developing diabetes. Nat. Med. *17*, 448–453. https://doi.org/10.1038/nm.2307.
9. Floegel, A., Stefan, N., Yu, Z., Mühlenbruch, K., Drogan, D., Joost, H.-G., Fritsche, A., Häring, H.U., Hrabě de Angelis, M., Peters, A., et al. (2013). Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. Diabetes *62*, 639–648. https://doi.org/10.2337/db12-0495.
10. Tillin, T., Hughes, A.D., Wang, Q., Würtz, P., Ala-Korpela, M., Sattar, N., Forouhi, N.G., Godsland, I.F., Eastwood, S.V., McKeigue, P.M., and Chaturvedi, N. (2015). Diabetes risk and amino acid profiles: cross-sectional and prospective analyses of ethnicity, amino acids and diabetes in a South Asian and European cohort from the SABRE (Southall and Brent REvisited) Study. Diabetologia *58*, 968–979. https://doi.org/10.1007/s00125-015-3517-8.
11. Ruiz-Canela, M., Guasch-Ferré, M., Toledo, E., Clish, C.B., Razquin, C., Liang, L., Wang, D.D., Corella, D., Estruch, R., Hernáez, Á., et al. (2018). Plasma branched chain/aromatic amino acids, enriched Mediterranean diet and risk of type 2 diabetes: case-cohort study within the PREDIMED Trial. Diabetologia *61*, 1560–1571. https://doi.org/10.1007/s00125-018-4611-5.
12. Chen, T., Ni, Y., Ma, X., Bao, Y., Liu, J., Huang, F., Hu, C., Xie, G., Zhao, A., Jia, W., and Jia, W. (2016). Branched-chain and aromatic amino acid profiles and diabetes risk in Chinese populations. Sci. Rep. *6*, 20594. https://doi.org/10.1038/srep20594.
13. Lu, Y., Wang, Y., Ong, C.-N., Subramaniam, T., Choi, H.W., Yuan, J.-M., Koh, W.-P., and Pan, A. (2016). Metabolic signatures and risk of type 2 diabetes in a Chinese population: an untargeted metabolomics study using both LC-MS and GC-MS. Diabetologia *59*, 2349–2359. https://doi.org/10.1007/s00125-016-4069-2.
14. Qiu, G., Zheng, Y., Wang, H., Sun, J., Ma, H., Xiao, Y., Li, Y., Yuan, Y., Yang, H., Li, X., et al. (2016). Plasma metabolomics identified novel metabolites associated with risk of type 2 diabetes in two prospective cohorts of Chinese adults. Int. J. Epidemiol. *45*, 1507–1516.
15. Vangipurapu, J., Fernandes Silva, L., Kuulasmaa, T., Smith, U., and Laakso, M. (2020). Microbiota-related metabolites and the risk of type 2 diabetes. Diabetes Care *43*, 1319–1325. https://doi.org/10.2337/dc19-2533.
16. Papandreou, C., Bulló, M., Zheng, Y., Ruiz-Canela, M., Yu, E., Guasch-Ferré, M., Toledo, E., Clish, C., Corella, D., Estruch, R., et al. (2018). Plasma trimethylamine-N-oxide and related metabolites are associated with type 2 diabetes risk in the Prevención con Dieta Mediterránea (PREDIMED) trial. Am. J. Clin. Nutr. *108*, 163–173. https://doi.org/10.1093/ajcn/nqy058.
17. Shan, Z., Sun, T., Huang, H., Chen, S., Chen, L., Luo, C., Yang, W., Yang, X., Yao, P., Cheng, J., et al. (2017). Association between microbiota-dependent metabolite trimethylamine–oxide and type 2 diabetes. Am. J. Clin. Nutr. *106*, 888–894. https://doi.org/10.3945/ajcn.117.157107.
18. Strand, E., Rebnord, E.W., Flygel, M.R., Lysne, V., Svingen, G.F.T., Tell, G.S., Løland, K.H., Berge, R.K., Svardal, A., Nygård, O., and Pedersen, E.R. (2018). Serum carnitine metabolites and incident type 2 diabetes mellitus in patients with suspected stable Angina pectoris. J. Clin. Endocrinol. Metab. *103*, 1033–1041. https://doi.org/10.1210/jc.2017-02139.
19. Sun, L., Liang, L., Gao, X., Zhang, H., Yao, P., Hu, Y., Ma, Y., Wang, F., Jin, Q., Li, H., et al. (2016). Early prediction of developing type 2 diabetes by plasma acylcarnitines: a population-based study. Diabetes Care *39*, 1563–1570. https://doi.org/10.2337/dc16-0232.
20. Qi, Q., Li, J., Yu, B., Moon, J.Y., Chai, J.C., Merino, J., Hu, J., Ruiz-Canela, M., Rebholz, C., Wang, Z., et al. (2022). Host and gut microbial tryptophan metabolism and type 2 diabetes: an integrative analysis of host genetics, diet, gut microbiome and circulating metabolites in cohort studies. Gut *71*, 1095–1105. https://doi.org/10.1136/gutjnl-2021-324053.
21. Lu, J., Wang, S., Li, M., Gao, Z., Xu, Y., Zhao, X., Hu, C., Zhang, Y., Liu, R., Hu, R., et al. (2021). Association of serum bile acids profile and pathway dysregulation with the risk of developing diabetes among normoglycemic Chinese adults: findings from the 4C study. Diabetes Care *44*, 499–510. https://doi.org/10.2337/dc20-0884.
22. Lu, J., Lam, S.M., Wan, Q., Shi, L., Huo, Y., Chen, L., Tang, X., Li, B., Wu, X., Peng, K., et al. (2019). High-coverage targeted lipidomics reveals novel serum lipid predictors and lipid pathway dysregulation antecedent to type 2 diabetes onset in normoglycemic Chinese adults. Diabetes Care *42*, 2117–2126. https://doi.org/10.2337/dc19-0100.
23. Parsons, L.S. (2004). Performing a 1:N Case-Control Match on Propensity Score. https://support.sas.com/resources/papers/proceedings/proceedings/sugi29/165-29.pdf.
24. Liu, R., Hong, J., Xu, X., Feng, Q., Zhang, D., Gu, Y., Shi, J., Zhao, S., Liu, W., Wang, X., et al. (2017). Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention. Nat. Med. *23*, 859–868. https://doi.org/10.1038/nm.4358.
25. Gu, Y., Wang, X., Li, J., Zhang, Y., Zhong, H., Liu, R., Zhang, D., Feng, Q., Xie, X., Hong, J., et al. (2017). Analyses of gut microbiota and plasma bile acids enable stratification of patients for antidiabetic treatment. Nat. Commun. *8*, 1785. https://doi.org/10.1038/s41467-017-01682-2.
26. Gunther, S.H., Khoo, C.M., Tai, E.S., Sim, X., Kovalik, J.P., Ching, J., Lee, J.J., and van Dam, R.M. (2020). Serum acylcarnitines and amino acids and risk of type 2 diabetes in a multiethnic Asian population. BMJ Open Diabetes Res. Care *8*, e001315. https://doi.org/10.1136/bmjdrc-2020-001315.
27. Lai, M., Liu, Y., Ronnett, G.V., Wu, A., Cox, B.J., Dai, F.F., Röst, H.L., Gunderson, E.P., and Wheeler, M.B. (2020). Amino acid and lipid metabolism in post-gestational diabetes and progression to type 2 diabetes: a

metabolic profiling study. Plos Med. *17*, e1003112. https://doi.org/10.1371/journal.pmed.1003112.

28. Wang, T.J., Ngo, D., Psychogios, N., Dejam, A., Larson, M.G., Vasan, R.S., Ghorbani, A., O'Sullivan, J., Cheng, S., Rhee, E.P., et al. (2013). 2-Aminoadipic acid is a biomarker for diabetes risk. J. Clin. Invest. *123*, 4309–4317. https://doi.org/10.1172/JCI64801.
29. Peddinti, G., Cobb, J., Yengo, L., Froguel, P., Kravić, J., Balkau, B., Tuomi, T., Aittokallio, T., and Groop, L. (2017). Early metabolic markers identify potential targets for the prevention of type 2 diabetes. Diabetologia *60*, 1740–1750. https://doi.org/10.1007/s00125-017-4325-0.
30. Liu, X., Zheng, Y., Guasch-Ferré, M., Ruiz-Canela, M., Toledo, E., Clish, C., Liang, L., Razquin, C., Corella, D., Estruch, R., et al. (2019). High plasma glutamate and low glutamine-to-glutamate ratio are associated with type 2 diabetes: case-cohort study within the PREDIMED trial. Nutr. Metab. Cardiovasc. Dis. *29*, 1040–1049. https://doi.org/10.1016/j.numecd.2019.06.005.
31. Cheng, S., Rhee, E.P., Larson, M.G., Lewis, G.D., McCabe, E.L., Shen, D., Palma, M.J., Roberts, L.D., Dejam, A., Souza, A.L., et al. (2012). Metabolite profiling identifies pathways associated with metabolic risk in humans. Circulation *125*, 2222–2231. https://doi.org/10.1161/CIRCULATIONAHA.111.067827.
32. Rebholz, C.M., Yu, B., Zheng, Z., Chang, P., Tin, A., Köttgen, A., Wagenknecht, L.E., Coresh, J., Boerwinkle, E., and Selvin, E. (2018). Serum metabolomic profile of incident diabetes. Diabetologia *61*, 1046–1054. https://doi.org/10.1007/s00125-018-4573-7.
33. McCann, A., Giil, L.M., Ulvik, A., Seifert, R., Rebnord, E.W., Pedersen, E.R., Svingen, G.F.T., Meyer, K., Strand, E., Dankel, S., et al. (2019). Plasma amino acids and incident type 2 diabetes in patients with coronary artery disease. Diabetes Care *42*, 1225–1233. https://doi.org/10.2337/dc18-2217.
34. Palmer, N.D., Stevens, R.D., Antinozzi, P.A., Anderson, A., Bergman, R.N., Wagenknecht, L.E., Newgard, C.B., and Bowden, D.W. (2015). Metabolomic profile associated with insulin resistance and conversion to diabetes in the Insulin Resistance Atherosclerosis Study. J. Clin. Endocrinol. Metab. *100*, E463–E468. https://doi.org/10.1210/jc.2014-2357.
35. Iida, S., Kamiya, H., Nakaya, A., Hayashi, Y., Sawada, A., Kaji, N., Baba, Y., and Harashima, H. (2013). Reduced plasma glucose by asparagine synthetase knockdown in the mouse liver. Biol. Pharm. Bull. *36*, 2009–2011. https://doi.org/10.1248/bpb.b13-00414.
36. Agus, A., Planchais, J., and Sokol, H. (2018). Gut microbiota regulation of tryptophan metabolism in health and disease. Cell Host Microbe *23*, 716–724. https://doi.org/10.1016/j.chom.2018.05.003.
37. Rebnord, E.W., Strand, E., Midttun, Ø., Svingen, G.F.T., Christensen, M.H.E., Ueland, P.M., Mellgren, G., Njølstad, P.R., Tell, G.S., Nygård, O.K., and Pedersen, E.R. (2017). The kynurenine:tryptophan ratio as a predictor of incident type 2 diabetes mellitus in individuals with coronary artery disease. Diabetologia *60*, 1712–1721. https://doi.org/10.1007/s00125-017-4329-9.
38. Chen, T., Zheng, X., Ma, X., Bao, Y., Ni, Y., Hu, C., Rajani, C., Huang, F., Zhao, A., Jia, W., and Jia, W. (2016). Tryptophan predicts the risk for future type 2 diabetes. PLoS One *11*, e0162192. https://doi.org/10.1371/journal.pone.0162192.
39. Yu, E., Papandreou, C., Ruiz-Canela, M., Guasch-Ferre, M., Clish, C.B., Dennis, C., Liang, L., Corella, D., Fitó, M., Razquin, C., et al. (2018). Association of tryptophan metabolites with incident type 2 diabetes in the PREDIMED trial: a case-cohort study. Clin. Chem. *64*, 1211–1220. https://doi.org/10.1373/clinchem.2018.288720.
40. Zhuang, R., Ge, X., Han, L., Yu, P., Gong, X., Meng, Q., Zhang, Y., Fan, H., Zheng, L., Liu, Z., and Zhou, X. (2019). Gut microbe-generated metabolite trimethylamine N-oxide and the risk of diabetes: a systematic review and dose-response meta-analysis. Obes. Rev. *20*, 883–894. https://doi.org/10.1111/obr.12843.
41. Heianza, Y., Sun, D., Li, X., DiDonato, J.A., Bray, G.A., Sacks, F.M., and Qi, L. (2019). Gut microbiota metabolites, amino acid metabolites and improvements in insulin sensitivity and glucose metabolism: the POUNDS Lost trial. Gut *68*, 263–270. https://doi.org/10.1136/gutjnl-2018-316155.
42. Shan, Z., Sun, T., Huang, H., Chen, S., Chen, L., Luo, C., Yang, W., Yang, X., Yao, P., Cheng, J., et al. (2017). Association between microbiota-dependent metabolite trimethylamine-N-oxide and type 2 diabetes. Am. J. Clin. Nutr. *106*, 888–894. https://doi.org/10.3945/ajcn.117.157107.
43. Qiu, G., Zheng, Y., Wang, H., Sun, J., Ma, H., Xiao, Y., Li, Y., Yuan, Y., Yang, H., Li, X., et al. (2016). Plasma metabolomics identified novel metabolites associated with risk of type 2 diabetes in two prospective cohorts of Chinese adults. Int. J. Epidemiol. *45*, 1507–1516. https://doi.org/10.1093/ije/dyw221.
44. Svingen, G.F.T., Schartum-Hansen, H., Pedersen, E.R., Ueland, P.M., Tell, G.S., Mellgren, G., Njølstad, P.R., Seifert, R., Strand, E., Karlsson, T., and Nygård, O. (2016). Prospective associations of systemic and urinary choline metabolites with incident type 2 diabetes. Clin. Chem. *62*, 755–765. https://doi.org/10.1373/clinchem.2015.250761.
45. Lee, H.J., Jang, H.B., Kim, W.H., Park, K.J., Kim, K.Y., Park, S.I., and Lee, H.J. (2019). 2-Aminoadipic acid (2-AAA) as a potential biomarker for insulin resistance in childhood obesity. Sci. Rep. *9*, 13610. https://doi.org/10.1038/s41598-019-49578-z.
46. Xu, W.Y., Shen, Y., Zhu, H., Gao, J., Zhang, C., Tang, L., Lu, S.Y., Shen, C.L., Zhang, H.X., Li, Z., et al. (2019). 2-Aminoadipic acid protects against obesity and diabetes. J. Endocrinol. *243*, 111–123. https://doi.org/10.1530/JOE-19-0157.
47. Rousseau, M., Guénard, F., Garneau, V., Allam-Ndoul, B., Lemieux, S., Pérusse, L., and Vohl, M.C. (2019). Associations between dietary protein sources, plasma BCAA and short-chain acylcarnitine levels in adults. Nutrients *11*, E173. https://doi.org/10.3390/nu11010173.
48. Floegel, A., von Ruesten, A., Drogan, D., Schulze, M.B., Prehn, C., Adamski, J., Pischon, T., and Boeing, H. (2013). Variation of serum metabolites related to habitual diet: a targeted metabolomic approach in EPIC-Potsdam. Eur. J. Clin. Nutr. *67*, 1100–1108. https://doi.org/10.1038/ejcn.2013.147.
49. Lloyd-Jones, D.M., Hong, Y., Labarthe, D., Mozaffarian, D., Appel, L.J., Van Horn, L., Greenlund, K., Daniels, S., Nichol, G., Tomaselli, G.F., et al. (2010). Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic Impact Goal through 2020 and beyond. Circulation *121*, 586–613. https://doi.org/10.1161/CIRCULATIONAHA.109.192703.
50. Lynch, C.J., and Adams, S.H. (2014). Branched-chain amino acids in metabolic signalling and insulin resistance. Nat. Rev. Endocrinol. *10*, 723–736. https://doi.org/10.1038/nrendo.2014.171.
51. Yoon, M.S. (2016). The emerging role of branched-chain amino acids in insulin resistance and metabolism. Nutrients *8*, E405. https://doi.org/10.3390/nu8070405.
52. Song, P., Ramprasath, T., Wang, H., and Zou, M.H. (2017). Abnormal kynurenine pathway of tryptophan catabolism in cardiovascular diseases. Cell. Mol. Life Sci. *74*, 2899–2916. https://doi.org/10.1007/s00018-017-2504-2.
53. Pavlova, T., Vidova, V., Bienertova-Vasku, J., Janku, P., Almasi, M., Klanova, J., and Spacil, Z. (2017). Urinary intermediates of tryptophan as indicators of the gut microbial metabolism. Anal. Chim. Acta *987*, 72–80. https://doi.org/10.1016/j.aca.2017.08.022.
54. Obrenovich, M.E., Tima, M., Polinkovsky, A., Zhang, R., Emancipator, S.N., and Donskey, C.J. (2017). Targeted metabolomics analysis identifies intestinal microbiota-derived urinary biomarkers of colonization resistance in antibiotic-treated mice. Antimicrob. Agents Chemother. *61*, e00477-17. https://doi.org/10.1128/AAC.00477-17.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
SAS version 9.4	SAS Institute	https://www.sas.com/
R version 3.3.1	R-Project	https://www.r-project.org/
STAR Methods. Variance analysis: R package variancePartition version 1.14.1	Bioconductor	http://www.bioconductor.org/packages/release/bioc/html/variancePartition.html
STAR Methods. Mediation analysis: R package mediation version 4.5.0	R Cran	https://cran.r-project.org/
Deposited data		
Metabolomics data-amino acids	NODE (The National Omics Data Encyclopedia)	https://www.biosino.org/node/project/detail/OEP003528
Metabolomics data-microbiota-related metabolites	NODE (The National Omics Data Encyclopedia)	https://www.biosino.org/node/project/detail/OEP003528
Biological samples		
Serum samples of cases and controls	4C BioBank	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Jieli Lu (jielilu@hotmail.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Data of amino acids and microbiota-related metabolites have been deposited at https://www.biosino.org/node/project/detail/OEP003528. Additional data reported in this manuscript will be shared. To request access, contact the lead contact jielilu@hotmail.com. No new code was generated in this study.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Samples used in this study were from the China Cardiometabolic Disease and Cancer Cohort (4C) Study, a nationwide, population-based, prospective cohort study with up to 5 years follow-up.[5,21,22]

Details of blood sample and data collection

Blood specimens were processed within 2 h after blood sample collection under a stringent quality control mechanism at the field center. Sera were aliquoted into 0.5-mL Eppendorf tubes within 2 h of blood collection and shipped by air in dry ice to the central laboratory of the study located at Shanghai Institute of Endocrine and Metabolic Diseases. Samples were stored at −80°C until use. Information on sociodemographic characteristics, lifestyle factors (including smoking or drinking status, healthy dietary score and physical activity), as well as medical history was collected using a standard questionnaire. Healthy dietary score was calculated according to the recommendation of the American Heart Association with replacement of whole grains with bean consumption.[49] All participants underwent measurements for obesity [body mass index (BMI), waist circumference (WC), hip circumference, waist to hip ratio (WHP)], systolic blood pressure (SBP) and diastolic blood pressure (DBP), glycemic measures [fasting plasma glucose (FPG), 2-h postload plasma glucose (2h-PG), and HbA1c), HOMA of insulin resistance, liver enzymes [alanine aminotransferase (ALT) and aspartate aminotransferase AST)], and lipids [triglycerides (TG), total cholesterol (TC), high-density lipoprotein cholesterol (HDL) and low-density lipoprotein cholesterol (LDL)].

The study protocol was approved by the Institutional Review Board of Ruijin Hospital affiliated to the Shanghai Jiao-Tong University School of Medicine. Informed consent was obtained from study participants.

METHOD DETAILS

Definition of diabetes

At both baseline and follow-up visits, all participants underwent an oral glucose tolerance test (OGTT), and plasma glucose was obtained at 0 and 2 h during the test. Plasma glucose concentrations were analyzed locally within 2 h of the blood collection under a

stringent quality control program using the glucose oxidase or the hexokinase method. Using the 1999 WHO cutoffs, incident diabetes was defined as the following: FPG $\geq$126 mg/dL, 2h-PG $\geq$ 200 mg/dL, or self-reported previous diagnosis of diabetes by physicians and the current use of antidiabetic medications.

Amino acids and microbiota-related metabolites measurement

Fasting serum samples collected at baseline were analyzed for amino acids and microbiota-related metabolites using the UPLC-MS/MS with multiple reactions monitoring methods in a Nexera ×2 system (Shimadzu, Japan) coupled to a triple quadrupole mass spectrometer 8050 (Shimadzu, Japan), as described previously.[24,25]

Amino acids

For sample preparation, 4 times volume of methanol containing internal standards with proper concentration (alanine-d_3 3 μg/mL, phenylalanine-d_5 4 μg/mL, valine-d_8 1 μg/mL, leucine-d_3 1 μg/mL) was added to 70 μL serum for protein precipitation and amino acid extraction. A 60 μL aliquot of supernatant was centrifuged to dry at vacuum condition. The AccQ-Tag derivation kit was used to derivatize amino acids. Briefly, 70 μL borate buffer (pH = 8.8) was added to the dried extract and vortexed for 30 s. Then, 20 μL AccQ-Tag reagent was added to the mixture. The resulting solution was kept at ambient temperature and then incubated at 55°C for 10 min. After centrifuged and the supernatant was ready for LC-MS analysis.

For liquid chromatography separation, an UPLC C18 column (100 × 2.1 mm i.d.; 1.7 μm; Waters, USA) was used at 55 °C and a flow rate of 0.35 mL/min. The mobile phases were 98.5% water, 1% acetonitrile, 0.5% formic acid and 20 mM ammonium formate (A) and 1.6% formic acid and 98.4% acetonitrile (B). The gradient elution was set at 1% (v/v) B for 1.08 min, linearly increased to 9.1% B during the next 10.4 min, increased to 21.2% B during the next 16.3 min, rapidly increased to 59.6% B during the next 0.6 min maintained at this composition for 1.2 min, decreased to 1% B during the next 0.18 min, and finally maintained at this composition for an additional 0.18 min. The injection volume was 0.1 μL. The mass spectrometer was used in electrospray ionization (ESI) negative mode.

Microbiota-related metabolites

For sample preparation, 100 μL of methanol containing internal standards with proper concentration (Carnitine C0-d_3 0.25 μg/mL, Phenylalanine-d_5 2 μg/mL, Hippuric acid-d_5 1 μg/mL, Tryptophan-d_5 1 μg/mL, Choline-d_4 30 μg/mL, Carnitine C8-d_3 0.1 μg/mL) was added to an aliquot of 20 μL serum for protein precipitation and metabolite extraction. After the mixture was vortexed and centrifuged, the supernatant was centrifuged to dry at vacuum condition. The dried residue was reconstituted with 50 μL 25% acetonitrile before LC-MS analysis.

For liquid chromatography separation, an UPLC HSS T3 column (50 × 2.1 mm i.d.; 1.8 μm; Waters, USA) was used at 40 °C and a flow rate of 0.25 mL/min. The mobile phases were 1% formic acid water (A) and 0. 1% formic acid acetonitrile (B). The gradient elution was set at 0% (v/v) B for 2 min, rapidly increased to 25% B during the next 1 min, linearly increased to 35% B during the next 4 min, increased to 100% B during the next 0.5 min, flushed the system for 2.5 min, decreased to 0% B during the next 0.5 min, and finally maintained at this composition for an additional 2.5 min. The injection volume was 1 μL. The mass spectrometer was used in ESI negative mode.

For quality control, we performed 1) injection of a mixed quality control (QC) sample every 10 samples, 2) insertion of a blank sample every 10 samples, 3) procedure randomized order of samples and inclusion of positive and negative controls in each running batch, and 4) blind of the case-control status to the person running the sequence.

QUANTIFICATION AND STATISTICAL ANALYSIS

The demographic and clinical characteristics of the study population were described as frequencies and percentages as well as mean ± SD Differences between cases and controls were evaluated using the Student's *t* test for continuous variables with normal distribution, paired Wilcoxon rank-sum test for those with skewed distribution, and the $\chi 2$ test for categorical variables. The levels of serum metabolites were log-transformed before analysis. The associations of metabolites and risk of T2DM were estimated using multivariable conditional logistic regression models with adjustment of age, gender, BMI, smoking status, alcohol intake, physical activity, education attainment, family history of diabetes, SBP, and FPG. ORs (95%CI) were presented as per SD increment of the metabolites. Quartile analysis used the values of control group as cutoff and the lowest quartile as the reference. p values for false discovery rate (FDR) were estimated using the Benjamini-Hochberg method.

To validate the reliability of the identified metabolites associated with incident diabetes, random sampling was performed as a sensitivity analysis, and 200 times of tests were conducted by randomly selecting 80% of all samples as a subset for each time. Moreover, Internal validation test were performed in two randomly split subsets (50% participants in each subset).

Spearman correlation analysis was used to assess the association of individual amino acids and microbiota-related metabolites with the baseline clinical parameters, including obesity measurements (BMI, WC, hip circumference, and WHR), blood pressure, glycemic measures (FPG, 2h-PG, and HbA1c), HOMA of insulin resistance, liver enzymes (ALT and AST), and lipids (TG, TC, HDL and LDL). We further assessed the variance in metabolite levels explained by each clinical risk factors and measurements at baseline using variance partitioning as implemented in the R package variancePartition (v1.14.1).

Causal mediation analysis was performed by using the R package mediation (version 4.5.0). The models were formulated as follows (Y as outcome-incident T2DM, X as exposure -metabolites, M as mediation -risk factors): $Y = cX + e1$, $M = aX + e2$ and $Y = c'X + bM + e3$, where $c = c' + ab$, c as total effect of metabolites, c' as direct effect of metabolites, and ab as indirect effect of metabolites mediated by risk factor. Relationship groups which meet the following criteria were defined as significant causal relationships: (1) total effect of metabolites is significant ($P_{total\ effect} < 0.05$), (2) Mediated proportion >10%, and (3) indirect effect is significant ($P_{mediation} < 0.05$).

A two-tailed p value < 0.05 was considered statistically significant. Statistical analyses were performed using SAS 9.4 software (SAS Institute, Cary, NC) and R 3.3.1 software.